Inequality *by* Infrastructure: How Regulatory Data Infrastructures Produce Infrastructural Inequalities

Abstract

Government agencies increasingly rely on data infrastructures to govern access to welfare, healthcare, mobility, and security. These Regulatory Data Infrastructures (RDIs) – such as digital identity systems, biometric databases, and health platforms - have become central for public administrations and citizens alike. While these systems are often justified in terms of efficiency and security, they risk reproducing existing inequalities in access to welfare, health, rights and mobility through infrastructural inequalities. In this conceptual paper, we examine how RDIs produce infrastructural inequalities. Drawing on critical data and infrastructure studies, we introduce the concept of infrastructural inequalities to describe forms of discrimination and exclusion that emerge from the material and institutional arrangements of RDIs. We use this analytical tool and adopt a research-as-assemblage approach, comparing two theory-building case studies: biometric border control systems in Europe and health data infrastructures in India. We develop the concept through comparative, theory-building case analysis based on documentary and secondary sources. We identify recurring mechanisms, intervening at the level of the infrastructure, that exacerbate inequality, including 'crimmigration', limited possibilities for opting out and redress, scope creep, data legibility requirements, and data poverty. We also show how the persistence of classification over time through interoperability protocols, data recursivity, and algorithmic prediction contributes to the long-lasting character of inequality *by* infrastructure.

Introduction

Across jurisdictions, government agencies rely on data infrastructures to govern access to welfare, healthcare, mobility, and security. Digital identity systems, biometric databases, health information platforms, and interoperable migration registers – are some of the data infrastructures that have become central to contemporary public administration. These data infrastructures are often justified in terms of efficiency, fraud prevention, and security. Yet they also signal a deeper transformation in how governance is exercised: data infrastructures increasingly operate as a mode of rule in their own right, shaping the possibilities of state action through the affordances and constraints of technical systems[1].

This paper argues that these developments demand sustained critical scrutiny. Regulatory Data Infrastructures (RDIs) identify 'data-tracking systems that produce data in an automated fashion, [feeding] (quasi-)real-time decision-making and population monitoring within the remit of state functions, such as public safety, education, public health and population management'[2]. As regulatory data infrastructures are embedded in routine administrative practices, they reorganise the relationship between states and populations around infrastructural capacities for identification and classification. Do these systems improve access to rights and services, or do they amplify existing inequalities?

Our point of departure is the observation that much existing scholarship has focused on algorithmic bias, automated decision-making, and the opacity of artificial intelligence systems[3]. While indispensable, these approaches often overlook the broader data ecosystems in which RDIs operate. A data ecosystem is an interlinked collection of datasets and data infrastructures that share related data services, held together by governance arrangements, partnerships, standards, and organisational routines[4]. In the case of RDI ecosystems, the administrative routines, political narratives, policy frameworks, legal mandates, procurement contracts, technical standards[5], and interoperability protocols — to name just a few elements that

[1] Stefania Milan, 'Afterword: From Number Politics to Infrastructure Politics: Notes on Context and Methods', *The Cambridge Journal of Anthropology* 42, no. 1 (2024): 118–26, https://doi.org/10.3167/cja.2024.420108.
[2] Milan, 'Afterword'.
[3] Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, n.d.; Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press, 2020), https://doi.org/10.18574/nyu/9781479833641.001.0001. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, First Picador edition (Picador St. Martin's Press, 2019).
[4] Rob Kitchin, *Critical Data Studies: An A to Z Guide to Concepts and Methods*, First edition (Polity press, 2025).
[5] In this article, we treat standards as part of the infrastructural context, but we do not analyse standard-setting processes or technical specifications; our focus is on the inequality effects of infrastructural connectivity, gateways, and accountability fragmentation.

contribute to shape how data is generated and mobilised for governance. We situate RDIs within debates on data assemblages and data ecosystems, emphasising that governance increasingly unfolds through infrastructural possibilities, that is to say what can be collected, linked, stored, and acted upon at scale.

Drawing on the emerging interdisciplinary field of critical data studies (CDS) and critical infrastructure studies (CIS), we examine how Regulatory Data Infrastructures produce infrastructural inequalities. We introduce the working concept of infrastructural inequalities to capture the discriminatory and exclusionary effects that arise from the materiality and infrastructural arrangements of regulatory data infrastructures. We operationalise this approach through two comparative case studies that examine different RDI assemblages - EU's interoperable biometric border infrastructure and India's Ayushman Bharat Digital Mission (ABDM), the national health data infrastructure. These cases were selected to illuminate how similar infrastructural logics operate across healthcare and security domains, and how inequalities manifest in each setting, as further explained in the case selection section of this paper.

Comparing the EU's biometric border infrastructure with India's health data infrastructure reveals two distinct but related modes of infrastructural inequality—hyper-legibility and uneven legibility.[6] Across both cases, we identify recurring mechanisms that sustain these inequalities: scope creep through interoperability, constrained opt-out and redress under diffuse accountability, standardisation around an 'ideal' data subject[7], data poverty[8], and the recursive persistence of classifications over time. In the EU case, crimmigration[9] emerges as an additional domain-specific mechanism linking migration control with criminal justice infrastructures.

With this article, we contribute to the interdisciplinary scholarship on infrastructural inequalities and algorithmic discrimination by identifying the mechanisms and elements

---

[6] By hyper-legibility we mean intensified, cross-domain visibility that makes suspicion durable; by uneven legibility we mean partial/conditional visibility that makes access contingent on producing 'recognised' data traces.

[7] Stefania Milan, 'Techno-Solutionism and the Standard Human in the Making of the COVID-19 Pandemic', *Big Data & Society* 7, no. 2 (2020): 2053951720966781, https://doi.org/10.1177/2053951720966781.

[8] Stefania Milan and Emiliano Treré, 'The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins', *Social Media + Society* 6, no. 3 (2020): 2056305120948233, https://doi.org/10.1177/2056305120948233.

[9] Nina Amelung, ''Crimmigration Control' across Borders: The Convergence of Migration and Crime Control through Transnational Biometric Databases', *Historical Social Research / Historische Sozialforschung* 46, no. 3 (2021): 151–77; Katja Franko, *The Crimmigrant Other: Migration and Penal Power*, 1st edn (Routledge, 2019), https://doi.org/10.4324/9781351001441.

through which RDIs risk exacerbating inequality. We extend algorithmic bias debates[10] by relocating inequality from the software to the infrastructural level, including the material level of data infrastructures – the hardware components – as well as the institutional arrangements – standards, policies and protocols that define the rules of functioning of regulatory data infrastructures. Likewise, we contribute to the scholarship on digital governance and Digital Public Infrastructures (DPIs)[11], by identifying some of the mechanisms through which infrastructural choices produce structural inequalities.

The article proceeds as follows. We first develop our theoretical framework, drawing on work in critical data studies and infrastructure studies. Then, we outline our research design, which treats research itself as an assemblage of knowledges, actors, and practices. We apply this framework to present the two case studies, biometric border control systems in Europe and health data infrastructures in India. In the concluding sections, we discuss the implications of infrastructural inequalities for data protection, accountability, and democratic governance.

Theoretical Background: Unveiling Infrastructural Inequalities

Datafication and Data Ecosystems

Our work builds on CDS, which starts from the premise that data and data practices are never neutral: they are embedded in sociotechnical imaginaries, institutional arrangements, and historically patterned relations of power and inequality[12]. CDS treats data as a primary research object and asks how data are produced, shared, governed, and made actionable, by whom, and for whose purposes[13]. Rather than adopting technological determinism or taking 'data-driven'

---

[10] Eubanks, *Automating Inequality*, First Picador edition (Picador St. Martin's Press, 2019); Buolamwini and Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*; Noble, *Algorithms of Oppression*.

[11] Lucas Michael Haitsma and Maarten Bouwmeester, 'Learning from Control Deficits in the Childcare Benefits Scandal: A Plea for Multi-Level Analysis in Law and Policy Research', *Recht Der Werkelijkheid* 44, no. 3 (2023): 57–68, https://doi.org/10.5553/RdW/138064242023044003004. Milan and Treré, 'The Rise of the Data Poor'; Ayona Datta, 'The Digital Turn in Postcolonial Urbanism: Smart Citizenship in the Making of India's 100 Smart Cities', *Transactions of the Institute of British Geographers* 43, no. 3 (2018): 405–19, https://doi.org/10.1111/tran.12225.

[12] Danah Boyd and Kate Crawford, 'CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication & Society* 15, no. 5 (2012): 662–79, https://doi.org/10.1080/1369118X.2012.678878. David Beer, 'Social Network(Ing) Sites…revisiting the Story so Far: A Response to Danah Boyd & Nicole Ellison', *Journal of Computer-Mediated Communication* 13, no. 2 (2008): 516–29, https://doi.org/10.1111/j.1083-6101.2008.00408.x.

[13] Andrew Iliadis and Federica Russo, 'Critical Data Studies: An Introduction', *Big Data & Society* 3, no. 2 (2016): 2053951716674238, https://doi.org/10.1177/2053951716674238; Rob Kitchin, *Critical Data Studies: An A to Z Guide to Concepts and Methods*, First edition (Polity press, 2025); Andreas Hepp et al., 'New Perspectives in Critical Data Studies: The Ambivalences of Data Power—An Introduction', in *New Perspectives*

claims at face value, we aim to produce reflexive and situated knowledge about what data do in society: how they reorder visibility, authority, accountability, and participation.

Datafication provides the first building block for our work. Van Dijck defines datafication as the transformation of social action into quantified data that can be tracked, compared, and analysed in real time, including for predictive purposes[14]. Hepp et al. emphasise the 'double processuality' of datafication: it is simultaneously a situated set of translation practices and a broader societal transformation as quantified data become pervasive[15]. This translation into data is never a simple 'capture' of reality; it entails complex, interest-driven operations and technical articulations of people and practices and relationships[16].

Crucially, the relation between data and the social world is recursive: data are used to act upon social life, and these interventions reshape the conditions under which future data are generated. For instance, when a biometric border system flags certain travel patterns as 'suspicious,' these classifications alter how individuals move, which documents they carry, and how authorities interact with them, thereby also changing the data traces that future algorithmic assessments will analyse. This recursive dynamic reproduces or intensifies inequality and surveillance[17], depending on how data infrastructures are configured and governed.

This recursive quality of datafication also reveals inherent tensions with foundational data protection principles – such as data minimization and purpose limitation, central in the regulatory frameworks of both contexts of interest. The phenomenon of 'purpose creep', for example, where data collected for one purpose are progressively repurposed for others, is not an aberration but a structural feature of data-driven systems[18], despite the principle of purpose limitation requires that data be collected for specified, explicit and legitimate purposes. This principle is fundamentally at odds with large-scale data ecosystems[19], such as those of

---

in Critical Data Studies, ed. Andreas Hepp et al., Transforming Communications – Studies in Cross-Media Research (Springer International Publishing, 2022), https://doi.org/10.1007/978-3-030-96180-0_1. Annika Richterich, *The Big Data Agenda: Data Ethics and Critical Data Studies* (University of Westminster Press, 2018), https://doi.org/10.16997/book14.

[14] Teun A. Van Dijk, 'Structures of Discourse and Structures of Power', *Annals of the International Communication Association* 12, no. 1 (1989): 18–59, https://doi.org/10.1080/23808985.1989.11678711.

[15] Hepp et al., 'New Perspectives in Critical Data Studies'.

[16] Kevin D. Haggerty and Richard V. Ericson, 'The Surveillant Assemblage', *The British Journal of Sociology* 51, no. 4 (2000): 605–22, https://doi.org/10.1080/00071310020015280. Evelyn Ruppert, 'Population Objects: Interpassive Subjects', *Sociology* 45, no. 2 (2011): 218–33, https://doi.org/10.1177/0038038510394027.

[17] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, First Picador edition (Picador St. Martin's Press, 2019). Noble, *Algorithms of Oppression*.

[18] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences', *Columbia Business Law Review*, 1 May 2019, 494–620 Pages, https://doi.org/10.7916/CBLR.V2019I2.3424.

[19] Tal Zarsky, *Incompatible: The GDPR in the Age of Big Data*, sect. 2, 47, no. 4 (2017), https://ssrn.com/abstract=3022646.

regulatory data infrastructures. Purpose limitation assumes data uses can be predetermined and contained, yet recursive datafication operates precisely through continuous repurposing: data generated from one governance intervention become inputs for subsequent interventions, creating feedback loops that extend far beyond initial specifications.

Datafication scales through interdependence – as exposed by data ecosystems. Data ecosystems identify the interlinked collection of datasets and infrastructures that share related data services, held together by institutional and governance arrangements and working partnerships[20]. Mapping a data ecosystem involves tracing data mobilities: how data move across organisations and technical layers, where linkages are established, where responsibilities are distributed or obscured. These distributed arrangements pose acute challenges for data protection's accountability frameworks which assume clearly delineated data controllers responsible for specific processing operations. When data flow across organisational boundaries and technical layers, determining accountability for ensuring lawfulness becomes fundamentally ambiguous.

This fragmentation of responsibility renders individual rights—such as the right to access or correct one's data—structurally difficult to exercise, as neither data subjects nor regulators can easily trace which actors control which data at which point in the ecosystem. We use the notion of data ecosystems as a lens to analyse the heterogeneity of these arrangements. In this paper, they allow us to trace how infrastructural inequalities may be generated when access, recognition, and contestation become increasingly contingent on what the infrastructure can see, link, and validate.

Regulatory data infrastructures: Towards a new mode of governance

In the aftermath of the Covid-19 pandemic, regulatory data infrastructures have become central state technologies for managing and monitoring populations[21]. Health dashboards, facial recognition systems, and digital identity systems – among other data-driven systems - operate through data assemblages that classify, count, and sort people. We use the term RDIs to describe data-driven infrastructures that do not merely store, process, or circulate data, but actively shape the polity.

[20] Kitchin, *Critical Data Studies*.
[21] Milan, 'Afterword'.

RDIs function as gateways to welfare, public services, and rights, mediating interactions between states and populations. Digital identity systems and biometric recognition are increasingly central to access health records, education, and employment. In some cases, for example UK's policy on eVisa for migrant, even mandatory[22]. When data infrastructures become mandatory gateways to essential services, the data protection requirement of freely given consent becomes structurally untenable. Recent EU-focused scholarship on cookie paywalls and 'pay-or-consent' models shows how access conditionality and refusal penalties erode the possibility of freely given consent.[23] This logic extends to RDIs: when biometric enrolment or digital identity registration becomes the only pathway to healthcare, education, or legal status, the legal fiction of 'consent' collapses entirely[24].

In this context, bureaucratic decision-making is increasingly replaced by systems presented as neutral, apolitical, and technical, framed primarily in terms of efficiency, security, and control[25]. Smith's notion of *data doxa* captures how digital data, alongside the platforms and devices that stage it, come to be perceived as normal, necessary, and enabling daily life activities within contemporary societies[26]. While data doxa captures how people rapidly become used to datafied realities, *dataism* reflects the cognitive (and political) belief that data and data-driven decision-making is neutral, objective, and ultimately reliable[27]. Together, data doxa and dataism operate as material and discursive justifications for data-driven governance, rendering it necessary and legitimate, while alternative modes of governance become difficult to imagine[28].

As documented by Virginia Eubanks[29], automated eligibility systems, coordinated databases, and predictive risk models increasingly replace human assemblages and bureaucratic systems in the assignation of rights, services, and resources. Although presented as objective, these systems produce patterned errors. A *false positive* occurs when individuals are wrongly

[22] Derya Ozkul and Marie Godin, *Exclusion by Design: Digital Identification and the Hostile Environment for Migrants* (ESRC Digital Good Network, 2025), https://doi.org/10.31273/9781911675211.
[23] Victor Morel et al., 'Your Consent Is Worth 75 Euros A Year - Measurement and Lawfulness of Cookie Paywalls', *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, 7 November 2022, 213–18, https://doi.org/10.1145/3559613.3563205; Nafiye Yücedağ et al., 'Pay or Consent Models in Europe: Already Outdated or an Overlooked Crisis in Freely Given Consent?', *Technology and Regulation* 2025 (July 2025): 336–54, https://doi.org/10.71265/8jjzhc21.
[24] Paul De Hert and Georgios Bouchagiar, 'Visual and Biometric Surveillance in the EU. Saying 'No' to Mass Surveillance Practices?12', *Information Polity* 27, no. 2 (2022): 193–217, https://doi.org/10.3233/IP-211525.
[25] Milan, 'Afterword'.
[26] Gavin J.D. Smith, 'Data Doxa: The Affective Consequences of Data Practices', *Big Data & Society* 5, no. 1 (2018): 2053951717751551, https://doi.org/10.1177/2053951717751551.
[27] Eubanks, *Automating Inequality*, First Picador edition (Picador St. Martin's Press, 2019).
[28] Eubanks, *Automating Inequality*. Smith, 'Data Doxa'.
[29] Eubanks, *Automating Inequality*, First Picador edition (Picador St. Martin's Press, 2019).

identified as matching a target, such as being flagged as fraudulent or criminal. Conversely, a *false negative* occurs when systems fail to recognise individuals who should be recognised, leading to exclusion and invisibility. If individuals are not legible to the infrastructure, they effectively do not exist, and when they are misclassified, opportunities for contestation are limited or absent. The three case studies analysed by Eubanks – state implementation of automated eligibility systems, coordinated databases, and predictive risk models in the US - show that these mistakes disproportionally affect the poor, reproducing layers of marginalisation whose effects last across generations.

As represented by Eubanks with the 'digital poorhouse', these unequal systems build upon long-lasting infrastructures and discourses that serve to 'target, track, and punish' the already marginalised[30]. The reproduction of class inequalities is embedded in the infrastructure. Similarly, Joy Buolamwini has shown that facial recognition systems systematically fail to recognise darker-skinned faces, a consequence of training datasets that underrepresent Black individuals[31]. As Buolamwini, Safiya Noble, and Ruha Benjamin argue, algorithmic systems tend to reproduce the same racist and sexist biases that exist in the societies where these systems where created[32].

The 2020 scandal in the Dutch childcare benefits illustrates the bias encoded in algorithmic design. An algorithmic system was implemented to identify fraudulent benefit claims through risk profiling, as an objective tool for protecting public resources. In fact, tens of thousands of parents, predominantly from low-income households, were falsely accused of fraud. Foreign nationality was explicitly included as a risk factor. Trained on historical datasets already shaped by institutional bias, the system reproduced historical class inequalities linked with colonial oppression, transforming them into algorithmic inevitability[33].

This raises a central question for this paper: How do you regulatory data infrastructures produce infrastructural inequalities? This question grounds our research. We hypothesise that as RDIs increasingly mediate access to public services and rights, they tend to materialise and reinforce pre-existing inequalities through infrastructural choices.

[30] Eubanks, *Automating Inequality*, First Picador edition (Picador St. Martin's Press, 2019).
[31] Buolamwini, Joy. *Unmasking AI: My mission to protect what is human in a world of machines*. Random House, 2024.
[32] Buolamwini, Joy. *Unmasking AI.* Noble, Safiya Umoja. 'Algorithms of oppression: How search engines reinforce racism.' *Algorithms of oppression*. New York university press, 2018. Benjamin, Ruha. 'Race after technology.' *Social Theory Re-Wired*. Routledge, 2023. 405-415.
[33] Haitsma and Bouwmeester, 'Learning from Control Deficits in the Childcare Benefits Scandal'.

Infrastructural inequalities: definition and concerns

Algorithmic racism captures discriminatory outcomes generated by automated systems that disadvantage racialised groups, often because of biased datasets or entrenched social hierarchies[34]. Similarly, technological redlining highlights how data-driven systems opaquely reproduce social discrimination[35], drawing on the historical practice in which banks used neighbourhood demographics —particularly race and ethnicity— to deny loans, rather than individual creditworthiness[36]. Data poverty refers to systematic absences in datasets that render individuals or communities invisible to algorithmic systems, and thus unable to make rights claims[37]. We argue that, however important, these concepts remain narrowly set on specific, sectorial dynamics of data-mediated classification and decision-making, failing to capture the systemic embedding of these dynamics in contemporary governance arrangements.

We therefore advance the concept of infrastructural inequalities. We define infrastructural inequalities the forms of discrimination, exclusion, and oppression that originate from and depend on infrastructural arrangements. From a Critical Data Studies perspective, the concept does three things. First, it foregrounds the structural character of inequalities produced through data infrastructures, showing how disadvantage is captured, organized, stabilized but also potentially produced through the arrangements and joint action of technical *and* institutional systems, rather than technical systems alone. Second, it extends narrower frameworks such as algorithmic racism, by directing attention to (a) the recurrent and recursive aftereffects of regulatory data infrastructures beyond individual systems or moments of decision-making, and (b) the fact that such effects can extend across the population, even as they are unevenly distributed and experienced. Thirds, it enables analysis across multiple layers – including, but not limited to, algorithmic processing – such as data production, storage, system design, and legal or policy mandates, rather than focusing on algorithmic mediation alone.

Interestingly, research on infrastructural inequalities first emerged within studies of networked systems in urban settings, such as water provision, transportation, electricity distribution, and waste disposal. The collection of essays edited by Grealy et al. (2019) exposes how settler/colonial regimes of governance, technologies of calculation and redistribution, and the

[34] Buolamwini, Joy. *Unmasking AI: My mission to protect what is human in a world of machines*. Random House, 2024.
[35] Noble, Safiya Umoja. "Algorithms of oppression: How search engines reinforce racism." *Algorithms of oppression*. New York university press, 2018.
[36] Catherine D'Ignazio and Lauren F. Klein, *Data Feminism*, Strong Ideas (The MIT Press, 2020).
[37] Stefania Milan and Emiliano Treré, 'Big Data from the South(s): Beyond Data Universalism', *Television & New Media* 20, no. 4 (2019): 319–35, https://doi.org/10.1177/1527476419837739.

political economy of public and private ownership result in deep inequalities in the distribution of resources, amenities, and opportunities[38]. Today, concerns over infrastructural inequalities spans critical data studies, human geography, and political science, and suggests that infrastructural inequalities are rising [39]. This interdisciplinarity is necessary because infrastructural problems affect more than one realm of human and governmental activity.

Large segments of the population experience inadequate access to employment, housing, education, nutrition, and healthcare[40]. Structural inequalities capture these durable asymmetries along lines of class, race, geography, or legal status. Infrastructural inequalities, however, direct attention to a distinct analytical layer: how disadvantage is captured, then organized, consequently stabilized and, we argue, reproduced (and even produced) through data infrastructures themselves. Inequality becomes embedded in cables, pipes, interfaces, databases, legal mandates, and institutional arrangements. For example, data centres exert extensive influence over digital activity while remaining spatially and politically distant from the communities they affect. Moreover, they often operate as opaque systems with limited public accountability[41]. Thus, control over data storage and processing is concentrated in a small number of states and corporations, reinforcing contemporary forms of data colonialism – which is to say, how global data extraction parallels historical colonial exploitation[42]. In this sense, infrastructural inequalities are not reducible to structural inequalities alone.

The contemporary shift toward data-driven governance has the potential to reproduce structural inequalities *by* infrastructure, and to produce new forms of inequality specific to the characteristics of these infrastructures. This article extends the line of inquiry concerned with algorithmic discrimination, to examine regulatory data infrastructures as assemblages, and by grounding infrastructural inequalities through comparative analysis.

Research-as-Assemblage Approach

In this study, we use infrastructural inequalities as a heuristic to examine how regulatory data infrastructures (RDIs) shape inequalities through the socio-technical conditions they establish

[38] Liam Grealy, Andrew Brooks, Astrid Lorange, Christen Cornell, and Tess Lea (2019). *Introduction: Tending a Social Infrastructure*. Infrastructural Inequalities Special Issue.
[39] Bhartendu Pandey et al., 'Rising Infrastructure Inequalities Accompany Urbanization and Economic Development', *Nature Communications* 16, no. 1 (2025): 1193, https://doi.org/10.1038/s41467-025-56539-w.
[40] David Machin and John E. Richardson, 'Renewing an Academic Interest in Structural Inequalities', *Critical Discourse Studies* 5, no. 4 (2008): 281–87, https://doi.org/10.1080/17405900802405148.
[41] Kynan Tan, *Polymorphism (Data Centre Simulation)*, n.d. Infrastructural Inequalities Special Issue.
[42] Nick Couldry and Ulises A. Mejias, 'Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject', *Television & New Media* 20, no. 4 (2019): 336–49, https://doi.org/10.1177/1527476418796632.

for accessing rights, services, and mobility. We adopt a comparative, case-based approach to advance the conceptual argument that RDIs can produce durable forms of inequality at the infrastructural level. The cases function as theory-building examples: they allow us to identify and compare recurrent mechanisms through which inequality becomes embedded in, and stabilised by, infrastructural arrangements. The analytical contribution is twofold: first, to develop the concept of regulatory data infrastructures by showing these are not merely technical systems but assemblages of institutional mandates, standards, and organisational arrangements; second, to advance the notion of infrastructural inequalities by examining how such assemblages stabilise existing structural inequalities as durable infrastructural conditions.

Consistent with our theoretical emphasis on data assemblages, and following established traditions in comparative qualitative research[43], as well as Critical Data Studies' call for situated and context-sensitive analysis[44], we treat cases as an analytical lens on a broader configuration of legal mandates, institutional arrangements, standards, and data practices through which governance is enacted and rendered operational. Following Kitchin, we conceptualise each case as "contingent, ontogenetic, and performative in nature", where ontogenetic refers to the fact that data infrastructures are always "in a state of becoming".[45] This perspective foregrounds RDIs as evolving ecosystems whose effects cannot be reduced to single systems or decision points, but emerge from how components connect across organisations, interfaces, and regulatory layers. The comparative approach therefore focuses on infrastructural linkages—such as interoperability, enrolment requirements, recursive data reuse, and distributed accountability—through which RDIs organise access, allocate responsibility, and shape possibilities for contestation and remedy.

The two cases—interoperable biometric border control infrastructures in the EU and health data infrastructures under India's Ayushman Bharat Digital Mission—were selected through theoretical sampling. They represent contrasting regulatory domains (security/migration and healthcare) and distinct political-institutional settings (a supranational regulatory regime and a large federal state), while sharing key characteristics that make them analytically comparable as RDIs. In both cases, access to entitlements is increasingly mediated through (quasi-)mandatory infrastructural gateways, and accountability is distributed across multiple

[43] Yin, Robert K., *Case Study Research: Design and Methods*, vol. 5 (SAGE Publications Ltd, 2009).
[44] Andrew Iliadis and Federica Russo, 'Critical Data Studies: An Introduction', *Big Data & Society* 3, no. 2 (2016): 2053951716674238, https://doi.org/10.1177/2053951716674238.
[45] Rob Kitchin, 'Thinking Critically about and Researching Algorithms', *Information, Communication & Society* 20, no. 1 (2017): 21, https://doi.org/10.1080/1369118X.2016.1154087.

public and private actors and layers. Both cases also operate against the backdrop of data-protection regimes that foreground principles such as purpose limitation, minimisation, and accountability (respectively, the EU GDPR and India's DPDP), which sharpens the analysis of how data-protection aims interact with infrastructural mechanisms and constraints.

This combination of difference and structural similarity allows us to examine how infrastructural inequalities emerge not only from domain-specific policy choices but from underlying infrastructural arrangements that travel across contexts. Comparing the two cases makes it possible to distinguish between different modes of infrastructural inequality—what we conceptualise as hyper-legibility and uneven legibility—while identifying a shared set of inequality-producing mechanisms. These include scope creep through interoperability, constrained opt-out and redress, standardisation around an "ideal" data subject, data poverty, and the recursive persistence of classifications over time. The cases are therefore presented not as exhaustive or exceptional, but as strategic, theory-building sites chosen because they foreground dynamics increasingly characteristic of governance by data infrastructure more broadly—enabling conceptual elaboration and mechanism identification rather than empirical coverage. The contribution lies in specifying where inequality is produced (at the infrastructural level), how it is produced, reproduced, and sustained through recurring mechanisms, and why it remains durable across domains and jurisdictions.

Case Studies: Tracing Infrastructural Inequalities

Biometric borders: Interoperability and Dataveillance for Crimmigration in the EU

Interoperability politics as the scaffolding of crimmigration

The interoperability initiative, adopted by the European Union in 2019 and implemented by Directorate-General for Migration and Home Affairs (DG HOME), seeks to connect all security and migration databases across the Union. These are the Entry-Exit System (EES), Visa Information System (VIS), European Travel Information and Authorisation System (ETIAS), European Asylum Dactyloscopie Database (Eurodac), the Schengen Information System (SIS), and European Criminal Records Information System - Third Country Nationals (ECRIS-TCN). The goal of this interoperable ecosystem is to create a supra-layer to allow for cross validation of identities and information for streamlined border crossing[46]. The idea is that border security officials would be able to get more complete and reliable information about

[46] Matthias Leese, 'Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU', *Geopolitics* 27, no. 1 (2022): 113–33, https://doi.org/10.1080/14650045.2020.1830764.

border crossers. Trauttmansdorff[47] explains that the justification for such an initiative is driven by narratives of efficiency and security: the fragmented databases are framed as error-prone, incomplete, and at risk of missing threats as they do not communicate with one another. Interoperability is instead presented as the solution: connecting databases improves fraud detection, allows for more complete security screenings, and faster border crossings through automated facial recognition.

From the EU perspective, interoperability appears as a neutral technical solution. For example, linking EES, which stores biometric data of third-country nationals entering and exiting the Schengen area, with SIS, which contains police alerts and arrest warrants, is presented as a means to prevent criminals from exploiting jurisdictional and informational gaps between policing institutions and member states[48]. However, this narrative fails to capture the full picture of interoperability's impact on fundamental rights of individuals. We argue that it reorganises EU's border regime at an infrastructural level where new forms of inequality can become embedded in the technical architectures.

One of the clearest consequences of these infrastructural inequalities is the worsening of crimmigration control. Crimmigration is 'the convergence of migration and crime control – together with the enactment of the figure of the so-called 'crimmigrant other': migrants are structurally positioned as potential criminals[49]. Amelung[50] illustrates how this is occurring at the EU level through Eurodac, a database originally designed to manage biometric identifiers for irregular migrants and asylum seekers, which is increasingly accessible to law enforcement authorities for criminal investigations. Through biometric matching across databases, asylum seekers and irregular migrants become permanent criminal suspects —even when they have done nothing wrong.

This increases the risk that these 'vulnerabilized' people[51] will be 'matched' by accident as biometrics dataveillance is tied to a probabilistic matching algorithm. More comparisons and

[47] Paul Trauttmansdorff, 'The Fabrication of a Necessary Policy Fiction: The Interoperability 'Solution' for Biometric Borders', *Critical Policy Studies* 17, no. 3 (2023): 428–46, https://doi.org/10.1080/19460171.2022.2147851.

[48] Amelung, ''Crimmigration Control' across Borders'; Matthias Leese, 'Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU', *Geopolitics* 27, no. 1 (2022): 113–33, https://doi.org/10.1080/14650045.2020.1830764; Trauttmansdorff, 'The Fabrication of a Necessary Policy Fiction'.

[49] Franko, Katja. *The crimmigrant other: Migration and penal power*. Routledge, 2019.

[50] Amelung, ''Crimmigration Control' across Borders'.

[51] Camryn M. Garrett and Rochelle Altman, 'Vulnerabilized: Revisiting the Language of the Vulnerable Populations Framework', *American Journal of Public Health* 114, no. 2 (2024): 177–79, https://doi.org/10.2105/AJPH.2023.307532.

larger databases lead to higher chances of false positives, especially if the system prioritises fast actionability over accuracy[52]. This is particularly detrimental as suspicion becomes 'sticky', meaning that a single match becomes part of the persons' data profile and will repeatedly be re-used in future risk assessments[53]. The burden then falls on the individual to correct the 'bad datapoint'. However, people' right to redress is reduced: interoperability distributes the data held on individuals across multiple entities and datasets, and the risk assessments across multiple interconnected systems and national jurisdictions, making it difficult for individuals and bureaucrats to correct the wrong entries[54]. In such cases, rather than eliminating bad error points, interoperability stabilizes them and makes them harder to repair through their propagation across a network of databases.

Biometric borders: infrastructures of compassionate repression

Thus, the interoperability system imposes certain political visions of governance into its technical architecture. It treats distinct regulatory domains -migration, border control, criminal justice- as one data environment, creating a security ecosystem where individuals, especially migrants, are continuously assessed and checked as criminals. Or with heightened scrutiny as the new default. This shift also places mass dataveillance as a dominant mode of border governance.

Dataveillance is 'the systematic and large-scale gathering, analysis and use of electronic information on persons'[55]. The logic of dataveillance is to have a standardised, efficient administrative process to assuring that border crossers are 'risk free' and belong in the EU, which requires identification to be able to link a physical person to their digital selves while continuously monitor their behaviours to make accurate risk assessments. Interoperability amplifies dataveillance by using biometric technologies to link identities across databases, thus repurposing data across contexts[56]. Moreover, it turns border checks into an ongoing process of surveillance that goes beyond the physical action of crossing a border, as each database updates in different contexts and times, with all these updates impacting the risk assessment for when the physical border crossing occurs. Therefore, individuals are no longer assessed on

[52] Leese, 'Fixing State Vision'.
[53] Julien Jeandesboz, 'Smartening Border Security in the European Union: An Associational Inquiry', *Security Dialogue* 47, no. 4 (2016): 292–309.
[54] Jeandesboz, 'Smartening Border Security in the European Union'; Leese, 'Fixing State Vision'.
[55] Jeandesboz, 'Smartening Border Security in the European Union'.
[56] Leese, 'Fixing State Vision'.

the basis of documents presented at a checkpoint but through continuously updated digital profiles[57].

Linking the increased dataveillance to structural inequality, Leurs & Shepherd argue that datafication is never neutral, it reflects and reinforces existing power relations[58]. In the context of the EU border regime, those deemed 'data-ready,' EU citizens and trusted travellers, will get faster border crossing lines through the automation of risk assessments. The 'other' is subjected to increased crimmigration and resulting extra dataveillance as a required condition for crossing the border. This means that mobility within Schengen becomes a continuous game of fitting into changing risk profiles that are recalculated across interoperable databases[59].

This result closely mirrors what Iazzolino[60] describes as infrastructures of compassionate repression. Using the case of biometric verification systems in the humanitarian context, a system justified through efficiency and reducing fraud, she demonstrates how these systems actually restrict the behaviours of the most vulnerable. Conditions to get access, in this case food, are never reflective of real life but simplifications and 'ideal cases', but real individuals who should be entitled might not necessarily conform to the ideal, due to contextual realities. As such, these 'benevolent technologies' actually restrict individuals by automating exclusion and ignoring contextual realities as no individual fits into the 'ideal case'.

In this sense, crimmigration and its consequences for the most vulnerable intersect with broader developments in interoperability and regulatory data infrastructures. Interoperability automates the decision-making process of who can enter and not, based on risk assessments that use data from previously separated databases. This infrastructural choice is responsible for increasing dataveillance from a single-event check to continuously updated checks, based on cross-sectional databases. As migrants' data is increasingly compared with criminal investigation, the process of crimmigration unfolds. This shifts inequality from discretionary decision-making to the infrastructural level, where unequal treatment is produced automatically through continuous surveillance, cross-database matching, and risk profiling that disproportionately targets migrants as default security subjects.

---

[57] Jeandesboz, 'Smartening Border Security in the European Union'; Koen Leurs and Tamara Shepherd, 'Datafication & Discrimination', in *The Datafied Society*, ed. Mirko Tobias Schäfer and Karin van Es, Studying Culture through Data (Amsterdam University Press, 2017), https://www.jstor.org/stable/j.ctt1v2xsqn.20.
[58] Leurs and Shepherd, 'Datafication & Discrimination'.
[59] Amelung, '"Crimmigration Control" across Borders'; Leese, 'Fixing State Vision'; Jeandesboz, 'Smartening Border Security in the European Union'.
[60] Gianluca Iazzolino, 'Infrastructure of Compassionate Repression: Making Sense of Biometrics in Kakuma Refugee Camp', *Information Technology for Development* 27, no. 1 (2021): 111–28, https://doi.org/10.1080/02681102.2020.1816881.

Health Data Infrastructure and Uneven Legibility in India: The Case of ABDM

The Context of Ayushmann Bharat Digital Mission

India's Ayushman Bharat Digital Mission, launched in September 2021, represents an ambitious national health data infrastructure and a core component of India's broader DPI agenda[61]. ABDM is framed by the state as an interoperable, federated, and consent-based digital infrastructure designed to enable portability of health records, continuity of care, and efficiency across India's highly fragmented healthcare system. Central components include Health IDs, standardised electronic health records (EHRs), and interoperability frameworks enabling data exchange across public and private actors (G20 GDPIR, n.d.).

ABDM does not directly provide healthcare services. Instead, it functions as a RDI that reorganises health governance around data flows, standards, and interoperability requirements. Participation is formally voluntary, and consent is positioned as a key safeguard. However, as ABDM progressively becomes embedded in insurance systems, hospital workflows, and public health programmes, participation becomes infrastructurally necessary for both institutions and patients seeking seamless access to care[62] . The case of ABDM therefor highlights how RDIs govern indirectly, and actively by shape the conditions under which actors can meaningfully participate in healthcare systems.

Health data poverty as structurally produced by institutional stratification

A central mechanism through which ABDM reproduces infrastructural inequality is data poverty of health data i.e., 'the inability for individuals, groups, or populations to benefit from a discovery or innovation due to a scarcity of data that are adequately representative'[63]. The generation of health data under ABDM depends on digitised hospitals, EHR-compatible systems, backend integration, and trained administrative staff. However, these infrastructural

[61] R. S. Sharma et al., 'The Ayushman Bharat Digital Mission (ABDM): Making of India's Digital Health Story', *CSI Transactions on ICT* 11, no. 1 (2023): 3–9, https://doi.org/10.1007/s40012-023-00375-0.
[62] Shubharanjan Jena et al., 'Integration of National Cancer Registry Program with Ayushman Bharat Digital Mission in India: A Necessity or an Option', *Public Health in Practice* 3 (June 2022): 100263, https://doi.org/10.1016/j.puhip.2022.100263.
[63] Milan and Treré, 'Big Data from the South(s)'; Catherine D'Ignazio and Lauren F. Klein, *Data Feminism*, Strong Ideas Series (The MIT Press, 2020); Hussein Ibrahim et al., 'Health Data Poverty: An Assailable Barrier to Equitable Digital Health Care', *The Lancet Digital Health* 3, no. 4 (2021): e260–65, https://doi.org/10.1016/S2589-7500(20)30317-4.

preconditions are unevenly distributed across India's healthcare landscape, stratified along urban-rural and formal-informal care axes.

Urban and formally institutionalised healthcare centres are structurally advantaged within ABDM. These sites tend to generate high volumes of standardised clinical data, operate digital information systems, and possess the financial and technical capacity to integrate with ABDM protocols. As a result, they produce dense, continuous data flows that are readily incorporated into data exchanges, insurance workflows, and policy analytics, reinforcing their visibility within health governance systems[64].

However, rural healthcare facilities, peripheral clinics, and informal or community-based care providers often rely on paper records, fragmented documentation practices, and limited digital infrastructure. Informal and community-based care, which is arguably central to healthcare delivery for large segments of the population, frequently remains non-interoperable and excluded from formal data flows[65]. These disparities produce data poverty not as an individual failure to participate, but as a structurally generated outcome of healthcare market stratification[66].

These exclusions are not merely technical gaps but reflect what scholars of healthcare digitalization describe as a reordering of moral and epistemic priorities, in which standardised, interoperable data forms are privileged while informal, relational, and paper-based modes of care are rendered invisible within digital systems[67]. Data infrastructures such as ABDM therefore institutionalise this hierarchy of legibility, in which certain forms of healthcare, and the populations reliant on them, are systematically underrepresented in health data governance[68].

Recursive data loops in public health governance

Infrastructural inequalities within ABDM are further reinforced through recursive data loops that are characteristic of contemporary data-driven public health governance. Data generated through healthcare encounters feeds back into policy decisions, resource allocation, clinical

[64] Sharma et al., 'The Ayushman Bharat Digital Mission (ABDM)'.
[65] Nisha B Jain and Samiran Nundy, 'Electronic Health Records in India: Challenges and Promises', *Journal of Medical Evidence* 2, no. 3 (2021): 278–79, https://doi.org/10.4103/JME.JME_94_21.
[66] Milan and Treré, 'Big Data from the South(s)'.i
[67] Ian P. McLoughlin et al., *The Digitalization of Healthcare* (2017).
[68] Kitchin, *Critical Data Studies*; James C. Scott, *Seeing Like a State* (Yale University Press, 1998), JSTOR, https://doi.org/10.2307/j.ctvxkn7ds.

priorities, and institutional incentives, which in turn shape future data production[69]. Populations that generate more data attract greater regulatory attention, funding, and infrastructural investment, enhancing their visibility furthermore.

Data-driven public health governance privileges what is measurable, traceable, and scalable, while treating absences as technical gaps rather than political problems[70]. Within ABDM, digitally integrated healthcare settings are more likely to be represented in dashboards, performance indicators, and policy analytics, solidifying their centrality within decision-making processes. Data visibility thus becomes a proxy for policy relevance, shaping how public health priorities are identified and addressed.

Conversely, data-poor populations experience reduced institutional recognition and diminished policy attention. Limited data production constrains their inclusion in risk profiling and resource planning, exacerbating marginalisation over time. These feedback loops stabilise inequality by governing through what infrastructures can see and process, making exclusion an infrastructural limitation rather than an outcome of regulatory and institutional design choices[71].

The case of ABDM highlights how regulatory health data infrastructures can reproduce infrastructural inequality even when framed as inclusive, consent-based, and efficiency-enhancing. Uneven legibility becomes a central mechanism through which these infrastructures govern populations differentially, privileging those capable of generating interoperable data while marginalising those whose health experiences fall outside dominant data standards.

Discussion

In this article, we define infrastructural inequalities as durable forms of discrimination, exclusion, and vulnerabilization produced through the material, technical, legal, and organisational arrangements of infrastructures. We investigate how these inequalities are embedded in regulatory data infrastructures – contemporary forms of governance infrastructures that rely on data-driven technologies – grounding our analysis in two case studies. The two cases show that RDIs generate infrastructural inequalities through different 'legibility regimes', with distinct rights and accountability consequences.

---

[69] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (The MIT Press, 1999), https://doi.org/10.7551/mitpress/6352.001.0001; Sheila Jasanoff, 'Virtual, Visible, and Actionable: Data Assemblages and the Sightlines of Justice', *Big Data & Society* 4, no. 2 (2017): 205395171772447, https://doi.org/10.1177/2053951717724477.

[70] Samuel Stehle and Rob Kitchin, 'Real-Time and Archival Data Visualisation Techniques in City Dashboards', *International Journal of Geographical Information Science* 34, no. 2 (2020): 344–66, https://doi.org/10.1080/13658816.2019.1594823.

[71] Bowker and Star, *Sorting Things Out*.

In the EU case, the dominant mode is hyper-legibility: interoperability and biometric matching enable cross-domain data re-use and amplify the durability of suspicion. Key infrastructural features include interconnected large-scale databases, probabilistic matching, and multi-agency access across jurisdictions. The central implications are heightened exposure to surveillance and error propagation, coupled with diffuse accountability that makes explanation, correction, and redress difficult. In the India case, the dominant mode is uneven legibility: enrolment logics, standardisation, and platform-mediated access make participation in care contingent on meeting data-intensive requirements. Key infrastructural features include identifiers, interoperability standards, and consent/enrolment interfaces. The central implications are exclusion and data poverty for those least able to produce 'recognised' data traces, alongside constrained choice where digital pathways become de facto gateways to essential services.

Across both cases, inequalities are produced through recurring infrastructure-level mechanisms, even where specific actors and policy rationales differ. In the EU case, three mechanisms are particularly salient: (1) scope creep via interoperability, as cross-system connectivity weakens practical purpose boundaries and expands consequential uses; (2) probabilistic matching and error, where misclassification becomes structurally possible and its harms scale through interconnection; and (3) constrained redress under diffuse accountability, as responsibilities are distributed across systems, agencies, and jurisdictions, making suspicion 'sticky' over time. In the India case, the dominant mechanisms differ: (1) data legibility requirements and the 'ideal' data subject, as technical standards and enrolment logics presuppose stable identity, documentation, connectivity, and administrative capacity; (2) limited opt-out in practice, where formal consent becomes structurally thin when participation is tied to service access; and (3) data poverty and uneven legibility, where fragmented documentation and intermittent access produce discontinuous data traces that reduce both inclusion and benefit while heightening exposure to exclusion.

Beyond merely enabling mobility and healthcare delivery, these infrastructures actively shape how (and whose) needs are rendered legible to the state, thereby influencing the distribution of regulatory attention and resource allocation[72]. This shift reframes citizenship, mobility, and welfare as conditional upon successful data legibility, transforming infrastructural compliance into a prerequisite for political and social belonging. The application of these governance

[72] Bowker and Star, *Sorting Things Out*; Jasanoff, 'Virtual, Visible, and Actionable'.

technologies first with marginalized populations, like migrants and asylum seekers, and then with the general population, assesses a structural (and institutional) form of exclusion in itself[73].

Regulatory data infrastructures must be understood as infrastructures in precisely this sense: they distribute resources, shape life chances, and allocate vulnerability through standards, interfaces, interoperability regimes, enrolment requirements, storage architectures, and accountability affordances. Algorithmic racism, data poverty, digital exclusion, technological redlining, and data colonialism are therefore not separate phenomena but analytically connected expressions of infrastructural inequality within datafied governance. From a Critical Data Studies perspective, infrastructural inequalities do not arise only from the ways infrastructures configure what can be seen, counted, linked, and acted upon, but also from how these configurations intersect with existing social hierarchies[74].

Conceptually, 'infrastructural inequalities' specifies both where inequality is produced and why it is durable. It shifts attention from discrete decision points (common in algorithmic bias accounts), from access or skills (typical in digital exclusion framings), and from macro-histories of extraction and dependency (often foregrounded in data colonialism) to the infrastructural arrangements that stabilise inequality across contexts: standards that encode an 'ideal' data subject; interoperability that enables scope creep and recursivity; mandatory gateways that render voluntariness tenuous; and accountability fragmentation that constrains contestation and repair. The comparison shows that these mechanisms can produce inequality through over-visibility (hyper-legibility) and under-/partial visibility (uneven legibility), but in both cases through the same underlying distribution of legibility, risk, and remedy.

These developments complicate dominant regulatory frameworks for data protection, particularly the principle of purpose limitation, according to which personal data should be collected for specific and restricted objectives. RDIs operate through continual data accumulation and interoperability across databases and systems, making it increasingly difficult to maintain clear boundaries around purpose. Health records, migration databases, and law enforcement systems are progressively interconnected, enabling new uses of data that were not envisioned at the moment of collection[75]. This dynamic of scope creep - such as the repurposing of migration data for criminal identification - raises fundamental questions about

[73] Ozkul and Godin, *Exclusion by Design*.
[74] Francesca Pilo', 'Negotiating Networked Infrastructural Inequalities: Governance, Electricity Access, and Space in Rio de Janeiro', *Environment and Planning C: Politics and Space* 39, no. 2 (2021): 265–81, https://doi.org/10.1177/2399654419861110.
[75] Amelung, ''Crimmigration Control' across Borders'.

whether existing data protection legislation remains adequate in environments characterized by large-scale integration and automated cross-checking.

Crucially, accountability becomes diffuse and redress is hindered. In the European case, data controllers are distributed across Member States, while supranational agencies administer core technical infrastructures as the 'data processor'. This institutional fragmentation complicates legal redress and obscures responsibility when errors occur. Individuals confronted with false positives or exclusionary classifications often face opaque bureaucratic pathways, with no single authority clearly accountable for correcting records or explaining decisions. Consent, meanwhile, becomes largely formalistic. When access to healthcare or cross-border mobility depends on biometric registration or digital identity systems, refusal carries severe consequences, undermining meaningful choice.

RDIs have long temporal effects. Algorithmic classifications and administrative risk scores are rarely ephemeral. As Eubanks[76] emphasizes, disadvantageous data points tend to be 'sticky,' persisting across systems and over time, and shaping future encounters with the state. Interoperability amplifies this durability: once recorded, a contested datapoint may circulate across sectors and jurisdictions, influencing decisions far removed from its original context. Infrastructural inequalities thus accumulate over generations, sedimenting past judgments into durable constraints on life opportunities.

The recursive character of datafication further intensifies this dynamic. As RDIs expand, the collection of ever more data becomes normalized and framed as necessary for improving accuracy, preventing fraud, or enhancing security. Yet our findings invite closer scrutiny of claims about necessity and proportionality. Additional data does not automatically produce fairer outcomes; it may instead entrench earlier errors[77] or extend surveillance to new domains[78]. Narratives of efficiency and security play a crucial legitimizing role here, presenting infrastructural expansion as technical optimization rather than political decisions[79].

This is particularly evident in contemporary border governance, increasingly reorganised as a technical infrastructure. Interoperable biometric systems reconfigure migration management into an exercise in data matching and risk scoring, reinforcing the dynamic of crimmigration,

[76] Eubanks, *Automating Inequality*, First Picador edition (Picador St. Martin's Press, 2019).
[77] D'Ignazio and Klein, *Data Feminism* (The MIT Press, 2020).
[78] Silvia Masiero, 'Digital Identity as Platform-Mediated Surveillance', *Big Data & Society* 10, no. 1 (2023): 20539517221135176, https://doi.org/10.1177/20539517221135176.
[79] Pilo', 'Negotiating Networked Infrastructural Inequalities'.

i.e. the entanglement of criminal investigation and immigration control [80] . Through interoperable databases, the figure of the 'crimmigrant other' is not merely rhetorically constructed but infrastructurally produced, as risk categories, alerts, and watchlists circulate across databases. These processes expose certain populations – often the already marginalized, like migrants and asylum seekers - to disproportionate risks of false positives, while making contestation difficult once classifications propagate across interconnected systems.

At the same time, infrastructural inequalities also manifest through invisibility. Data poverty is likely to generate false negatives that render individuals unrecognizable to administrative systems, hindering access to welfare or care. In health infrastructures, standardized data formats and algorithmic assessment contribute to the construction of 'ideal data subjects', which do not reflect the complex reality of individuals and their lived experience[81]. Those who do not conform to these expectations are more likely to fall outside the system's field of vision. The prioritization of data legibility over citizen needs becomes itself a mechanism of marginalization.

Taken together, these findings suggest that RDIs should be understood not merely as tools that operate within unequal societies, but as infrastructures that actively participate in the production of inequality. By foregrounding infrastructural inequalities, we aim to shift analytical and regulatory attention from isolated instances of algorithmic bias toward the broader architectures through which data-driven governance is assembled, justified, and normalized. This perspective highlights the need for regulatory approaches that go beyond individual rights and technical audits, and instead interrogate interoperability regimes, institutional accountability, standards of necessity and proportionality, and the political imaginaries that underpin contemporary datafication.

If inequality is infrastructural, governance must target infrastructural levers rather than only downstream outcomes. This includes treating interoperability as a high-stakes design choice that requires enforceable purpose boundaries, minimisation, and monitoring of cross-domain re-use; auditing matching practices for error propagation across connected systems; and building accessible, single-entry pathways for explanation, correction, and remedy that do not require navigating multiple agencies. Where RDIs operate as gateways to essential services,

[80] Amelung, ''Crimmigration Control' across Borders'.
[81] Goriunova, Olga. *Ideal Subjects: The Abstract People of AI*. Vol. 76. U of Minnesota Press, 2025.

safeguards should prioritise non-digital alternatives and prevent ‘consent’ from becoming a formal label attached to structurally mandatory participation.

This paper foregrounds infrastructure-level mechanisms using documentary and secondary sources; it therefore captures design rationales, institutional arrangements, and system logics more than lived experience at the point of encounter. Future research should add ethnographic and participatory work to examine how people navigate, resist, or reconfigure these infrastructures in practice, and comparative extensions across additional geographies and technology families to test the portability of the mechanism set and refine the concept of infrastructural inequalities.

Conclusion

This article argues that regulatory data infrastructures (RDIs) produce inequality not primarily through isolated biased outputs, but through infrastructural arrangements that distribute legibility, risk, and remedy across populations. Comparing the EU’s interoperable biometric border infrastructures with India’s ABDM health data infrastructure shows two distinct but related modes of infrastructural inequality—hyper-legibility and uneven legibility—each sustained by recurring mechanisms.

The analysis is limited by its focus on two cases and its primary reliance on documentary and secondary sources, which illuminate system design, regulatory narratives, and institutional arrangements more than lived experience at points of encounter. It also foregrounds the infrastructural layer rather than adjacent political-economic dimensions such as procurement practices, cost models, vendor lock-in, and the role of commercial intermediaries—factors that may further shape how RDIs scale and entrench inequality.

Future research should extend this comparative approach to additional geographies and technology families and combine mechanism tracing with ethnographic and participatory methods to examine how infrastructures are navigated, contested, and repaired in practice. In parallel, infrastructure-focused forms of scrutiny are needed, including audits of interoperability arrangements (and their downstream error propagation), and empirical study of redress pathways across fragmented institutional settings.

The central implication is that governing RDIs requires moving beyond system-by-system compliance and outcome-only fairness checks toward assessing the infrastructural conditions that make inequality durable. Regulators and institutions should treat interoperability as a high-

stakes political and technical choice subject to robust impact assessment; ensure meaningful opt-out or non-digital alternatives where infrastructures become gateways to essential services; strengthen purpose limitation and data minimisation to prevent recursivity and scope creep; and build clear, accessible, single-entry routes for explanation, correction, and remedy across interconnected systems. In short, protecting rights in datafied societies increasingly depends on governing the infrastructures that render people visible, actionable, and contestable in the first place.

References

Amelung, Nina. '"Crimmigration Control" across Borders: The Convergence of Migration and Crime Control through Transnational Biometric Databases'. *Historical Social Research / Historische Sozialforschung* 46, no. 3 (2021): 151–77.

Beer, Dr David. 'Social Network(Ing) Sites…revisiting the Story so Far: A Response to Danah Boyd & Nicole Ellison'. *Journal of Computer-Mediated Communication* 13, no. 2 (2008): 516–29. https://doi.org/10.1111/j.1083-6101.2008.00408.x.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. The MIT Press, 1999. https://doi.org/10.7551/mitpress/6352.001.0001.

Boyd, Danah, and Kate Crawford. 'CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon'. *Information, Communication & Society* 15, no. 5 (2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Buolamwini, Joy, and Timnit Gebru. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. n.d.

Couldry, Nick, and Ulises A. Mejias. 'Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject'. *Television & New Media* 20, no. 4 (2019): 336–49. https://doi.org/10.1177/1527476418796632.

Datta, Ayona. 'The Digital Turn in Postcolonial Urbanism: Smart Citizenship in the Making of India's 100 Smart Cities'. *Transactions of the Institute of British Geographers* 43, no. 3 (2018): 405–19. https://doi.org/10.1111/tran.12225.

De Hert, Paul, and Georgios Bouchagiar. 'Visual and Biometric Surveillance in the EU. Saying "No" to Mass Surveillance Practices?12'. *Information Polity* 27, no. 2 (2022): 193–217. https://doi.org/10.3233/IP-211525.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas. The MIT Press, 2020.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. The MIT Press, 2020.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Picador edition. Picador St. Martin's Press, 2019.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Picador edition. Picador St. Martin's Press, 2019.

Fox, Nick J., Pam Alldred, and Adrian Franklin. *The More-than-Human Micropolitics of the Research Assemblage*. 1st edn. Routledge, 2024. https://doi.org/10.4324/9781003262619-28.

Franko, Katja. *The Crimmigrant Other: Migration and Penal Power*. 1st edn. Routledge, 2019. https://doi.org/10.4324/9781351001441.

Garrett, Camryn M., and Rochelle Altman. 'Vulnerabilized: Revisiting the Language of the Vulnerable Populations Framework'. *American Journal of Public Health* 114, no. 2 (2024): 177–79. https://doi.org/10.2105/AJPH.2023.307532.

Grealy, Liam, Andrew Brooks, Astrid Lorange, Christen Cornell, and Tess Lea. *Introduction: Tending a Social Infrastructure*. n.d.

Haggerty, Kevin D., and Richard V. Ericson. 'The Surveillant Assemblage'. *The British Journal of Sociology* 51, no. 4 (2000): 605–22. https://doi.org/10.1080/00071310020015280.

Haitsma, Lucas Michael, and Maarten Bouwmeester. 'Learning from Control Deficits in the Childcare Benefits Scandal: A Plea for Multi-Level Analysis in Law and Policy Research'. *Recht Der Werkelijkheid* 44, no. 3 (2023): 57–68. https://doi.org/10.5553/RdW/138064242023044003004.

Hepp, Andreas, Juliane Jarke, and Leif Kramp. 'New Perspectives in Critical Data Studies: The Ambivalences of Data Power—An Introduction'. In *New Perspectives in Critical Data Studies*, edited by Andreas Hepp, Juliane Jarke, and Leif Kramp. Transforming Communications – Studies in Cross-Media Research. Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-96180-0_1.

Iazzolino, Gianluca. 'Infrastructure of Compassionate Repression: Making Sense of Biometrics in Kakuma Refugee Camp'. *Information Technology for Development* 27, no. 1 (2021): 111–28. https://doi.org/10.1080/02681102.2020.1816881.

Ibrahim, Hussein, Xiaoxuan Liu, Nevine Zariffa, Andrew D Morris, and Alastair K Denniston. 'Health Data Poverty: An Assailable Barrier to Equitable Digital Health Care'. *The Lancet Digital Health* 3, no. 4 (2021): e260–65. https://doi.org/10.1016/S2589-7500(20)30317-4.

Iliadis, Andrew, and Federica Russo. 'Critical Data Studies: An Introduction'. *Big Data & Society* 3, no. 2 (2016): 2053951716674238. https://doi.org/10.1177/2053951716674238.

Iliadis, Andrew, and Federica Russo. 'Critical Data Studies: An Introduction'. *Big Data & Society* 3, no. 2 (2016): 2053951716674238. https://doi.org/10.1177/2053951716674238.

Jain, Nisha B, and Samiran Nundy. 'Electronic Health Records in India: Challenges and Promises'. *Journal of Medical Evidence* 2, no. 3 (2021): 278–79. https://doi.org/10.4103/JME.JME_94_21.

Jasanoff, Sheila. 'Virtual, Visible, and Actionable: Data Assemblages and the Sightlines of Justice'. *Big Data & Society* 4, no. 2 (2017): 205395171772447. https://doi.org/10.1177/2053951717724477.

Jeandesboz, Julien. 'Smartening Border Security in the European Union: An Associational Inquiry'. *Security Dialogue* 47, no. 4 (2016): 292–309.

Jena, Shubharanjan, Venkatarao Epari, and Krushna Chandra Sahoo. 'Integration of National Cancer Registry Program with Ayushman Bharat Digital Mission in India: A Necessity or an Option'. *Public Health in Practice* 3 (June 2022): 100263. https://doi.org/10.1016/j.puhip.2022.100263.

Kitchin, Rob. *Critical Data Studies: An A to Z Guide to Concepts and Methods*. First edition. Polity press, 2025.

Kitchin, Rob. 'Thinking Critically about and Researching Algorithms'. *Information, Communication & Society* 20, no. 1 (2017): 14–29. https://doi.org/10.1080/1369118X.2016.1154087.

Law, John. 'Actor Network Theory and Material Semiotics'. 25 April 2007. http://www.heterogeneities. net/publications/Law2007ANTandMaterialSemiotics.pdf.

Leese, Matthias. ‘Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU’. *Geopolitics* 27, no. 1 (2022): 113–33. https://doi.org/10.1080/14650045.2020.1830764.

Leurs, Koen, and Tamara Shepherd. ‘Datafication & Discrimination’. In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es. Studying Culture through Data. Amsterdam University Press, 2017. https://www.jstor.org/stable/j.ctt1v2xsqn.20.

Masiero, Silvia. ‘Digital Identity as Platform-Mediated Surveillance’. *Big Data & Society* 10, no. 1 (2023): 20539517221135176. https://doi.org/10.1177/20539517221135176.

McLoughlin, Ian P., Ian P./Garrety McLoughlin Kari, and Rob Wilson. *The Digitalization of Healthcare*. 2017.

Milan, Stefania. ‘Afterword: From Number Politics to Infrastructure Politics: Notes on Context and Methods’. *The Cambridge Journal of Anthropology* 42, no. 1 (2024): 118–26. https://doi.org/10.3167/cja.2024.420108.

Milan, Stefania. ‘Techno-Solutionism and the Standard Human in the Making of the COVID-19 Pandemic’. *Big Data & Society* 7, no. 2 (2020): 2053951720966781. https://doi.org/10.1177/2053951720966781.

Milan, Stefania, and Emiliano Treré. ‘Big Data from the South(s): Beyond Data Universalism’. *Television & New Media* 20, no. 4 (2019): 319–35. https://doi.org/10.1177/1527476419837739.

Milan, Stefania, and Emiliano Treré. ‘The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins’. *Social Media + Society* 6, no. 3 (2020): 2056305120948233. https://doi.org/10.1177/2056305120948233.

Morel, Victor, Cristiana Santos, Yvonne Lintao, and Soheil Human. ‘Your Consent Is Worth 75 Euros A Year - Measurement and Lawfulness of Cookie Paywalls’. *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, 7 November 2022, 213–18. https://doi.org/10.1145/3559613.3563205.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2020. https://doi.org/10.18574/nyu/9781479833641.001.0001.

Ozkul, Derya, and Marie Godin. *Exclusion by Design: Digital Identification and the Hostile Environment for Migrants*. ESRC Digital Good Network, 2025. https://doi.org/10.31273/9781911675211.

Pandey, Bhartendu, Christa Brelsford, and Karen C. Seto. ‘Rising Infrastructure Inequalities Accompany Urbanization and Economic Development’. *Nature Communications* 16, no. 1 (2025): 1193. https://doi.org/10.1038/s41467-025-56539-w.

Pilo’, Francesca. ‘Negotiating Networked Infrastructural Inequalities: Governance, Electricity Access, and Space in Rio de Janeiro’. *Environment and Planning C: Politics and Space* 39, no. 2 (2021): 265–81. https://doi.org/10.1177/2399654419861110.

Richterich, Annika. *The Big Data Agenda: Data Ethics and Critical Data Studies*. University of Westminster Press, 2018. https://doi.org/10.16997/book14.

Ruppert, Evelyn. ‘Population Objects: Interpassive Subjects’. *Sociology* 45, no. 2 (2011): 218–33. https://doi.org/10.1177/0038038510394027.

Scott, James C. *Seeing Like a State*. Yale University Press, 1998. JSTOR. https://doi.org/10.2307/j.ctvxkn7ds.

Sharma, R. S., Aishwarya Rohatgi, Sandeep Jain, and Dilip Singh. ‘The Ayushman Bharat Digital Mission (ABDM): Making of India’s Digital Health Story’. *CSI Transactions on ICT* 11, no. 1 (2023): 3–9. https://doi.org/10.1007/s40012-023-00375-0.

Smith, Gavin Jd. ‘Data Doxa: The Affective Consequences of Data Practices’. *Big Data & Society* 5, no. 1 (2018): 2053951717751551. https://doi.org/10.1177/2053951717751551.

Stehle, Samuel, and Rob Kitchin. ‘Real-Time and Archival Data Visualisation Techniques in City Dashboards’. *International Journal of Geographical Information Science* 34, no. 2 (2020): 344–66. https://doi.org/10.1080/13658816.2019.1594823.

Tan, Kynan. *Polymorphism (Data Centre Simulation)*. n.d.

Trauttmansdorff, Paul. ‘The Fabrication of a Necessary Policy Fiction: The Interoperability “Solution” for Biometric Borders’. *Critical Policy Studies* 17, no. 3 (2023): 428–46. https://doi.org/10.1080/19460171.2022.2147851.

Van Dijk, Teun A. ‘Structures of Discourse and Structures of Power’. *Annals of the International Communication Association* 12, no. 1 (1989): 18–59. https://doi.org/10.1080/23808985.1989.11678711.

Wachter, Sandra, and Brent Mittelstadt. ‘A Right to Reasonable Inferences’. *Columbia Business Law Review*, 1 May 2019, 494-620 Pages. 494–620 Pages. https://doi.org/10.7916/CBLR.V2019I2.3424.

Yin, Robert K. *Case Study Research: Design and Methods*. Vol. 5. SAGE Publications Ltd, 2009.

Yücedağ, Nafiye, Elif Beyza Akkanat Öztürk, and Şehriban İpek Aşıkoğlu. ‘Pay or Consent Models in Europe: Already Outdated or an Overlooked Crisis in Freely Given Consent?’ *Technology and Regulation* 2025 (July 2025): 336–54. https://doi.org/10.71265/8jjzhc21.

Zarsky, Tal. *Incompatible: The GDPR in the Age of Big Data*. Sect. 2. 47, no. 4 (2017). https://ssrn.com/abstract=3022646.