Openness

Abstract
Data openness has emerged as a crucial requirement for fostering innovation, transparency, and collaborative progress across sectors in an increasingly digital world. This report examines the multifaceted landscape of openness, exploring its fundamental principles, challenges, and potential solutions. We analyze the organizational, regulatory, ethical, and technical hurdles that organizations face when implementing open data initiatives while presenting practical solutions to overcome these obstacles. The discussion encompasses both theoretical frameworks and real-world applications, providing insights for stakeholders seeking to embrace and implement data openness strategies effectively.

1. Introduction

The concept of 'open' has proliferated to all sectors, from public organizations, non-profits, and private organizations. 'Open' is not only a description: it has become an organizational driver. Business models are designed around 'open' innovation; software is developed in the 'open' through collaboration; scientific knowledge is provided as 'open' access; spaces are designed to be 'open' to encourage conversation; communities are forming around the world to gather 'open' intelligence to make an inventory of and verify war actions and abuses.

Data is a fundamental element in facilitating this 'openness', but there are many views on what data are. Before defining open data, we must first explore the concept of data.

1.1 On Data

The predominant view considers data as an objective representation of an pre-existing reality. Such view sees data as 'raw materials' that result from the abstraction of that reality thorugh numbers, categories, or measures (Kitchin 2014). Because of such objective view, data is often thought to be neutral and help reduce bias that is located in the social world and particularly humans. Predictive policing systems, for example, are portrayed as reducing bias (Ferguson 2020). Data also have political meaning because what can be measured can be controlled (Scott state).

This view of data is contested by many researchers (e.g., Kitchin, Monteiro, Pargmiaggnia, Aaltonen). Data is often embedded in processes of measure and calculation that are far from neutral (Atalas of AI). As Kitchin (2014) argues, "data are in fact framed technically, economically, ethically, temporally, spatially and philosophically". AI systems, for example, are dependent on exploitative processes that aim to 'moderate its contents' (Wired/Guardian Kenya article).

Not all data are the same. There are many descriptive categories of data. Some data are quantitative and refer to numerical records such as ratios. Some data are qualitative such as video statements about an observation. Some data are structured and organized in categories that hold meaning. For example, in Decide Madrid, one of the most downloaded citizen participation portal, participation is coded by a number of categories (e.g., proposal description, number of votes, number of comments, etc.). According to Zikopoulos et al. (2012 from Kitchin), semi- and unstructured are growing faster than structured data, requiring new ways to store and access such data such as noSQL (Kitchin 2014). Semi and unstructured data would, for example, are either loosely structured or have no clear way to identify a structure. An uploaded video on youtube without metadata, for example, is semistructured, with just the title helping with classification.

Data is also generated (see Kitchin 2014). Observable events need to be measured to be transformed into data. This kind of data is 'captured', sometimes through a protocol. When we take a photograph with our phones, the device embeds a number of algorithms that measure light sensors to regulate the camera's aperture to improve the quality of our photos. Other data are called 'exhaust' data because they are created through our use of systems. For example, cars in the future may measure certain driving data such as speed or driving behaviour which are not essential to driving but may become important if there is an accident. Some data are transient, meaning that they are discarded because not seen as valuable. Other data is derived and not measured directly. In the example on the participation platform, we can derive through machine learning that one of the main topics of citizen proposals are ecological (Miguel XXX).

Data matters in this informational age because it is a fundamental obbject in decision-making. A linear view of the value of data sees it as a building block towards higher forms. Data need to be processed and turned into information, which after analysis can be turned into knowledge. Big data follows this linear porcess. The value of big data – data that is characterised by the four Vs (high volume, high velocity, high variance, high V...) – depends on the availability of computing power, large storage facilities, algorithms, and immense amounts of data. One item in that big data is irrelevant, but taken togethre and processed can derive information (e.g., most people in Madrid write proposals about ecological issues), which can then be turned into knowledge (e.g., let's create a community that can create a unifying proposal that synthesises everyone's demands). This pyramidal view of data has a lot of purchase beyond big data. Evidence-based management and best-practices also implicitly assume such pyramidal knowledge. To know what to do, we need to gather data to turn it into evidence (e.g., information) that can then be reapplied to multiple future situations.

Not all data follows this value chain. Some data only local and contextual meaning. Such data are called 'thick' because they have thick social meaning attributed to it (Smith 2018). Smith contrasts big data with thick data: "The kinds of data gathered, the methods used, how it gets interpreted, what gets overlooked, the context in which it is generated, and by whom, and what to do as a result, are all choices that shape the facts of a matter. For experts building Big Data city platforms, one sensor in one square is simply a data point. On the other side of that point, however, are residents connecting that data to life in all its richness in their square. Anthropologist Clifford Geertz argued many years ago that situations can only be made meaningful through 'thick description'. Noise data in Plaza del Sol was becoming thick with social meaning. Collective data gathering proved more potent than decibel levels alone: it was simultaneously mobilising people into changing the situation. Noise was no longer an individual problem, but a collective issue. And it was no longer just noise. The data project arose through face-to-face meetings in a physical workshop space. Importantly, this meant that neighbours got to know one another better, and had reasons for discussing life in the square when they bumped into one another."

1.2 Why open data?

A recent trend has seen open data gaining relevance. There are five main inter-related reasons why open data has been explored. First, with increasing computing power, the accumulation of data gained much more value contributing to the elaboration of derived data. Open datasets can be linked together, scaling up the value of the datasets (Lindman 2014). Second, capabilities to store and manipualte data has increased: people have machines that can access store on the Internet and compute algorithms that necessitated significant computing power before. Moreover, computing knowledge has diffused beyond organizations, allowing a wider number of stakeholders to generate, capture, and derive meaning from data. Fourth, data has become an inescapable way to conduct politics. As Powell (XXXX) recounts, many civic groups have learned to use data to push political agendas. Fifth, there are importatn ethical and normative reasons for dta to be open: data should be free because it influences people. Opaque systems with incorrect data may have inadvertant or harmful effects on people (see Australia's robodebt system). The debates on making data open, thus, show us that it is a site of struggle and not a raw materials describing an outside reality and can become a source of emancipation (Ignazio and klein).

Pollock (2006, also from Kitchin 2014) suggests that 'data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike'. This simple definition implies that the term 'open data' holds different meaning to people. For Pollock, for example, the requirement is that it has a certain license (i.e., give attribution and/or share-alike). But there are many other ways to undertand open data. For example, Spanish hacker movements tend to refer to the notion of Free culture as part of their process of working with open data (Corsin Jimenez and Estalella 2014). Instead of the license, a state, 'opening' is an action allowed through collective expressions of Free culture that speak of autonomy and sustainability.

The concept of (data) openness not only holds different meaning to different people and groups, but it has also evolved significantly in recent years, transforming from a niche interest into a fundamental principle driving innovation and transparency. across public and private sectors. A consensual definition would refer toopen data as a practice of making data freely available, accessible, and usable while ensuring it remains machine-readable and interoperable. Beyond this definition, views on what, how, and why we should open may diverge. This divergence results from the very nature of openness: data can be reproduced, taken out of its context, or infringe privacy expectations (we discuss challenges in section XXX). Not all would agree, for example, that all should be open by default (open manifesto XXX). Open data creates so much debate precisely because digital data is given meaning in the way by the way we collect it, measure it, describe it, link it, and so on.

1.3 Multiple traditions

One of the reasons why there are different views on what ‘open’ means or should mean is because the evolution of data openness is traced back through several key developments, traditions, and communities of practice. These include the open source software movement, the push for government transparency, and the rise of open science initiatives. These movements have demonstrated that when data is made accessible and reusable, it can generate significant value for society. The importance of open data initiatives has been particularly evident in government sectors (Davies, 2010), where open data platforms have demonstrated significant potential for public sector reform and democratic engagement. However, successful implementation requires careful consideration of benefits and potential barriers (Janssen et al., 2012).

1.4 Potential of open data

Open data has gained prominence for its potential, because of the scale it can achieve or as a legitimate process to create ethical ‘thick’ descriptions of issues and solutions. Hossain et al.’s (2016) review the literature and classify the benefits of open data according to organizations and citizens. Organizations recognize the potential of shared data to catalyze innovation, improve decision-making, and foster collaboration across boundaries. For private firms, open data can enable better customer interaction and access to dispersed information from a single repository. This can lead to improved decision-making and enhanced services through visualization and mash-ups. In the context of government, open data can increase transparency by disclosing datasets related to government spending and statistics, allowing citizens to hold institutions accountable and fight corruption. Open data can also empower citizens, policymakers, social analysts, and advocacy groups by providing access to primary data for better policy-making.

The benefits of open data extend beyond governments and private firms to society as a whole. By making public facilities' datasets available, citizens can enhance the quality of their lives through improved services such as healthcare and transportation. Open data also enables civic engagement in areas like policing, law enforcement, and social disorder monitoring. Open data also offer a way to contest unethical systems. Opaque systems with close data can take unfair decisions without people knowing. Because we know what it is, we can trace algorithmic actions back to the kind of data we have, data is performed, so we need to know how it is measured. Furthermore, value creation through open data can generate wealth by creating new jobs and innovative services, while government agencies can save costs on report rendering and application development. In academic research, open data can speed up research, reduce redundant work, and promote collaboration and reproducibility of results.

1.5 Values of open data

Open data is associated to many values including transparency (Poirier 2024), empowerment, economic growth through innovation, and social value. Transparency is often an important consideration, as the disclosure of datasets related to government activities fosters accountability and enhances trust among stakeholders. This value is complemented by the notion of empowerment, whereby access to information previously inaccessible enables individuals to effectuate their own interests and participate in decision-making processes. Furthermore, open data has been recognized as a catalyst for economic growth through innovation, as it facilitates entrepreneurship and job creation by enabling third-party developers to craft novel applications and services. Social value is another salient aspect of the concept, as open data can lead to improved public service delivery, enhanced civic engagement, and better-informed citizens, thereby contributing to the realization of a more just and equitable society. Others have stronger hopes with open data and call on data or informational justice (Ignazio and Klein, Johnson 2014).

2. Challenges

2.1 Organizational Challenges

The implementation of data openness initiatives faces significant organizational hurdles that can impede progress and adoption. Research on open government data programs has demonstrated that successful implementation requires an ecosystem approach (Dawes et al., 2016), considering multiple stakeholders and their interactions. Studies of platform ecosystems in Latin American cities have shown that effective governance mechanisms are crucial for cultivating sustainable open data initiatives (Bonina & Eaton, 2020). There is often a strong initial cultural resistance to open practices. Research has suggested that many within organizations feel that they are not prepared for data sharing even when it is technically possible (Runeson and Olsson 2020). For open data to work, an organization must have standard processes to collect and share data. It can be the case that different layers (e.g., the strategic and operational levels) in the organization feel they are not aligned with the consequences and requirements for setting up open data processes.

Cultural resistance manifested strongly within the healthcare environment. Medical staff expressed concerns about potential misinterpretation of their clinical decisions, while senior physicians demonstrated hesitancy in adopting new data-sharing protocols. The variation in medical practices across countries led to persistent disagreements about data standardization. The challenges of data governance and standardization across different organizational cultures align with findings from recent studies on open government data quality (Alexopoulos et al., 2023).

It is not a matter of freeing the data that counts, but how it is made available (Jetzek 2016). There are many challenges that organizations must face to make open data useful. Zuiderwijk et al. (2012) make a comprehensive list of obstacles, many of which should be solved or are the responsibility of organizations. They identify eight main themes: how open data is made available or can be accessed (e.g., are datasets updated, how to deal with duplicated datasets, etc.); the ability to discover the data sets in the first place (e.g., is there advanced search capabilities? Dataset storage is fragmented, etc.); the usability of the open data (is it trustworthy?); how the data is contextualized or communicated (e.g., data are not visualised or there is a lack of knowledge in interpreting the data); data quality (e.g., inaccurate data or data whose quality is difficult to ascertain); how easy it is to link datasets together (e.g., is the dataset interoperable with others? Lack of tools to link data); comparability (e.g., multiple and conflicting definitions of data); lack of metadata; interactions with data providers; and finally, difficulties in opening or uploading data (e.g., threats of privacy violation by publishing data).

The hospital case showcased many of these issues. One of the main challenges was to determine data ownership and control. Each hospital maintained different data ownership structures, while department heads showed considerable reluctance to share their research data. Creating a unified governance framework across different organizational cultures proved particularly challenging. A departmental approach to data sharing created different procedures which had the potential to affect data quality, discoverability, and availability, significantly reducing the potential benefits from data openness.

Open data, thus, requires the organizational deployment of proportionate resources. These go beyond, but include deploying a a strategy and identification of organizational data needs to define a coherent, long-term programme. This will inform the elaboration of a 'soft data infrastructure' that determines the licenses used, but potential costs and value evaluations of opening data up. These strategic considerations lead to more operational concerns such as how or whether such data can be used internally, how they should improve services, or how they provide value to the public and how citizens should become engaged (Jetzek 2016).

The resources required to overcome all these challenges and define a coherent strategy are a major hurdle for smaller organizations. Indeed, in our study, smaller hospitals within the consortium struggled with implementation costs, while IT departments found themselves understaffed for the increased workload. Emergency care resources often took priority over data initiative needs, and training costs for new systems consistently exceeded initial budgets.

2.2 Regulatory/Ethical Challenges

The regulatory and ethical landscape surrounding data openness presented complex challenges that required careful navigation. Wessels et al. (2014) consider a number of legal and ethical challenges. One such challenge is etermining when data should be opened. There are important legal requirements for organizations limiting or guiding what and how data can be opened. For example, intellectual property may pose an upper limit to what data can be opened for an organization, while in the US, much government work is considered open and not copyrighted (though there may be prevalent organizational issues). Citizens opening data or designs up may very well give away their rights to copyright. Scientific data still remains the captured by institutional processes of ranking and evaluation in the hands of lalrge publishing companies (Postill XXX). Scientific publishing process place embargoes on when data can be made available, where data can be uploade (e.g., some forbid storing articles in 3rd-party or institutional repositories to complicate access).

The ethical considerations in data openness extend beyond legal compliance, requiring structured frameworks for ethical decision-making (Open Data Institute, 2018). The healthcare consortium's experience aligns with known GDPR challenges in research contexts (Cagnazzo, 2021; Staunton et al., 2019), particularly regarding secondary data usage (Peloquin et al., 2020).

Privacy concerns stood at the forefront of these challenges within the healthcare consortium. Patient data required different levels of anonymization across countries, and GDPR compliance needed varying approaches in each jurisdiction. For certain data, there are difficult questions regarding whether consent should be opt-in or opt-out (e.g., anonymised software telemetry).

Even though the GDPR was a significant advancement in regulating the use of data and the role of consent, many issues remain that are relevant to opening data up. One such issue is the typical conflict between transparency and openness on one hand, and confidentiality and privacy on the other. Not all data can easily be opened up (e.g., patient data requires additional work). There is a lot of research on this topic, including the creation of 'synthetic data' that can accoutn for privacy and confidetiality issues through data governance policies (Young et al. 2019). Another option is to provide accreditationt to specific people who are authorized to carry out data analysis (Guimarães 2023), leading to semi-open data (Bargh et al. 2016). Semi-data are data that only partially meet the criteria for open data. Regardless, the complexity of existing legal frameworks add another layer of difficulty to open data initiatives.

In our case, ethical considerations emerged throughout the project's implementation. Concerns arose about insurance companies potentially using data to discriminate, while pediatric data required special protection measures. AI diagnostic tools showed bias in certain demographic groups, raising questions about responsibility for secondary research findings. In other words, difference chains in the process had a different regulatory (and ethical) approach to processing data, which further complicated data handling.

There are also potential conflict between the various values and meaning held by open data (Gonzalez-Zapata and Heeks 2015). For example, openness does not necessarily bring transparency, and these two values may sometimes need to be balanced (Shaikh and Vaast). Openness is also not just an ideal that is attained by only making data open. There are different dimensions that are negotiated between open practitioners including whether openness also includes (inclusive) participation, processes for making a product open, or indeed whether open is compatible with rigour (Curto-Millet and Shaikh 2018). As Hyong (2017) forcefully shows this contrast. On one hand, open data "order[s] data (through capturing, structuring, aggregating, and visualizing) forms part and parcel of ordering society as well as eradicating irrationalities, inefficiencies, and corruption." On the other, "open Data also supports datafication and the algorithmic governance of targeted populations and markets [...] to control which populations are subjected (Deleuze 1992; Lyon 2001)." The ideals of open data, in other words, may be subverted for harmful purposes, an outcome that Gonzalez-Zapata and Heeks (2015) refers to as 'adverse digital incorporation'.

2.3 Technical Challenges

Technical challenges presented significant barriers to implementing effective data openness initiatives. Quality measurement frameworks for open government data have identified multiple dimensions that must be addressed to ensure data usability (Vetrò et al., 2016). These quality concerns are particularly evident in healthcare settings, where standardization across systems remains a significant challenge (Van de Vyvere & Colpaert, 2022).

Data quality and standardization issues emerged immediately, as different hospitals used varying diagnostic coding systems. Metadata standards varied across institutions, and legacy systems in smaller hospitals struggled to handle new data formats. Infrastructure requirements posed substantial challenges throughout the implementation. Imaging data storage exceeded initial capacity estimates, while real-time data processing requirements strained existing systems. Network bandwidth bottlenecks delayed data transfers, and security systems needed continuous upgrades to maintain effectiveness.

Interoperability issues persisted throughout the project's lifetime. Each hospital maintained different IT infrastructure, and medical device data formats varied by manufacturer. Technical expertise proved difficult to secure and maintain, as finding qualified data engineers presented a consistent challenge.

3. Solutions

3.1 Technical Solution Architecture

Modern data openness initiatives require robust technical infrastructure that can handle diverse data types while maintaining security and accessibility. The foundation begins with a data lake architecture implementing delta lake principles for version control and ACID compliance, ensuring data consistency while enabling flexible access patterns.

API management includes well-documented gateways supporting both REST and GraphQL interfaces, with rate limiting, request validation, and OAuth 2.0 authentication. Data standardization employs industry-standard formats and protocols, complemented by ETL pipelines for real-time processing. Security solutions implement end-to-end encryption and zero-trust architecture principles.

3.2 Data Governance Structure

The governance structure requires clear roles and responsibilities implemented through a carefully designed hierarchical structure. Recent studies on open government data quality have emphasized the importance of robust governance structures (Alexopoulos et al., 2023). Successful implementation requires careful consideration of both technical and organizational factors, with governance mechanisms playing a crucial role in platform ecosystem development (Bonina & Eaton, 2020; Dawes et al., 2016). This, in turn, requires deploying resources strategically with a coherent plan. There are a number of elements that we highlight here. Any open data plan should reflect on each of these in turn.

At the highest level, the Data Governance Board provides strategic direction and establishes key policies. This board, comprising representatives from senior management, legal departments, IT divisions, and key business units, sets the overall data strategy and ensures alignment with organizational objectives. Supporting this top-level guidance, the Data Stewardship Council manages operational oversight, implementing governance policies and overseeing day-to-day data management activities. This council includes data stewards from each business unit, technical experts, and compliance officers who collectively manage data quality standards, access controls, and metadata management. The Data Quality Team forms the third pillar of this structure, focusing exclusively on maintaining accuracy and consistency across all data systems and processes.

The compliance framework must comprehensively address various regulatory requirements through an integrated approach. This begins with privacy compliance mapping to major regulations, ensuring that all data handling processes align with relevant legal requirements. Organizations must implement security controls based on established standards, creating a robust foundation for data protection. Regular audits and monitoring procedures ensure ongoing compliance, while automated compliance checking tools provide continuous oversight and early warning of potential issues.

Implementation follows three key phases: foundation-building (establishing governance structures and basic infrastructure), core implementation (deploying API management and security controls), and enhancement (expanding capabilities to meet evolving needs).

Organizations must implement comprehensive *risk management* through systematic assessment and mitigation. This includes evaluating potential threats, developing clear mitigation strategies, and maintaining effective incident response planning. Continuous monitoring ensures risk management strategies remain effective and adapt to changing circumstances.

Quality assurance relies on automated control checks, clear metrics, and monitoring processes to track data quality. Well-defined remediation procedures and regular stakeholder feedback ensure continuous improvement of quality control measures.

This is not a linear process and multiple iterations are expected. Each of these elements tackle organizational, technical, and ethical issues and dilemma, engaging stakeholders to reflect deeply on their process of opening up data.

4. Conclusion

Data openness represents both a significant opportunity and a complex challenge for organizations across sectors. While the implementation journey requires careful navigation of organizational, regulatory, and technical hurdles, the potential benefits of increased innovation, improved decision-making, and enhanced collaboration make these efforts worthwhile. Success depends on strong organizational commitment, robust infrastructure, clear governance frameworks, and effective stakeholder engagement, all supported by continuous monitoring and improvement processes. As organizations continue to evolve their approaches to data openness, the focus must remain on balancing accessibility with security while creating sustainable systems that can adapt to changing needs and requirements, ultimately enabling them to harness the full potential of open data while effectively managing associated risks.

References

Alexopoulos, C., Saxena, S., Rizun, N., Shao, D. (2023). A framework of open government data (OGD) e-service quality dimensions with future research agenda. Records Management Journal, 33: 1, pp. 20-32

Bargh, M. S., Choenni, S., and Meijer, R. 2016. "Meeting Open Data Halfway: On Semi-Open Data Paradigm," in Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, Montevideo Uruguay: ACM, March, pp. 199–206.

Bonina, C., & Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. Government Information Quarterly, 37(3), 101479.

Cagnazzo, C. (2021). The thin border between individual and collective ethics: the downside of GDPR. www.thelancet.com/oncology Vol 22 November 2021

Curto-Millet, D. and Shaikh, M. (2018). The emergence of openness in open-source projects: the case of openEHR. Journal of Information Technology (32:4), pp. 361-379.

Davies, T. (2010). Open data, democracy and public sector reform: A look at open government data use from data.gov.uk. Oxford Internet Institute, University of Oxford.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly, 33(1), 15-27.

D'Ignazio, C., and Klein, L. F. 2020. Data Feminism, Ideas Series, Cambridge, Massachusetts: The MIT Press.

Deleuze, G. (1992) Postscript on the societies of control , in "October" 59, Winter, pp. 3-7.

Ferguson, A. G. 2020. Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.

Gonzalez-Zapata, F., and Heeks R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. Government Information Quarterly (32: 4), October 2015, pp. 441-452.

Guimarães, P. 2023. "Reproducibility With Confidential Data: The Experience of BPLIM," Harvard Data Science Review (5:3), The MIT Press.

Hagner, M. 2018. "Open Access, Data Capitalism and Academic Publishing," Swiss Medical Weekly (148:0708), pp. w14600–w14600.

Hoyng, R. 2017. "Transparency in the Rupture? Open Data and the Datafication of Society," Tecnoscienza (8), pp. 190–195.

Hossain, M. A., Dwivedi, Y. K., and Rana, N. P. 2016. "State-of-the-Art in Open Data Research: Insights from Existing Literature and a Research Agenda," Journal of Organizational Computing and Electronic Commerce (26:1–2), pp. 14–40.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258-268.

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. Government Information Quarterly 33 (1), pp. 89-104

Jiménez, A. C., and Estalella, A. 2023. Free Culture and the City: Hackers, Commoners, and Neighbors in Madrid, 1997–2017, Cornell University Press.

Johnson, J. A. 2014. "From Open Data to Information Justice," Ethics and Information Technology (16:4), pp. 263–274.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society 1(1).

Lindman, J. and Nyman, L. (2014). The businesses of open data and open source: Some key similarities and differences. Technology Innovation Management Review, Vol January 2014 , pp. 12-17 .

Lyon, D. (2001). Surveillance society. Open University Press, Buckinham, UK.

Lyon, D. (2001) Facing the future: seeking ethics for everyday surveillance, in "Ethics and Information Technology", 3, pp. 171-181.

Open Data Institute. (2018). The Data Ethics Canvas. Open Data Institute. https://theodi.org/article/data-ethics-canvas/

Peloquin, D., DiMaio, M., Bierer, B., Barnes, M. (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. European Journal of Human Genetics (2020) 28:pp. 697–705.

Poirier, L. (2024). Enacting data context: Fixing meaning in transparency data initiatives. Big Data & Society 11(3).

Powell, A. 2021. Undoing Optimization: Civic Action in Smart Cities, New Haven, CT: Yale University Press.

Runeson, P. and Olsson, T. (2020). Challenges and Opportunities in Open Data Collaboration – a focus group study. 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Portoroz, Slovenia, 2020, pp. 205-212.

Shaikh, M., and Vaast, E. 2016. “Folding and Unfolding: Balancing Openness and Transparency in Open Source Communities,” Information Systems Research (27:4), INFORMS, pp. 813–833.

Smith, G. and Sandberg, J. (2018). Barriers to innovating with open government data: Exploring experiences across service phases and user types. Information Polity 23 (2018), pp. 249–265

Staunton, C., Slokenberga, S., and Mascalzoni, D. (2019). The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks. European Journal of Human Genetics (2019) 27:1159–1167.

Van de Vyvere, and Colpaert, P. (2022). Using ANPR data to create an anonymized linked open dataset on urban bustle. European Transport Research Review, (2022) 14:17.

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. Government Information Quarterly, 33(2), 325-337

Wessels, B., Finn, R. L., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood, R., Taylor, M. J., Tsoukala, V., Wadhwa, K., and Wyatt, S. 2014. “Issues in the Development of Open Access to Research Data,” Prometheus (32:1).

Young, M., Rodriguez, L., Keller, E., Sun, F., Sa, B., Whittington, J., and Howe, B. 2019. “Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing,” in Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta GA USA: ACM, January 29, pp. 191–200.

Zikopoulos, P. C., C. Eaton, D. deRoos, T. Deutsch and G. Lapis (2012). Understanding big data. Analytics for enterprise class hadoop and streaming data. New York, McGraw-Hill.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R. and Sheikh Alibaks, R. (2012). Socio-technical Impediments of Open Data. EJEJ 10 (2).