Received 5 June 2025
Revised 19 July 2025
Accepted 3 September 2025

Ethical concerns of open government data

Fredrick Ishengoma and Deo Shao
Information Systems and Technology, The University of Dodoma College of Informatics and Virtual Education, Dodoma, United Republic of Tanzania

Abstract

Purpose – This study aims to investigate the ethical challenges associated with open government data (OGD) initiatives.

Design/methodology/approach – A systematic literature review (SLR) based on the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 protocol was adopted to synthesize peer-reviewed and grey literature on ethical issues in OGD. The SLR is augmented by an illustrative analysis of real-world examples drawn from the synthesized literature, serving to demonstrate practical manifestations of ethical dilemmas.

Findings – The study identifies four primary ethical domains in OGD ecosystems: privacy risks, data misuse, digital inequality and environmental sustainability. It reveals that technical safeguards like anonymization are often insufficient against re-identification threats. Real-world examples highlight how public–private data synergies contribute to surveillance and systemic bias, particularly in predictive policing and healthcare systems. In addition, unequal access to data resources exacerbates the digital divide, and environmental impacts of data-intensive technologies such as blockchain and artificial intelligence training are emphasized.

Practical implications – The study proposes a set of policy and practical interventions including differential privacy mandates, algorithmic accountability frameworks, community-centered data literacy programs and energy-efficient data infrastructure guidelines.

Originality/value – This study contributes to the literature by developing a comprehensive ethical taxonomy for OGD by integrating normative ethical theories.

Keywords Open government data, Ethics, PRISMA, Bias, Data misuse

Paper type Research paper

1. Introduction

The open government data (OGD) initiative stands as a fundamental element of present-day governance as it seek to enable citizen empowerment and drive public service improvement through open access to government data sets while also supporting innovation (Attard *et al.*, 2015; Shao, 2024). The widespread adoption of OGD platforms like data.gov and the European Data Portal shows an increasing understanding of how data-driven governance can create transformative changes. The swift growth of OGD initiatives has triggered ethical debates about privacy issues and the misuse of data which also contributes to widening social disparities (Zuiderwijk *et al.*, 2020; Martin *et al.*, 2021). Advanced data analytics techniques with artificial

Journal of Ethics in Entrepreneurship and Technology
Vol. 5 No. 2, 2025
pp. 206-222
Emerald Publishing Limited
e-ISSN: 2633-7444
p-ISSN: 2633-7436
DOI 10.1108/JEET-06-2025-0035

intelligence (AI) and Internet of Things amplify existing challenges by making it possible to extract sensitive data from data sets that appear harmless (Tene and Polonetsky, 2020; Shao *et al.*, 2023).

OGD supports evidence-based policymaking, yet it also enables surveillance and discriminatory practices through malicious exploitation (Kitchin, 2014; Zhang, 2024). Disparities in data access and literacy aggravate social and economic inequalities which weaken the inclusive capabilities of OGD projects (Safarov *et al.*, 2017). Moreover, studies (Rocher *et al.*, 2019; Zhang, 2024) highlight that it is possible to re-identify people from anonymized databases while predicting private attributes and behavioral patterns which questions the effectiveness of existing privacy protection measures. The merging of OGD with AI decision-making frameworks has developed fresh risks that includes algorithmic bias and discrimination which disproportionately affect marginalized communities (Eubanks, 2018; O'Neil, 2016).

The expanding academic literature on OGD still lacks sufficient exploration into ethical aspects especially with regard to new technological developments. Moreover, most of the existing research efforts have concentrated on technical and organizational barriers to OGD deployment while ignoring wider societal effects (Zuiderwijk *et al.*, 2020; Martin *et al.*, 2021). Several researchers applied justice and fairness principles to evaluate OGD policies (Safarov *et al.*, 2017; Zhang, 2024), but other scholars concentrated on the ethical duties of data processors and the inherent worth of information according (Floridi, 2013; Tene and Polonetsky, 2020). The differing perspectives demonstrate the necessity of an integrated framework that combines theoretical insights with empirical evidence. This study aims to explore the taxonomy of ethical concerns for OGD through a Rawlsian justice and Floridi's information ethics framework. The specific study research questions are:

RQ1. What are the primary ethical concerns associated with OGD, particularly in relation to privacy risks, data misuse, inequality and environment?

RQ2. How can Rawlsian justice and Floridi's information ethics provide a theoretical foundation for understanding and addressing these ethical challenges?

RQ3. What policy and technological interventions can mitigate the ethical risks of OGD while promoting transparency, accountability and equity?

This study's primary contribution lies in developing a comprehensive ethical taxonomy for OGD that integrates normative ethical theories with real-world examples. This approach offers a structured analytical framework, systematically linking diverse ethical concerns with their underlying technical mechanisms, ethical implications and real-world consequences. By applying the theoretical lenses of Rawlsian justice and Floridi's information ethics, the study provides interconnected understanding of why certain OGD practices are ethically problematic, thereby moving beyond descriptive accounts and offer an evaluative framework for governance and policy.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical foundations of the study, covering ethical principles in data governance and other related works. Section 3 presents the methodology used in the study systematic literature review (SLR) and real-world examples approach. Section 4 presents the taxonomy of ethical concerns in OGD. Section 5 presents the discussion of the study findings. Section 6 presents the policy and practical implications of the study and conclusion and future research is presented in Section 7.

2. Theoretical foundations

2.1 Ethical principles in data governance

Ethical OGD management requires strong governance system based on principles that harmonize transparency needs with accountability requirements while maintaining fairness and protecting privacy rights. These principles serve as basic building blocks for data ethics and hold significant positions in public administration and human rights discussions.

Transparency stands as the foundational principle of OGD which allows data to be accessible and understandable to different stakeholders, thus, building trust and promoting civic engagement (Zuiderwijk *et al.*, 2020). The implementation of transparency requires deliberate adjustment to prevent unexpected negative outcomes like sensitive data disclosure or increased surveillance opportunities (Tene and Polonetsky, 2020; Martin *et al.*, 2021). Researchers benefit from granular data sets release which simultaneously creates heightened re-identification risks when auxiliary data sources come into play (Rocher *et al.*, 2019; Zhang, 2024).

Proper accountability mechanisms are needed to maintain ethical standards for OGD usage by making data processors and policymakers responsible. Rawlsian justice requires that OGD projects avoid favoring privileged groups while excluding disadvantaged populations. The principal gains importance from research findings which show that biased data sets and algorithmic systems maintain systemic inequalities (Rocher *et al.*, 2019; Benjamin, 2019). Thus, the gap between those with access to technology and data skills continues to obstruct fair utilization of OGD in rural areas and among low-income populations (Shao, 2024).

Privacy stands as a top concern as the detailed nature and networked structure of OGD increases the chance of individuals being re-identified and their data misused (Zhang, 2024). Floridi's information ethics reinforces the ethical duty to safeguard data subjects by establishing privacy as a basic right for today's digital world (Floridi, 2013; Taddeo and Floridi, 2018). The latest developments in privacy-preserving technologies including differential privacy and federated learning present effective remedies to address these challenges (Abowd, 2018; Dwork *et al.*, 2017). Theoretical frameworks such as Rawlsian justice and Floridi's information ethics provide foundational ethical lenses for evaluating OGD initiatives. These frameworks guide the interpretation of technical issues by developing principles that supports equitable access and the inherent value of data, informing how data governance should operate.

2.2 Literature review

Research into OGD ethics is extensive yet scattered as this field requires input from multiple disciplines. Fo instance, public administration scholars have directed most of their research toward organizational and technical challenges involved in OGD implementation which includes data quality, interoperability and stakeholder engagement (Attard *et al.*, 2015; Shao and Saxena, 2019). Research regarding data anonymization has yet to comprehensively confront how traditional methods fall short against sophisticated re-identification techniques (Rocher *et al.*, 2019).

The field of data ethics has developed substantial theoretical advancements regarding the ethical aspects of data governance by using frameworks such as Floridi's information ethics and Rawlsian justice (Rawls, 1971; Floridi, 2013; Taddeo and Floridi, 2018). These theoretical frameworks establish the inherent worth of information entities while also demanding fair allocation of data advantages. The application of these theories to OGD has not yet reached full development as only limited research exists on how they can guide practical policy measures (Martin *et al.*, 2021). The existing human rights literature acknowledges that OGD can help drive social justice agendas but fails to address adequately the dangers posed by data misuse and surveillance (Tene and Polonetsky, 2020). The application of OGD by both governmental and nongovernmental entities for mass surveillance purposes generates essential debates regarding the equilibrium between transparency and individual privacy (Eubanks, 2018; O'Neil, 2016). Researchers has pinpointed distinct challenges like re-identification risks and algorithmic bias but these investigations remain isolated and overlook the complex relationships among ethical issues (Rocher *et al.*, 2019; Eubanks, 2018). The swift development of AI and machine learning technologies has generated emerging ethical issues like OGD-based predictive policing and social scoring which current research has yet to fully examine (O'Neil, 2016; Shao and Saxena, 2019).

3. Methodology

This study primarily uses a SLR based on the PRISMA protocol to synthesize existing peer-reviewed and gray literature on the ethical challenges associated with OGD initiatives. To enhance the practical relevance of the findings, the SLR is augmented by an illustrative analysis of real-world examples drawn from the synthesized literature.

3.1 Systematic literature review with PRISMA-P

The 2015 PRISMA-P statement consists of guidelines and item checklist to uphold methodological rigor while ensuring reports of systematic reviews remain reproducible and clear (Moher *et al.*, 2015).

3.1.1 Identification phase. The study conducted systematic examination of pertinent studies through diverse academic databases alongside gray literature sources during the identification phase. Four major academic databases were queried: Scopus, Web of Science, IEEE Xplore and PubMed. Google Scholar served as a tool to find gray literature including government reports, policy documents and working papers that frequently offer important perspectives on the practical complexities of OGD implementation. Boolean operators (AND, OR) and truncation symbols (*) were used to refine the search and capture variations in terminology. The following keywords and phrases were used: "open government data," "open data," "public sector data," "government transparency," "privacy risks," "data misuse," "digital inequality," "algorithmic bias," "surveillance," "re-identification," "Rawlsian justice," "Floridi's information ethics," "data ethics," "human rights," "artificial intelligence," "machine learning," "differential privacy" and "data governance." The initial search yielded 1,250 records.

3.1.2 Screening phase. This phase evaluated titles and abstracts of all 1,250 records to determine their relevance to the research questions. Studies that did not cover OGD and those which focused purely on technical or operational aspects without exploring ethical implications were not considered. After evaluation the data set was narrowed down to 450 relevant records which undergo full-text review. Research that did not maintain a distinct ethical focus or contained redundant information and which could not be accessed in full text was removed from consideration. The elimination process brought the data set down to 200 records.

3.1.3 Eligibility phase. Assessment of 200 records determined their compliance with the study's quality and relevance standards. The quality assessment examined their methodological rigor together with their contribution and relevance to OGD ethics.

3.1.4 Inclusion phase. The review included 55 studies selected from 200 records after a thorough evaluation process. The included studies represent a diverse range of perspectives on OGD ethics and are categorized into the following themes: The selected studies provide multiple viewpoints on OGD ethics which fall under several thematic categories including privacy risks, data misuse, inequality and environmental and sustainability concerns.

3.1.5 Data analysis. The synthesis of the included studies' findings used thematic analysis using the Braun and Clarke (2006) protocol and NVivo qualitative data analysis software. The process involved three stages of coding, theme development and refinement.

3.2 Illustrative analysis of real-world examples

To enhance the empirical evidence base of the SLR findings, an illustrative analysis of real-world examples was conducted. The selection of these examples focused on their clear connection to the main ethical issues identified in the literature, such as privacy risks, data misuse and inequality. Each selected example demonstrates particular ethical challenges and their real-world consequences, offering enhanced insight into the practical manifestations of these issues. A structured analysis framework was used to evaluate each example, examining its context, ethical dilemmas, outcomes and lessons learned. This method provided a comprehensive examination of how technical elements interact with organizational structures and societal influences within OGD initiatives.

4. Taxonomy of ethical concerns in open government data

The examination of ethical issues in OGD introduces a detailed classification system divided into four main areas including privacy risks, data misuse, inequality challenges and environmental effects.

4.1 Privacy risks

4.1.1 Re-identification and de-anonymization. The development of machine learning and data linkage technology has made conventional anonymization techniques outdated. For instance, the Netflix Prize data set revealed the potential to identify individuals through the combination of anonymized movie ratings and public information (Munson *et al.*, 2025). Research indicates that data sets that seem highly anonymized remain at risk of re-identification when combined with external data sources (Rocher *et al.*, 2019). Although k-anonymity and differential privacy offer potential solutions to these risks their effectiveness is constrained by the complexity and size of current data sets (Dwork *et al.*, 2017). The proliferation of open data repositories together with more advanced re-identification algorithms has intensified these issues which makes protecting individual privacy in public data sets more difficult (Zhang, 2024).

The process of re-identification violates informed consent principles because people do not know their data can be traced to them. Vulnerable populations face surveillance and discrimination along with other adverse effects due to this exposure (Clayton *et al.*, 2021). Insurers might use re-identified health data to refuse coverage while employers could make biased hiring choices based on such information (Martin *et al.*, 2021). The ethical dilemma emerges from trying to find the right balance between making data accessible and keeping personal privacy protected (Zuiderwijk *et al.*, 2020; Clayton *et al.*, 2021). The difficulties escalates in anonymizing complex data sets and the necessity for strong privacy protection methods (Safarov *et al.*, 2017).

To approach re-identification, current methods assess the vulnerability of anonymized data sets by attempting to link them with external data sources using techniques like *k*-anonymity or differential privacy, though these have known limitations . Measuring the success of de-anonymization often involves quantifying the number of re-identified individuals or the confidence level of attribute inference from supposedly private data.

4.1.2 Surveillance capitalism. Private entities can use OGD for mass surveillance through the process of combining public data with corporate AI technologies (Ishengoma *et al.*, 2022). Using OGD companies have developed predictive algorithms which track consumer activity to deliver targeted advertisements to individuals (Zuboff, 2019). The term "surveillance capitalism" describes a system where personal information becomes a commodity and businesses apply sophisticated analytics tools to influence consumer actions (Eubanks, 2018). The combination of OGD with corporate surveillance methods poses ethical concerns as it mixes public and private data usage and challenges the fundamental goals of OGD (Tene and Polonetsky, 2020).

OGD used for surveillance activities damages public trust in government programs while increasing power disparities between corporations and citizens (Benjamin, 2019). Private companies use OGD to monitor user movements and behaviors which results in the generation of detailed profiles for purposes like targeted advertising and discriminatory practices (O'Neil, 2016; Ishengoma, 2022; Alexopoulos *et al.*, 2023). The practice breaches personal privacy while simultaneously rise ethical debates about governmental duties to oversee private use of public data.

For instance, the application of OGD through predictive policing systems in Chicago demonstrates how surveillance practices and social inequality intersect whereby public crime data training the algorithm resulted in biased targeting of low-income minority communities which sustained systemic discrimination (Benjamin, 2019). This demonstrates how using OGD to develop AI systems presents ethical concerns when the training data embodies historical biases and social inequalities. Moreover, the work by Eubanks (2018) stresses the importance of transparency and accountability throughout the development and deployment processes of predictive policing systems. Technical approaches to surveillance capitalism involve analyzing the integration of OGD with corporate AI technologies to build detailed behavioral profiles (Figure 1). Measuring its impact includes evaluating the extent to which public data contributes to targeted advertising, predictive consumer analytics and the erosion of individual autonomy. Figure 1 illustrates the various privacy risks that come with OGD.

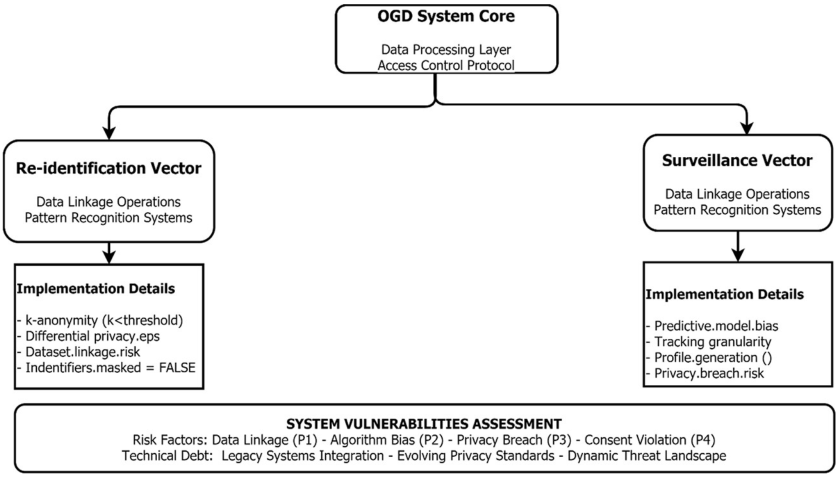

Figure 1. Privacy risks with OGD
Source: Figure by authors

4.2 Misuse of data

4.2.1 Malicious exploitation. With OGD, publicly available data sets can be misused for fraudulent activities, identity theft and discriminatory practices (Zhang, 2024). Voter registration data released with OGD initiatives has been misused to target people with deceptive political ads while also attempting to suppress voter turnout by spreading misinformation (Tufekci, 2018). Without strong protective and monitoring systems open data ecosystems become vulnerable to malicious activities since adversaries can take advantage of systemic weaknesses (Zhang, 2024).

Data ethics requires compliance with the principle of nonmaleficence which the misuse of OGD violates as it demands data initiatives to avoid harming people and communities. The exploitation of OGD for harmful purposes undermines public confidence in open data systems because people start to view their personal data as vulnerable and become hesitant to engage in data-sharing activities (Zuiderwijk *et al.*, 2020; Zhang, 2024). The misuse of data not only affects individuals but develop broader societal problems such as weakening democratic institutions and reinforcing social disparities.

Moreover, political disinformation through OGD manipulation damages electoral system integrity and fuels division which poses risks to democratic governance principles (Persily, 2017; Matasick *et al.*, 2020). For instance, the 2016 US presidential election witnessed a typical example of OGD misuse through the exploitation of voter registration data. Political operatives used public voter data to execute precise disinformation campaigns which leveraged OGD granularity to sway voter behavior and affect election results (Persily, 2017).

The situation illustrates the ethical dangers of open data usage during politically sensitive times and demonstrates the critical necessity for enhanced protective measures against data misuse. This situation prompts essential ethical inquiries regarding governmental duties in maintaining OGD system security to protect democratic institutions when data becomes a tool for subversion (Tufekci, 2018). Detecting malicious exploitation requires developing monitoring systems and strong protective measures to identify and prevent the misuse of publicly available data sets for fraudulent activities or identity theft. Measuring this involves tracking unauthorized data access, unusual usage patterns and the spread of misinformation derived from manipulated OGD.

4.2.2 Misinterpretation and misrepresentation. Data sets released without appropriate documentation and quality controls can lead users to misinterpret data which then results in flawed analyses and decisions. Some users face challenges in accurately interpreting OGD as data complexity and absence of standardized formats and metadata make it difficult to understand (Zuiderwijk *et al.*, 2020). The limited technical knowledge that normal citizens possess adds to existing complications which result in data misuse and misrepresentation according (Martin *et al.*, 2021).

Misinterpretation of OGD can lead to broad negative impacts such as resource misallocation and increased social inequalities when used to guide public policy (Eubanks, 2018). When health data gets misinterpreted, it can results to negative consequence in public health policies that can negatively impact vulnerable groups (Clayton *et al.*, 2021). To overcome these obstacles organizations must dedicate themselves to improving data literacy while providing the necessary support to enable users to use OGD effectively (Kempeneer *et al.*, 2025).

For instance, misinterpretation and politicization of COVID-19 statistics in Brazil contributed to delayed or inappropriate public health responses, exacerbating the severity of the crisis in certain regions. Misinterpretations of public health data caused authorities to underestimate COVID-19 which led to misinformation spreading across populations and unnecessary fatalities (Castro *et al.*, 2021). This example demonstrates how misinterpretation poses ethical dangers and shows why strong quality control measures and thorough documentation are essential for maintaining OGD accuracy and dependability. Figure 2 illustrates the ethical issues that emerge from the misuse of OGD.

Approaching misinterpretation involves ensuring data sets are released with proper contextual metadata and standardized formats to reduce user confusion. Measuring misrepresentation can be done by assessing the accuracy of derived analyses and decisions against the original data's intent and limitations, particularly when technical skills are lacking among users.

4.3 Exacerbation of inequality

4.3.1 Digital divide. Disparities in technology access and data literacy form a critical barrier to equitable OGD use among different demographic and geographic populations resulting from the digital divide (Kempeneer *et al.*, 2025). Rural and low-income areas experience the most severe digital divide as inadequate infrastructure including limited broadband connectivity and restricted access to computing resources obstructs effective OGD utilization (Shao, 2024). A lack of dedicated programs and capacity-building initiatives increases these disparities as marginalized groups frequently do not have the essential abilities to understand and use OGD for civic decisions and involvement (Safarov *et al.*, 2017; Zuiderwijk *et al.*, 2020).

The digital divide persists in violation of Rawls' difference principle as it states inequalities should only exist when they serve to help the least advantaged people in society (Rawls, 1971). OGD initiatives that overlook the digital divide unintentionally sustain existing power imbalances and block marginalized communities from accessing both social and political advantages of open data. The exclusionary effects of these initiatives counteract the fundamental principles of equity and inclusion which OGD data projects aim to advance supporting the UN Sustainable Development Goals (Ishengoma and Shao, 2025).

India's rural-urban disparities demonstrate a clear example of the digital divide in OGD initiatives. Urban areas benefit greatly from OGD programs through better public services and civic engagement but rural communities struggle with accessing these data sets because of insufficient digital infrastructure and connectivity issues (Bhattacharya *et al.*, 2020). Rural

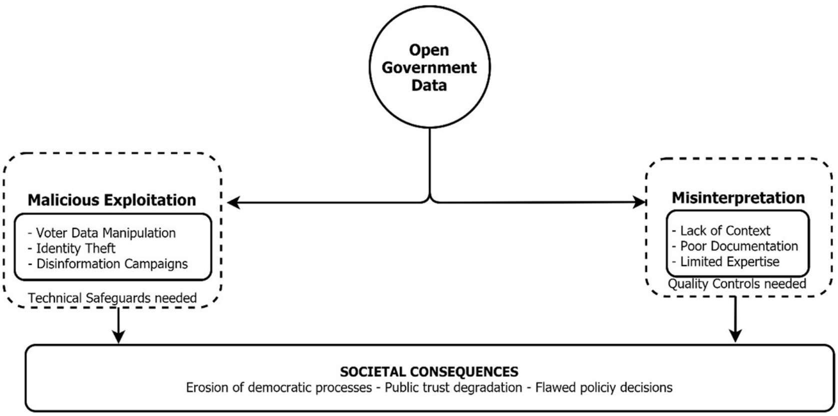

Figure 2. Misuse of data in OGD

Source: Figure by authors

populations stay excluded from the benefits of OGD which causes ongoing regional inequalities. The Indian situation demonstrates how context-specific solutions are necessary to overcome distinct obstacles rural and low-income communities face to make OGD initiatives fair and inclusive (Shao, 2024). Measuring the digital divide involves quantifying disparities in broadband connectivity and access to computing resources, especially in rural and low-income areas. Addressing this requires assessing the effectiveness of targeted capacity-building programs designed to improve data literacy among marginalized groups.

4.3.2 Algorithmic bias. The presence of algorithmic bias within OGD presents major challenges for fair decision-making processes when integrated with AI systems. The inherent biases present in OGD data sets become magnified through machine learning algorithms which then produce discriminatory results that negatively impact marginalized groups. Studies reveal that predictive policing systems which use biased historical crime data end up wrongly identifying minority neighborhoods as high-crime areas which leads to increased policing and reinforces systemic discrimination (O'Neil, 2016; Benjamin, 2019). The observed effects emphasize why fairness and accountability must be central when developing AI systems that operate based on openly provided government data.

The presence of algorithmic bias within OGD presents challenges for fair decision-making processes when integrated with AI systems. The inherent biases present in OGD data sets become magnified through machine learning algorithms which then produce discriminatory results that negatively impact marginalized groups. Studies reveal that predictive policing systems which use biased historical crime data end up wrongly identifying minority neighborhoods as high-crime areas which leads to increased policing and reinforces systemic discrimination (O'Neil, 2016; Benjamin, 2019).

Studies have shown that, algorithmic bias in OGD can manifests through predictive policing systems used and that algorithms trained on historical crime data from OGD systems tend to display racial bias by forecasting more crime in neighborhoods populated mainly by minorities (Wang *et al.*, 2023). As a result of these practices disadvantaged communities face intensified policing and surveillance which adds to their marginalization (Benjamin, 2019). These reveal the ethical dangers of using biased OGD for AI training and demonstrate why algorithmic decision-making processes need transparency and accountability (O'Neil, 2016). To approach algorithmic bias, it's necessary to analyze the historical biases embedded within OGD data sets used for training machine learning models. Measuring bias involves auditing algorithmic outputs for disproportionate or discriminatory impacts on specific demographic groups, such as in predictive policing systems.

Figure 3 illustrates how OGD contributes to growing ethical challenges regarding inequality.

4.4 Environmental and sustainability concerns

4.4.1 Energy consumption of data centers. The rising ethical issue surrounding the environmental impact of OGD becomes more pressing as climate change continues to intensify. The main challenge of this issue stems from the significant energy demand of data centers that manage large volumes of government data storage and processing. Contemporary data centers manage petabytes of data which demands extensive computing resources and cooling systems that add to their environmental impact through carbon emissions. Google operated data centers which provide support for multiple OGD projects across the globe used around 7.6 billion kWh of electricity during the year 2019 according to Google's 2020 report. Google has advanced its use of renewable energy sources while numerous other data centers continue to depend on nonrenewable energy sources. The need for robust cooling systems arises because these facilities run high-performance computing

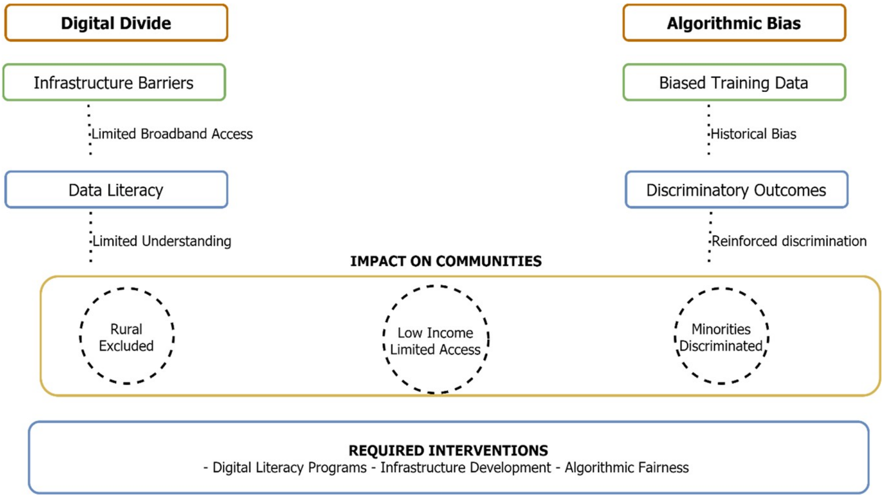

Figure 3. Exacerbation of inequality in OGD
Source: Figure by authors

hardware continuously which creates significant heat output. The inefficiency of traditional air-cooling systems has driven data centers to implement liquid cooling technologies which provide improved thermal management but require higher upfront costs.

When OGD initiatives grow larger they generate more data storage and processing needs which increases environmental damage and requires sustainable data center solutions that use renewable energy sources. Measuring the energy consumption of data centers involves quantifying their electricity demand, particularly for high-performance computing and cooling systems. Technical approaches focus on implementing energy-efficient infrastructure and transitioning to renewable energy sources to reduce carbon emissions.

4.4.2 Blockchain technologies and their energy demands. Blockchain technologies used in OGD ecosystems to maintain data integrity and traceability add to the sustainability issues beyond conventional data centers. Proof-of-work (PoW) blockchain networks like Bitcoin and Ethereum require intensive computational work for mining which leads to high energy consumption. PoW demands miners to solve cryptographic puzzles to confirm transactions and append new blocks to the blockchain which requires substantial computational power and results in excessive energy use. The city of Zug in Switzerland serves as an important example as it integrated blockchain technology into its public service operations. The city received negative feedback regarding its blockchain infrastructure's environmental footprint even though it provided improved transparency and security benefits (Zug City Council, 2018).

Proof-of-stake (PoS) mechanisms target energy reduction by prioritizing network stake over computational power but still face limited adoption in the blockchain industry. However, introducing blockchain technology into OGD frameworks creates further complexity due to its increased energy usage and the resulting doubts about whether these technologies can scale and remain viable in environmentally friendly governance models.

Measuring the energy demand of blockchain technologies in OGD ecosystems necessitates quantifying the terahashes per second (TH/s) and associated electrical power (watts) consumed by PoW consensus mechanisms, such as those in Bitcoin and Ethereum. Technical approaches to mitigate this involve transitioning from computationally intensive PoW, which requires miners to solve complex cryptographic puzzles, to lower-energy alternatives like PoS, where network validation is based on staked cryptocurrency rather than raw processing power, aiming to reduce energy consumption comparable to entire nations.

4.4.3 Artificial intelligence-driven analytics and resource intensity. AI-based analytics that process large-scale data sets intensify the environmental impact of OGD initiatives through their resource-intensive nature. Deep learning structures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) need exhaustive training processes using extensive data sets to reach peak performance levels. The training procedure requires multiple forward and backward network passes which result in substantial computational power consumption and heat generation. Strubell *et al.* (2019) report that the carbon footprint of training one large-scale AI model reaches beyond 626,000 pounds of carbon dioxide which matches the entire lifetime emissions produced by five average American vehicles. The UK's National Health Service (NHS) demonstrates practical AI use by implementing predictive analytics to improve patient care pathways. The project attracted scrutiny because of its substantial energy consumption throughout the training phase of the model (NHS Digital, 2020). The deployment stage for AI models in OGD systems requires continuous inference activities that consume less energy than training but nonetheless add to environmental impact. The integration of OGD with AI-driven analytics demands energy-efficient algorithmic development and hardware enhancements through the use of specialized accelerators such as Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) to minimize their environmental footprint. Measuring the resource intensity of AI-driven analytics involves quantifying the energy consumption and carbon footprint associated with training large-scale deep learning models. Technical approaches focus on developing energy-efficient algorithms and leveraging specialized hardware accelerators like GPUs and TPUs. Figure 4 illustrates how OGD affects environmental and sustainability issues.

Figure 5 presents the summary diagram of how Rawlsian Justice and Floridi's Information Ethics underpin the four thematic domains of the taxonomy developed in this study. These theoretical perspectives do not appear as abstract moral ideals but serve as interpretive lenses that give structure and ethical depth to the classification of concerns identified through the empirical synthesis. Rawlsian Justice supports the analysis of digital inequality, algorithmic bias and environmental externalities by emphasizing fairness, equitable access and the moral priority of the least advantaged. This framing guides the evaluation of how OGD practices can exacerbate social and infrastructural asymmetries. Conversely, Floridi's Information Ethics provides the normative grounding for assessing privacy risks, re-identification and data misuse. It treats information subjects as moral entities, whose autonomy and dignity are compromised through surveillance, commodification and inadequate safeguards. Taken together, these theories enable a cohesive interpretation of the taxonomy – not as a descriptive inventory of harms, but as an ethically reasoned structure that links empirical manifestations of OGD challenges to normative principles of justice, respect and accountability.

5. Discussion

This study identifies multiple ethical challenges within OGD. This research supports existing literature and expands it but also identifies diverging areas that require additional investigation. The combination of Rawlsian justice principles with Floridi's information

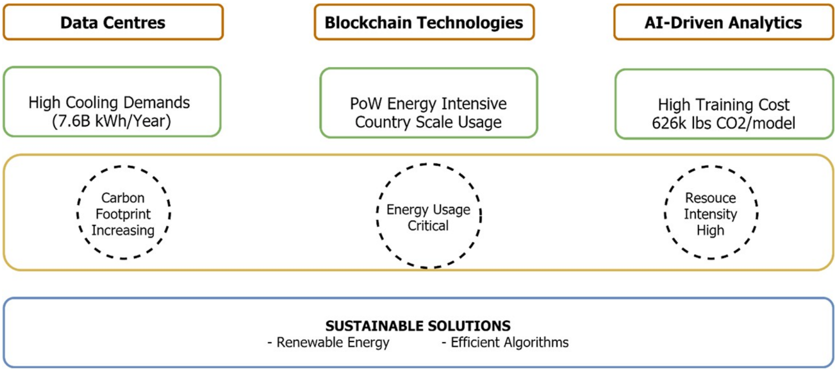

Figure 4. Environmental and sustainability concerns in OGD
Source: Figure by authors

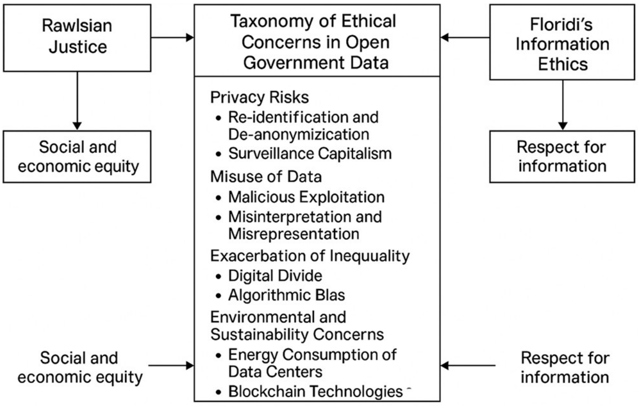

Figure 5. Normative foundations of the ethical taxonomy in OGD (how Rawlsian Justice and Floridi's Information Ethics underpin the four thematic domains of ethical concerns)
Source: Figure by authors

ethics establishes a strong theoretical framework that helps understand ethical challenges and guides policymakers and practitioners with practical solutions.

The research demonstrates that re-identification and de-anonymization results align with earlier studies including Netflix Prize data set re-identification risk findings (Munson *et al.*, 2025). Both Rocher *et al.* (2019), Zhang (2024) have pointed out how traditional anonymization techniques fail to withstand advanced data linkage methods. This research differs from prior studies by demonstrating how contemporary privacy-preserving technologies, including differential privacy frameworks, fall short when confronting today's extensive and complex data environments. This study demonstrates that the application of differential privacy according to Dwork *et al.* (2017) results in trade-offs which restrict OGD's usefulness for research and policymaking. This tension reveals the necessity for privacy protection measures that balance data set accessibility with individual rights protection.

The study investigation into surveillance capitalism resonates with Zuboff's (2019) examination of personal data commodification and Eubanks' (2018) study of OGD's potential misuse for widespread surveillance. Although Zuboff's analysis centers on corporate exploitation of data, this study expands the discussion to examine governmental regulation of public data usage by private companies. Recent occurrences demonstrate this point through examples like Chicago's predictive policing algorithms use of OGD which resulted in disproportionate targeting of marginalized groups (Benjamin, 2019). The results dispute the positive perspective of OGD as a democratizing factor and emphasize the necessity for stricter regulatory systems to protect civil rights.

Research shows that OGD exploitation through voter records misuse during the 2016 presidential election matches Tufekci's study about data weaponization for political misinformation. This study expands the discussion by demonstrating that data misuse results in broader societal impacts which degrade democratic functions and reinforce existing social disparities. Eubanks (2018), Benjamin (2019) both demonstrate that biased data sets and AI systems serve to maintain existing systemic inequalities. This study presents a different approach from prior research by focusing on tailored context-specific solutions like community-driven digital literacy initiatives and accessible OGD platforms to tackle these disparities. The rural-urban gap in India (Bhattacharya *et al.*, 2020) demonstrates that specialized solutions are essential to tackle specific obstacles experienced by marginalized groups. The literature's standard one-size-fits-all strategy faces criticism while there is a need for more inclusive and equitable OGD initiatives.

The study results regarding algorithmic bias support O'Neil's (2016) analysis of biased data leading to unequal results especially in predictive policing systems. Although O'Neil examines the technical elements of algorithmic bias this study highlights the ethical duties of governmental bodies and data handlers to enforce justice and responsibility. Meanwhile, predictive policing algorithms demonstrate the essential requirement for transparent systems and participatory monitoring in creating and implementing AI technologies that use OGD.

The findings regarding the environmental sustainability align with several studies that highlight the significant energy consumption and carbon footprint associated with these technologies. For instance, Shehabi *et al.* (2016) underscore the substantial electricity demand of data centers, noting that they account for approximately 1% of global electricity usage, which resonates with our observation on Google's data centers consuming 7.6 billion kWh in 2019. Similarly, Strubell *et al.* (2019) emphasize the high energy demands of training AI models, paralleling to the case study of the NHS's predictive analytics project, where model training was criticized for its environmental impact. These convergent findings highlight a pressing need for sustainable practices within OGD ecosystems. However, some studies present divergent perspectives; for example, De Vries (2018) focuses primarily on blockchain technologies like Bitcoin, whose energy consumption is disproportionately higher due to PoW consensus mechanisms.

While our analysis acknowledges the energy intensity of blockchain, we argue that it is not universally representative of all OGD applications, as many OGD initiatives do not rely heavily on blockchain. Instead, their primary environmental concerns stem from data storage and AI processing. This distinction is crucial because it suggests that targeted interventions—such as renewable energy adoption for data centers and energy-efficient algorithm design for AI—are more relevant solutions than broad dismissals of OGD's viability based solely on blockchain's unsustainable practices. Therefore, while the overall trend points towards significant environmental impacts, nuanced analyses reveal that specific components of OGD ecosystems require tailored sustainability strategies rather than blanket condemnations.

Beyond the commonly cited benefits OGD holds profound value as a public good, facilitating societal progress in myriad ways. Its utility extends to fostering scientific discovery, enabling evidence-based policymaking and driving innovation across sectors. For instance, granular public health data sets can be instrumental in identifying emergent disease patterns, evaluating the efficacy of public health interventions and even "uncovering successful but less-well known treatments in medicine" by allowing researchers to analyze real-world outcomes on a large scale. Similarly, open economic data can empower citizens and businesses to make informed decisions, stimulate economic growth through new applications and services and enhance market efficiency by reducing information asymmetry.

The ethical challenge in OGD governance lies in navigating the inherent and often competing tension between maximizing data utility for public benefit and safeguarding individual privacy and rights. This is not a simple contrast but a dynamic trade-off, where increasing one often comes at the expense of the other. The findings, particularly concerning re-identification risks and surveillance capitalism, highlight that while greater data granularity and linkage might unlock richer insights, they simultaneously amplify the potential for privacy breaches and discriminatory practices. The current limitations of technical safeguards like anonymization and differential privacy in complex, interconnected data sets further underscore this delicate balance, as over-anonymization can render data useless, while insufficient protection invites misuse. Thus, OGD initiatives must move beyond a mere release-and-forget model towards a sophisticated understanding of this trade-off.

Constructing effective mechanisms for valuing utility against privacy requires integrating legal, ethical and technical considerations. Conceptually, this involves a "privacy by design" philosophy where privacy protections are baked into the data lifecycle from collection to dissemination, rather than being an afterthought. Practically, this translates to implementing dynamic access control protocols that allow varying levels of data granularity based on user legitimacy and intended purpose. Ethical frameworks, such as the principles of proportionality and necessity derived from human rights discourse, can guide decisions on what data to release, to whom and under what conditions, ensuring that any intrusion on privacy is justified by a clear and substantial public good that cannot be achieved through less privacy-invasive means.

6. Policy and practical implications

Governments need to focus their efforts on developing and applying cutting-edge privacy protection technologies such as differential privacy and federated learning to address re-identification and de-anonymization threats. Clear guidelines on data anonymization and ethical release procedures needs to support these technologies. Policymakers need to set up accountability systems to make sure that data processors and AI developers follow ethical guidelines especially in critical applications like predictive policing and public health.

The digital divide demands specific measures to make OGD available to all while promoting data literacy. To address this, governments need to prioritize infrastructure development for broadband connectivity and computational resources in rural and low-income regions. Marginalized communities need to gain the power to engage in data-driven governance through subsidized technology access and community-driven digital literacy programs. These initiatives will make certain that the advantages of OGD spread fairly across society according to Rawlsian justice concepts and equity principles. Moreover, the government should develop and implement laws that control how private organizations use OGD especially when the data has potential uses in surveillance or disinformation campaigns. Governments must facilitate international partnerships to solve the worldwide ethical challenges of OGD which involve cross-border data sharing and AI-driven decision-making ethics.

Addressing the environmental concerns, policymakers must advocate for the adoption of green data center practices, including the utilization of renewable energy sources to offset the carbon footprint. Incentives for data center operators to invest in energy-efficient infrastructure can further drive reductions in energy consumption. Additionally, regulatory measures should be implemented to encourage the use of low-energy consensus mechanisms within blockchain technologies. For AI-driven analytics, policies should mandate the inclusion of environmental impact assessments as part of the model development lifecycle, ensuring that researchers and practitioners consider the ecological ramifications of their work.

7. Conclusions and future work

The study presents a complete examination of ethical issues related to OGD while categorizing them into privacy risks alongside data misuse and their effects on inequality and the environment. The research combines Rawlsian justice principles with Floridi's information ethics to establish a detailed theoretical structure for grasping these issues while delivering practical suggestions for policymakers and practitioners. The research results demonstrate that OGD operates as a dual-use tool which can benefit public welfare and cause private damage while emphasizing the essential requirement for strong governance to reduce ethical threats.

Future research needs to investigate how emerging technologies like generative AI and blockchain create ethical challenges within OGD systems. Blockchain technology presents opportunities for better data transparency and accountability yet generates ethical concerns due to its environmental effects and scalability challenges. Future research needs to analyze participatory governance models in OGD projects to determine how citizen involvement improves transparency, accountability and equity. Long-term societal effects of OGD initiatives require longitudinal studies for evaluation particularly about privacy, inequality and democratic governance. Through a focus on these research gaps future studies can help build OGD ecosystems that are both inclusive and ethical while safeguarding individual rights and preventing social inequalities in data-driven governance.

References

Abowd, J.M. (2018), "The U.S. Census bureau adopts differential privacy", *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2867-2867.

Alexopoulos, C., Saxena, S., Janssen, M. and Rizun, N. (2023), "Whither the need and motivation for open government data (OGD) promotional strategies?", *Digital Policy, Regulation and Governance*, Vol. 25 No. 2, pp. 153-168.

Attard, J., Orlandi, F., Scerri, S. and Auer, S. (2015), "A systematic review of open government data initiatives", *Government Information Quarterly*, Vol. 32 No. 4, pp. 399-418.

Benjamin, R. (2019), *Race after Technology: Abolitionist Tools for the New Jim Code*, Polity Press.

Bhattacharya, S., Singh, A. and Bhattacharya, S. (2020), "Digital divide in India: a barrier to open government data", *Journal of Information Technology and Politics*, Vol. 17 No. 3, pp. 234-248.

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101.

Castro, M.C., Kim, S., Barberia, L., Ribeiro, A.F., Gurzenda, S. and Ribeiro, K.B. (2021), "The impact of covid-19 on inequalities in Brazil", *The Lancet Public Health*, Vol. 6 No. 2, pp. e92-e93.

Clayton, E.W., Evans, B.J., Hazel, J.W. and Rothstein, M.A. (2021), "The ethical implications of re-identification in health data. Journal of law", *Medicine and Ethics*, Vol. 49 No. 1, pp. 22-34.

De Vries, A. (2018), "Bitcoin's growing energy problem", *Joule*, Vol. 2 No. 5, pp. 801-805.

Dwork, C., Roth, A. and Vadhan, S. (2017), "The algorithmic foundations of differential privacy", *Foundations and Trends® in Theoretical Computer Science*, Vol. 9 Nos 3-4, pp. 211-407.

Eubanks, V. (2018), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press.

Floridi, L. (2013), *The Ethics of Information*, Oxford University Press.

Ishengoma, F. (2022), "Exploring critical success factors towards adoption of m-government services in Tanzania: a web analytics study", *App and Website Accessibility Developments and Compliance Strategies*, pp. 225-253, IGI Global.

Ishengoma, F.R., Shao, D., Alexopoulos, C., Saxena, S. and Nikiforova, A. (2022), "Integration of artificial intelligence of things (AIoT) in the public sector: drivers, barriers and future research agenda", *Digital Policy, Regulation and Governance*, Vol. 24 No. 5, pp. 449-462.

Ishengoma, F. and Shao, D. (2025), "A framework for aligning e-government initiatives with the sustainable development goals", *Journal of Innovative Digital Transformation*, Vol. 2 No. 1, pp. 73-89.

Kempeneer, S., Wolswinkel, J. and van Maanen, P.B.G. (2025), "Beyond the portal: Increasing citizen understanding of OGD with storytelling", *Information Polity*, Vol. 30 No. 1, pp. 65-81.

Kitchin, R. (2014), *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, SAGE Publications.

Martin, C., Blommestein, K. and Klievink, B. (2021), "The ethics of open government data: a systematic literature review", *Government Information Quarterly*, Vol. 38 No. 3, pp. 101-115.

Matasick, C., Alfonsi, C. and Bellantoni, A. (2020), "Governance responses to disinformation: How open government principles can inform policy options", *OECD Working Papers on Public Governance*, Vol. 39, pp. 1-45.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M.,… Stewart, L.A. (2015), "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement", *Systematic Reviews*, Vol. 4 No. 1, pp. 1-9.

Munson, J., Cummins, B. and Zosso, D. (2025), "An introduction to collaborative filtering through the lens of the netflix prize", Knowl Inf Syst, Vol. 67 No. 4, pp. 3049-3098, doi: 10.1007/s10115-024-02315-z.

NHS Digital (2020), "AI in healthcare: opportunities and challenges", available at: https://digital.nhs.uk/data-and-information/information-services/ai-in-healthcare

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishing Group.

Persily, N. (2017), "The 2016 U.S. election: can democracy survive the internet?", *Journal of Democracy*, Vol. 28 No. 2, pp. 63-76.

Rawls, J. (1971), *A Theory of Justice*, Harvard University Press.

Rocher, L., Hendrickx, J.M. and de Montjoye, Y.A. (2019), "Estimating the success of re-identifications in incomplete datasets using generative models", *Nature Communications*, Vol. 10 No. 1, pp. 1-9.

Safarov, I., Meijer, A. and Grimmelikhuijsen, S. (2017), "Utilization of open government data: a systematic literature review of types, conditions, effects, and users", *Information Polity*, Vol. 22 No. 1, pp. 1-24.

Shao, D. (2024), "Empirical analysis of open government data usage in Tanzania", *Information Discovery and Delivery*, Vol. 52 No. 1, pp. 73-84.

Shao, D. and Saxena, S. (2019), “Barriers to open government data (OGD) initiative in Tanzania: Stakeholders’ perspectives”, *Growth and Change*, Vol. 50 No. 1, pp. 470-485.

Shao, D., Ishengoma, F.R., Alexopoulos, C., Saxena, S., Nikiforova, A. and Matheus, R. (2023), “Integration of IoT into e-government”, *foresight*, Vol. 25 No. 5, pp. 734-750.

Shehabi, A., Smith, S.J., Masanet, E., Koomey, J., Brown, R., Horner, N.,… Wong, L. (2016), *United States Data Center Energy Usage Report*, Lawrence Berkeley National Laboratory.

Strubell, E., Ganesh, A. and McCallum, A. (2019), “Energy and policy considerations for deep learning in NLP”, arXiv preprint arXiv:1906.02243.

Taddeo, M. and Floridi, L. (2018), “Regulate artificial intelligence to avert cyber arms race”, *Nature*, Vol. 556 No. 7701, pp. 296-298.

Tene, O. and Polonetsky, J. (2020), “Privacy in the age of big data: a time for big decisions”, *Stanford Law Review Online*, Vol. 64, pp. 63-69.

Tufekci, Z. (2018), “Engineering the public: big data, surveillance, and computational politics”, *First Monday*, Vol. 19 No. 7.

Wang, F., Zhu, H., Wu, Y., He, D. and Liu, X. (2023), “From data to action: Leveraging open data to drive knowledge-based intelligent governance”, *Proceedings of the Association for Information Science and Technology*, Vol. 60 No. 1, pp. 845-849.

Zhang, X. (2024), “A more secure framework for open government data sharing based on federated learning”, *Government Information Quarterly*, Vol. 41 No. 4, p. 101981.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs.

Zug City Council (2018), “Blockchain initiative update”, available at: www.zug.ch/en/verwaltung/stadtpräsidialdienst/news/2018/blockchain-initiative-update.html

Zuiderwijk, A., Chen, Y.C. and Salem, F. (2020), “Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda”, *Government Information Quarterly*, Vol. 38 No. 3, pp. 101-115.

Further reading

Google (2020), “Environmental report 2020”, available at: www.google.com/sustainability/environmental-report/

Johnson, P.A., Robinson, P.J. and Philpot, S. (2020), “Participatory data governance: a framework for citizen engagement in data-driven decision-making”, *Journal of Urban Technology*, Vol. 27 No. 1, pp. 1-20.

Corresponding author
Fredrick Ishengoma can be contacted at: ishengomaf@gmail.com