online first

Data Fairness in Trusted Research Environments (TREs): Towards New Ethical and Procedural Approach

Equità dei dati nei Trusted Research Environments (TREs): verso un nuovo approccio etico e procedurale

FRANCESCO TAVA
Francesco.Tava@uwe.ac.uk
KYLE ALVES
Kyle.Alves@uwe.ac.uk
ELIZABETH GREEN
Elizabeth7.Green@uwe.ac.uk

AFFILIAZIONE
University of the West of England, Bristol (UK)

online first
3 novembre 2025

ABSTRACT

This paper examines how the ethical principle of fairness can be embedded within the governance and operational models of Trusted Research Environments (TREs). While TREs are designed to ensure data security and privacy, the authors argue that ethical governance should extend beyond trust to include fairness in the distribution of data-related risks and benefits. They propose a fairness model grounded in public value, combining defensive motivations (non-discrimination, redistribution, equality of opportunity) with progressive ones (transparency, participation, right to justification). The paper outlines how fairness can be operationalized through process design, management practices, and policy frameworks, offering practical recommendations for more equitable and transparent TRE operations. By conceptualizing fairness as a dynamic and measurable principle, the study bridges ethical theory and procedural application, aiming to enhance public legitimacy, stakeholder confidence, and the overall justice of data governance systems.

SOMMARIO

Questo articolo esplora come il principio etico di equità (fairness) possa essere integrato nella governance e nei modelli operativi dei Trusted Research Environments (TREs). Nati per garantire sicurezza e privacy, i TREs non dovrebbero basarsi solo sul principio di fiducia, ma anche sulla giustizia nella distribuzione di rischi e benefici derivanti dall'uso dei dati. Si propone così un modello di equità fondato sul valore pubblico, che unisce motivazioni difensive (non discriminazione, redistribuzione, uguaglianza di opportunità) e progressive (trasparenza, partecipazione, diritto alla giustificazione). L'articolo mostra come l'equità possa essere resa operativa attraverso la progettazione dei processi gestionali e decisionali dei TREs, fornendo raccomandazioni per politiche e pratiche più giuste e trasparenti. In tal modo, l'equità diventa un principio dinamico e misurabile, indispensabile per rafforzare la legittimità etica e la fiducia pubblica nei sistemi di governance dei dati.

KEYWORDS

Data fairness
Trust
Ethical governance
Public value
Data justice
Trusted research environments

PAROLE CHIAVE

Equità dei dati
Fiducia
Governance etica
Valore pubblico
Giustizia dei dati
Trusted Research Environments

DOI: 10.53267/20250104

CC BY NC ND

online first
3 novembre 2025

1. INTRODUCTION

This paper explores how the ethical principle of fairness can be integrated into the governance and operational models of Trusted Research Environments (TREs), with the goal of enhancing their functioning and utilization of public value in data management.

Whilst there is no universally agreed-upon definition of a TRE, multiple initiatives are actively working to establish clearer standards and frameworks. In essence, a TRE is generally understood as a secure environment that enables approved researchers to access and analyze sensitive or de-identified data, while safeguarding privacy and confidentiality through technical, procedural, and governance controls[1].

In the UK, the Office for National Statistics (ONS) Secure Research Service (SRS)—a prime example of a TRE—operates under the Five Safes framework, which ensures that data access is granted only to safe people, for safe projects, in safe settings, using safe data, and that outputs are checked for disclosure risk[2].

Terminology varies both within the UK (e.g., Secure Data Environment, Data Safe Haven) and internationally (e.g., microdata laboratories, virtual data labs, safe havens)—reflecting considerable variation in design, governance, and functionality[3]. For instance, while many TREs incorporate comprehensive output checking, others may omit this component, leading to differences in how effectively privacy is protected.

This definitional ambiguity has significant implications for the principle of fairness: inconsistent use of the term "TRE" risks creating misleading assumptions about the protections offered to individuals, communities, and researchers. The SATRE specification aims to develop a broad and inclusive definition: a TRE encompasses not only the computing and technological infrastructure but also the information governance and data management processes required to support secure research with sensitive data[4]. Here, sensitive data refers to any data requiring disclosure control measures (for instance, personally identifiable information or other data with privacy implications). By articulating these assumptions explicitly, SATRE seeks to advance a transparent and consistent foundation for assessing fairness in TRE design and implementation.

Recent research from the UK Statistics Authority (UKSA) and the Economic and Social Research Council (ESRC) indicates that TREs are generally trusted by the public[5]. However, emerging academic literature has challenged the assumption that TREs are about trust[6].

This paper partially endorses this argument acknowledging that TREs are *not solely* about trust. This raises a follow-up question: *What additional principle(s)—beyond trust—should guide an ethically sound management of TREs?*

While we do not claim to offer a definitive answer, we argue that future research should focus on identifying and operationalizing a set of guiding principles to underpin the governance of TREs. In this paper, we highlight fairness as one such principle—central to ensuring that data access and use within TREs not only protects privacy but also promotes public value in an equitable and transparent manner.

In summary, this paper aims to critically examine and enhance the role of fairness within TREs. Specifically, it seeks to:

- Investigate the concept of fairness in the context of data access, governance, and public benefit generation within TREs.
- Propose a new fairness model that complements existing trust frameworks by focusing on how TREs deliver public value.
- Develop practical process designs to operationalize fairness in TREs, offering clear guidance on how engagements in such environments can be structured and managed.

These aims are guided by the central hypothesis that TREs need to be not only *safe* but also *fair* in order to be fully trusted by the public and stakeholders. This hypothesis builds on the growing consensus that trust in data environments is multidimensional. Being seen to be fair in procedures and practices (often referred to as "procedural fairness") is increasingly recognized as a core component of institutional trust[7].

This paper is guided by three overarching research questions. First, it asks how fairness and trust interact within TREs, and how these two concepts might be co-optimized.

This question explores the relational dynamics between fairness and trust, recognizing that perceptions of equity and justice increasingly underpin the legitimacy of data governance systems.

Second, the paper investigates which design features and decision-making processes are most effective in fostering fairness in the governance and operation of TREs. This includes an examination of institutional arrangements, procedural safeguards, and participatory mechanisms that can embed fairness into the everyday functioning of these environments.

Finally, the research considers the broader benefits of integrating fairness into data access and governance practices. By doing so, it aims to assess not only the ethical principle of fairness, but also its practical implications for public confidence, stakeholder engagement, and the delivery of public value.

The study is expected to make several key contributions to both scientific discourse and policy development. It will offer a conceptual framework for understanding fairness within TREs, situating this principle within broader debates on trust, governance, and data justice. Building on this foundation, the paper will propose a novel fairness model, explicitly oriented toward the generation of public benefit, and designed to complement existing trust-based frameworks.

In addition to theoretical contributions, the research will develop a set of practical recommendations for policy and organizational design. These will provide actionable guidance on how fairness can be operationalized through the structure, management, and oversight of TREs. Finally, the work aims to inform ongoing national and international conversations about how to design data access infrastructures that are not only secure and efficient, but also demonstrably fair and trustworthy.

By framing fairness as both a measurable and actionable principle, the paper seeks to bridge the gap between normative ideals and practical implementation—supporting the creation of more equitable and credible research environments.

2. BACKGROUND AND RATIONALE

Graham *et al.*[8] suggest that TREs do not foster public trust but instead minimize the *need* for it. Trust inherently involves vulnerability and the possibility of disappointment—elements that TREs are explicitly designed to eliminate through strong privacy and security controls. By offering technical and procedural safeguards, TREs seek to replace reliance on trust with assurance and risk mitigation.

This perspective has generated important debate. In response, Affleck *et al.*[9] contend that while TREs do mitigate certain risks, they do not address all public concerns. From this view, TREs serve not as replacements for trust but as tools that help people assess whether the institutions managing their data are trustworthy. In other words, while TREs can facilitate trust, they do not remove the need for it. Similarly, Jesudason[10] argues that verification can reduce uncertainty, but nevertheless increase feelings of trust.

In response to this criticism, Graham *et al.*[11] emphasized the need for linguistic precision—particularly in distinguishing between *trust* and *reliance*. They argue that while critics often conflate the two, the difference is significant: reliance involves predictable, risk-managed systems, whereas trust implies a willingness to be vulnerable. According to their view, TREs are designed to ensure reliability, not to cultivate trust directly. As they conclude, “TREs are (still) not about trust, although they may be part of building a health data research system that is trustworthy” (p. 660).

The landscape has grown increasingly complex as TREs are now expected to regulate access not only to traditional datasets but also to emerging assets such as artificial intelligence models[12]. While TREs are often valued for the continuity and predictability they offer, research has highlighted how they are also being drawn into uncharted territory, where risks remain poorly understood and insufficiently mapped[13]. This evolving context makes it increasingly difficult to calibrate the balance between security and trust within these infrastructures, and it reopens critical questions around vulnerability and, as we shall explore further, fairness.

This paper partially endorses the argument put forward by Graham et al., recognizing that TREs are *not solely* about trust. We argue that the ethical governance of TREs should be grounded in a broader set of guiding

online first
3 novembre 2025

principles. Identifying and operationalizing these principles will be a central objective for future scholarship in data ethics and governance.

In this paper, we focus on one such principle: fairness—a concept we contend is essential for ensuring that TREs are not only secure and efficient, but also just and publicly legitimate.

3. THE CASE FOR FAIRNESS IN TRES

Existing scholarship on the ethics of risk management highlights the limitations of deterministic approaches attempting to eliminate all risk by planning for worst-case scenarios. As Hansson[14] argues, most real-life decisions are made under conditions of uncertainty. Yet, in data access and governance (including TREs), a natural tendency exists to simplify this uncertainty by reframing it as calculable risk, leading to decision-making that often underestimates the complexity of both human and technological systems.

This complexity is particularly evident in data environments, where no system can fully eliminate the possibility of harm. In medical data sharing, for example, no disclosure control technique can guarantee absolute protection against re-identification[15]. Nonetheless, the societal benefits of responsible data use—such as breakthroughs in medical science—are widely acknowledged. Accordingly, a strict "zero-risk" approach is not only impractical but potentially counterproductive, as it could hinder progress in public health and research.

Importantly, risk acceptance is only justifiable within a fair and equitable framework. Hansson argues that "exposure of a person to a risk is acceptable if and only if this exposure is part of an equitable social system of risk-taking that works to her advantage"[16]. This principle underscores two crucial dimensions of fairness in data governance:

1. Reciprocal Benefit: Individuals accept certain risks because they anticipate collective or personal benefits—such as improved healthcare—within a shared social system.
2. Equity: These risks and benefits must be distributed fairly across the population. No group should disproportionately shoulder the burdens or be excluded from the rewards.

Public sentiment suggests a lack of confidence in how this fairness is currently managed by government and public and private firms. A 2018 survey by Imperial College's Institute of Global Health found that while UK respondents were more open to sharing medical data with research institutions than with commercial entities, only half were willing to do so. In the US, the figure dropped to one-quarter[17]. This hesitancy points not necessarily to a rejection of medical research, but to a deeper concern about how the benefits of data sharing are distributed.

Recent controversies have further eroded public trust. In 2023, *The Observer* reported that UK Biobank shared sensitive donor data with insurance companies, despite prior assurances that such information would only support academic research[18]. Similar breaches involving public health institutions and private companies—including the NHS, drug and insurance firms[19], Google DeepMind[20], and Palantir[21]—have raised questions not just of privacy, but of justice and fairness. This aligns with research findings in agricultural industries where farmers, after initially sharing data about their land and food production, are now more reluctant due to concerns about fairness—particularly in inequality of the sharing of benefits emerging from the use of their data[22].

While some benefit may have accrued to the public, financial advantages were clearly skewed in favour of corporate stakeholders. These incidents illustrate a systemic failure: a breakdown in the fair distribution of benefits and a violation of the trust that underpins public participation in data ecosystems.

Such failures highlight that technical safeguards alone are insufficient. What is needed is a more comprehensive model that incorporates fairness as a core evaluative criterion in data governance. Public willingness to share data increases when the purposes are transparent, the benefits are evident, and—critically—when those benefits are perceived as fairly shared. This is especially relevant in TREs, which are designed to securely manage access to sensitive data for research purposes.

4. TOWARD A FAIRNESS MODEL BASED ON PUBLIC BENEFIT

In the previous section, we introduced fairness as a fundamental principle in data governance. Fair-

ness, in this context, refers to a moral principle closely associated with social justice and the equitable redistribution of resources. Treating others fairly involves recognizing their moral equality, acknowledging their inherent worth as individuals, and avoiding wrongful discrimination while ensuring mutual respect.

Scholarly literature has highlighted several key aspects of fairness, such as "fair equality of opportunity" and the "right to justification"[23]. It is widely acknowledged that a fair social system should regulate the distribution of burdens and benefits among its members while managing socio-economic inequalities in a non-discriminatory manner. Focusing on fairness as fair equality of opportunity involves going beyond mere non-discrimination; it requires creating conditions that allow every moral agent to fully realise their potential. For instance, ensuring fairness in a democratic election process would mean not only guaranteeing voting rights to all eligible citizens, but also making sure that every voter has access to correct information and the necessary capacities to interpret it. This application of fairness extends into sensitive moral domains such as education, media, and communication.

Additionally, conceiving fairness through the right to justification implies that in a fair social system, every person should be respected as an individual who both offers and demands justification. In this view, one fundamental right is not only to receive fair treatment but also to claim it on the basis of legitimate and justifiable arguments. Such a process cannot occur in isolation. On the contrary, fairness requires an ongoing interplay among moral agents who must be recognized, included, and able to participate in the fair distribution of resources. On this premise, intersubjective relations must be structured with the aim of empowering people to form their own judgements on issues of concern and demand fair treatment. This process becomes especially important in situations of perceived injustice. People may choose to enact their right to justification by, for example, campaigning against an unfair rule and demanding better treatment. From this perspective, fairness is seen less as an acquired right and more as a moving target, which political action must continually pursue and reinforce.

Drawing upon this emphasis on its relational nature, Giovanola and Tiribelli have recently described fairness as "fair equality of relationship" in the sense that, in conditions of fairness, human relations ought to "foster particular individuals' agency, triggering genuine attachments, commitments, values and ends", instead of enabling potentially detrimental phenomena such as political polarization and prejudice[24].

In data governance, fair conduct implies that personal data should not be collected or shared if the associated risks disproportionately target specific individuals or groups. Additionally, ensuring an equal distribution of the benefits (not just the risks) from data sharing is essential for fair data governance. This involves two distinct motivations: one defensive and one progressive:

- The defensive motivation for fair data governance focuses on non-discrimination, redistribution, and equality of opportunity. If there is a reasonable possibility that data sharing could lead to an unequal distribution of risks, a fair decision would be to halt data sharing until these risks are adequately mitigated.
- The progressive motivation for fair data governance emphasizes data agency, transparency, and mutual benefits. According to the principles of fairness as the right to justification and fair equality of relationship, individuals must be informed about their role as data producers. They should be able to demand full explanations and justifications for how their personal data is used, the benefits it generates (whether economic, commercial, or public), and how the resulting data wealth is redistributed within society.

We argue that combining these defensive and progressive motivations can serve as a potent antidote to both data misuse and data skepticism. It empowers individuals to engage in data sharing more actively and conscientiously while mitigating the risks of privacy violation and identity disclosure.

The ethical approach we outlined can help enforce concepts of data justice by ensuring fairness in how people are made visible, represented, and treated as data producers. This conception is particularly beneficial in the context of national and supranational law-making processes, where the challenges surrounding the status and risks associated with personal

online first
3 novembre 2025

data sharing are becoming increasingly complex. According to Taylor[25], legal frameworks must achieve three fundamental goals to secure data justice. First, they must provide individuals with the legal capacity to know about the collection and use of their personal data. This aspect pertains to agency and transparency, aligning with the progressive motivation for fair data governance mentioned earlier. Second, legal frameworks must enable the detachment of personal data from automated commodification on global data markets while encouraging the analysis of big data for the common good. This goal relates to the need to anchor data access to the creation and redistribution of public benefits, which is essential for countering people's skepticism about personal data sharing. Third, the law must counteract technical conditions that might lead to intentional or unintentional discrimination. This aspect underscores the necessity of not only distributing data benefits but doing so fairly. It recognizes the social and political aspects, such as the inclusion of data producers from minority groups. By addressing these three goals, legal frameworks can better ensure data justice and promote a fairer, more transparent data governance system.

such as TREs. Delivering on the promise of fairness requires embedding points of action into the regular activities associated with sharing data in a TRE. From this perspective, we can draw on research from the disciplinary field of Management for guidance on how to operationalize fairness in these data sharing transactions.

In this section, we offer support for an operational view of how fairness can be delivered by drawing on empirical evidence. We first present (in Figure 1) a model created during an engagement in January 2024 that included two of the authors and several data custodian organizations, including the UK Office of National Statistics (ONS), UK Health Data Research (UK HDR), and Administrative Data Research UK (ADR UK) among others. The model presents an illustrative example of the general process of a typical query against a data set held in a TRE. This model also suggests a set of fairness considerations appropriate for each step of the process. Further, we suggest the connection to specific actors who perform that process step, indicating the roles of those who participate in the specific tasks in the process where fairness can be operationalized.

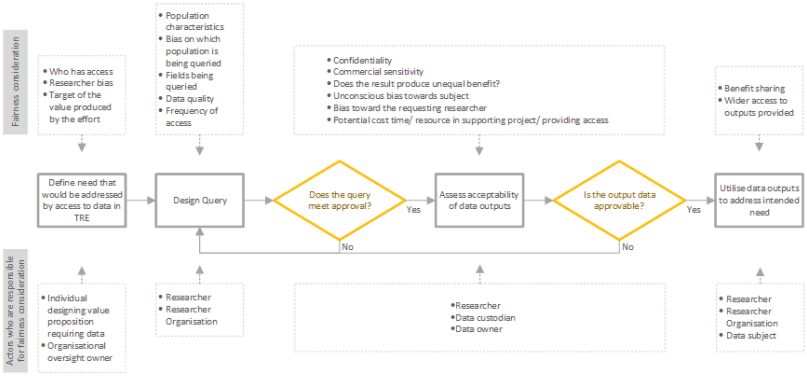

Figure 1

General process for querying data in a TRE, with suggested relevant fairness factors for each step shown above each step, and the actors performing each step attached below.

5. PROCESS DESIGN FOR OPERATIONALIZING FAIRNESS IN TRES

Having clarified our understanding of data fairness and outlined the core principles of the proposed fairness model within data governance, we now turn to the practical dimension of this framework by examining how fairness can be operationalized within secure research infrastructures

We might start by employing a framework that distinguishes between data governance, governing, and management. Janssen *et al*. defines 'governance' as an overarching organising logic that influences decision making for any data-focused activity[26]. They further

suggest that 'governing' describes the collective actions to execute this logic, while the individual actions themselves are 'management' tasks. Actions of governing (*doing the management tasks*) are then required to operationalize the delivery of fairness as an outcome. An illustration of this framework, as applied to the general TRE process drawn from practice, is provided in Figure 2. This suggests that the concept of fairness is an influencing factor affecting the goal-orientation of the actions associated with governance.

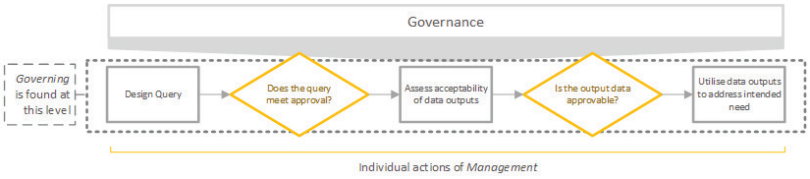

Figure 2

Governance logic shown at top which influences decision making; governing is shown as reflecting the establishment of a collection of actions to execute the logic; and management tasks in TREs are shown in the process steps.

There are strong parallels here between what is described as 'management' and the general understanding of process management. Processes are simply a collection of interconnected tasks and activities that deliver a specific outcome or strategy, both within and across organizational boundaries[27].

In empirical examinations of processes of sharing data and information between organizations, Kembro *et al.*[28] identify the importance of a fair benefit-sharing model and the protection of confidential information as antecedents enabling sharing. Their recommendations include agreed performance measures amongst data sharing partners to ensure equitable benefit distribution; and contracts which embed fairness principles to eliminate potential opportunistic behavior by any single player. More generally, management research shows a positive correlation between perceptions of fairness and higher levels of trust between organizational entities in a working relationship.

In the domain of agriculture, Wiseman *et al.*[29] also identified fears of unequal distribution of data-related benefits and the exposure of confidential information. Farmers in their study expressed concerns of bearing an unfair balance of risk and vulnerability, further stating that they did not benefit adequately from the rewards associated with data sharing.

Recommendations from these cases shift focus to the need for ongoing dialogue for adjusting processes associated with data sharing to create fairness and ultimately greater trust. Such a balance, however, is in danger of being overwhelmed by partners with greater levels of power, effectively giving the less powerful no option but to enter unfair agreement.

The existence of a *'digital data divide*' was noted, describing the divide between those who contribute data and those who control, aggregate and share that data[30].

To address the power imbalance, regulation now exists within the EU to ensure that those who offer services requiring data, and those who contribute their data to that service both have 'data literacy and awareness' as a way for both parties to expose potential value of any particular data set. A similar requirement is in place for the operators of artificial intelligence (AI) systems using any data from contributors to train that AI. Mandating increased data literacy and awareness, however, is a *'toothless tiger'*. Such regulation relies on the assumption fairness consistently emerges when participants have higher levels of data literacy and awareness. Others take a more direct, operational approach.

In a case example from Australia, more transparent contracts were developed that emphasize fairness between data sharing participants[31]. These contracts rely on co-created, agreed metrics that measure process-level performance. Others move past contractual adjustments and metrics to directly address the need for process-focused fairness adjustments directly. Jakku *et al.*

online first
3 novembre 2025

describe adjustments at the process level, calling for organizations to improve the "everyday practices and decisions" that would enable fairness in data sharing[32]. Further, their research findings suggest the approach of building-in fairness at the process level through increased cooperative development of data curation and evaluation processes involving all participants.

Increased levels of fairness in organizational process design and execution have been shown to contribute to the overall perception of 'organizational justice'[33]. Using Colquitt's constructs in assessing performance of projects, Unterhitzenberger and Lawrence show how work processes that follow fairness principles are an important element to creating fair outcomes in managing change[34]. They state a need for organizations to embed fair procedures while also supporting team members in implementing those fair processes and procedures.

In practice, many of these recommendations reflect the approach suggested by one of the more relevant frameworks related to operating TREs. The widely used data management approach The Five Safes[35] itself is fundamentally a process ensuring that no confidential or sensitive data is exposed for analysis or publication. At each step of the Five Safes, those abiding its guidance are asked to make decisions shaped by principles of fairness and security. It details a set of checks and questions to rigorously monitor the process of monitoring the production and use of TRE data and outputs. This is a clear example of ensuring that fairness as a guiding principle in TREs requires the embedding of change at the process level.

Getting fairness right, however, demands that the process be designed for trial-and-error cycles for governing. Ostrom's extensive investigations into common-pool resources like shared sets of data in TREs led to the observation that the development of governance processes requires the freedom and acceptance of "a considerable amount of trial-and-error learning"[36]. Ongoing efforts to refine the delivery of fairness can lean on a robust body of research and a recent history of success in practice that comes from improvements at the process level. Continuous improvement using the various proven approaches of Total Quality Management (TQM)[37], Business Process Management (BPM)[38], and change management methods all rely on action at the process level to deliver organisational goals.

6. POLICY AND PRACTICE RECOMMENDATIONS

To enhance the legitimacy and ethical behaviours of TREs this paper argues that fairness must be embedded as a foundational principle and at the core of both processes and frameworks. Currently, the focus is heavily shifted towards prioritising privacy, security, and verifiability and often overlook or silo perception of equity, justice and public participation. Policies should be adapted to formally recognise fairness as a core, ever evolving part of TRE activity on par with traditional risk management. To help facilitate change and bring fairness to the forefront this section will outline policy and practice recommendations. The approach outlined in this section draws on the principles associated with continuous improvement cycles from the process management literature[39].

1. Determine the goal of fairness and how that goal can be either perceived and/or measured.
   Fairness must not be seen as a fixed goal but an emergent outcome—requiring active monitoring, data collection, and public feedback. It should not be a one-shot activity that is somehow completed once done. We suggest fairness should be a foundational core principle in TRE governance and standards should explicitly include fairness as a dimension and guiding design principle and reflected in the actions of governing the operation.

2. Establish what exists: illustrate the *'As-Is'*.
   Process mapping helps organizations to simultaneously gain sight of where decisions are made and actions are taken, while also identifying the responsible person/persons in the organization. The use of process mapping to support decision making has already been empirically established internationally in this environment[40]. We propose the same process examination would help TRE operators map out potential fairness blind spots, while understanding how and when fairness is actively being judged. This objective view can then expose potential power imbalances or

inequities at the process level, where they can be explicitly acted upon.

3. Understand that greater transparency in governance is possible by operationalizing TRE processes.
This transparency underpins clear accountability and oversight to each stage of governing the process. Once a process is established, emphasis could then be placed on the creation of a visible and transparent audit trail to evidence fairness, inviting external individuals to be public auditors. Embedded in this interaction of external review should be mechanisms which allow individuals to challenge fairness judgements. A clear picture of the above considerations can then be used to design the future *'to be'* processes, carrying momentum in continuous improvement to monitor the emergent operation's performance to ensure proper intended governance. An illustrative example of this approach, described above, is the newly developed Standardized Architecture for Trusted Research Environments (SATRE). SATRE proposes greater transparency in TRE process design to enable assessments of fairness.

4. Embed actions of capacity building
Training of TRE staff and data owners to better understand fairness and transparency in decision-making processes. This includes the activities and tasks of creating templates and records to capture socio-political, ethical and procedural justice dimensions of data use. Regular reflection and action, as part of the continuous improvement cycle, will then integrate captured insight into the management actions of governing. This may likely include the creation of fairness officers who can at as a liaison between public oversight and those responsible for TRE governance. Use of a process-centered approach has already demonstrated improvements in capacity building in the UK ONS and Eurostat[41], however these implementations were without a focus on fairness as described in this paper.

7. CONCLUSIONS

In this paper, we first argued for the inclusion of fairness as a core tenet in the management of Trusted Research Environments (TREs), alongside trust and other foundational principles. We then outlined a specific ethical conception of fairness, drawing on recent philosophical research, and adapted this concept to the distinctive landscape of data management. Building on this foundation, we proposed a process model supported by empirical examples and cases through which fairness can be effectively operationalized, closely aligning it with the concept of 'governing' within a Management framework. Finally, we distilled our findings into four policy and practice recommendations aimed at experts and researchers working in and with TREs.

What this paper has not addressed—and what constitutes a future objective of our research—is the integration of feedback from such experts on how the proposed model might function in practice. Moreover, as we emphasised at the outset, fairness represents only one among several ethical principles that can support more robust data access and governance practices. Future work will focus on identifying and discussion additional principles, such as solidarity, transparency, and reciprocity, with the aim of developing a comprehensive ethical toolbox for professionals in the field and for training purposes.

NOTE

1. Esme Mansouri-Benssassi, Simon Rogers, Jim Smith, and Felix Ritchie. "Machine Learning Models Disclosure from Trusted Research Environments (TRE), Challenges and Opportunities". arXiv:2111.05628 [Preprint], (2021). Available at: https://arxiv.org/abs/2111.05628 [Accessed: September 19, 2025].
Aridhia (n.d.) What is a Trusted Research Environment?. Available at: https://www.aridhia.com/what-is-a-trusted-research-environment/ [Accessed: September 19, 2025].

2. Office for National Statistics (ONS) (n.d.) *About the Secure Research Service*. Available at: https://www.ons.gov.uk/aboutus/whatwedo/statistics/secure-research-service [Accessed: September 19, 2025).

online first
3 novembre 2025

ADR UK (n.d.) *Trusted Research Environments- ADR UK*. Available at: https://www.adruk.org/trusted-research-environments/ [Accessed: September 19, 2025].

Five Safes (2025) Five Safes- decision-making framework. Available at: https://www.fivesafes.org/ [Accessed: September 19, 2025].

3. Esme Mansouri-Benssassi, Simon Rogers, Jim Smith, and Felix Ritchie. "Machine Learning Models Disclosure from Trusted Research Environments (TRE), Challenges and Opportunities". arXiv:2111.05628 [Preprint], (2021). Available at: https://arxiv.org/abs/2111.05628 [Accessed: September 19, 2025].

DNAnexus (n.d.) Trusted Research Environment (TRE) Explained. Available at: https://blog.dnanexus.com/2022-08-30-trusted-research-environment-tre-explained/ [Accessed: September 19, 2025].

Lifebit (n.d.) What is a Trusted Research Environment?. Available at: https://lifebit.ai/resources/trusted-research-environment/ [Accessed: September 19, 2025].

4. Standard Architecture for Trusted Research Environments (SATRE). https://satre-specification.readthedocs.io/en/stable/ [Accessed June 29, 2025].

5. B. Goldacre, J. Morely, "Better, broader, safer: using health data for research and analysis. A review commissioned by the Secretary of State for Health and Social Care", *Department of Health and Social Care* (2022). Available at: https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis [Accessed June 29, 2025].

6. More on this point hereby in section two.

7. On procedural fairness, with special focus on law courts, see Emily Gold La Gratta, Phil Bowen, "To be fair: procedural fairness in courts", *Centre for Justice Innovation* (November 2014): https://justiceinnovation.org/publications/be-fair-procedural-fairness-courts [Accessed June 29, 2025].

8. Mackenzie Graham, Richard Milne, Paige Fitzsimmons, and Mark Sheehan, "Trust and the Goldacre Review: why trusted research environments are not about trust", *Journal of Medical Ethics* 49, (2023): 670–673. https://doi.org/10.1136/jme-2022-108435

9. Paul Affleck, Jenny Westaway, Maurice Smith, and Geoff Schrecker, "Trusted research environments are definitely about trust", *Journal of Medical Ethics* 49, (2023): 656–657. https://doi.org/10.1136/jme-2022-108678

10. Edwin Jesudason, "Verification and trust in healthcare", *Journal of Medical Ethics* 49, (2023): 223–224. https://doi.org/10.1136/jme-2022-108634

11. Mackenzie Graham, Richard Milne, Paige Fitzsimmons, and Mark Sheehan, "TREs are still not about trust", *Journal of Medical Ethics* 49, (2023): 658–660. https://doi.org/10.1136/jme-2023-109031

12. Emily Jefferson, James Liley, Maeve Malone, *et al.*, "GRAIMATTER Green Paper: Recommendations for disclosure control of trained Machine Learning (ML) models from Trusted Research Environments (TREs)", *Zenodo* (2022). https://doi.org/10.5281/zenodo.7089491

13. Charalampia Kerasidou, Maeve Malone, Angela Daly, and Francesco Tava, "Machine learning models, trusted research environments and UK health data: ensuring a safe and beneficial future for AI development in healthcare", *Journal of Medical Ethics* 49, 12 (2023): 838–843. https://doi.org/10.1136/jme-2022-108696

14. Sven Ove Hansson, "Ethical Criteria of Risk Acceptance", *Erkenntnis* 59, (2003): 291–309. https://doi.org/10.1023/A:1026005915919

15. See on this Jim Smith, Maha Albashir, Seb Bacon, *et al.*, "SACRO: Semi-Automated Checking of Research Outputs", Zenodo (2023). https://doi.org/10.5281/zenodo.10055365

16. Sven Ove Hansson, "Ethical Criteria of Risk Acceptance", *Erkenntnis* 59, (2003): 291–309. https://doi.org/10.1023/A:1026005915919, p. 305

17. Justine Alford, "Public trust in health data sharing has sharply declined, survey reveals", *Imperial News*, https://www.imperial.ac.uk/news/200436/public-trust-health-data-sharing-sharply/ [Accessed June 24, 2025]

18. Shanti Das, "Private UK Health Data Donated for Medical Research May Be Shared with Insurance Companies", *The Guardian*, November 12, 2023. https://www.theguardian.com/technology/2023/nov/12/pri-

vate-uk-health-data-donated-medical-research-shared-insurance-companies. [Accessed June 24, 2025].

19. Randeep Ramesh, "NHS Patient Data to Be Made Available for Sale to Drug and Insurance Firms", *The Guardian*, January 19, 2014, https://www.theguardian.com/society/2014/jan/19/nhs-patient-data-available-companies-buy. [Accessed June 24, 2025]

20. Hal Hodson, "Revealed: Google AI Has Access to Huge Haul of NHS Patient Data", *New Scientist*, May 4, 2016, https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/. [Accessed June 24, 2025]

21. Andrea Chipman, "NHSE to Investigate Palantir for Possible Breach of FDP Contract", *Digital Health*, January 18, 2024, https://www.digitalhealth.net/2024/01/nhse-to-investigate-palantir-for-possible-breach-of-fdp-contract/. [Accessed June 24, 2025]

22. See on this Leanne Wiseman, Jay Sanderson, Airong Zhang, Emma Jakku, "Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming", *NJAS: Wageningen Journal of Life Sciences* 90–91(1), (2019): 1–10. https://doi.org/10.1016/j.njas.2019.04.007; Emma Jakku, Bruce Taylor, Aysha Fleming, *et al.*, "'If they don't tell us what they do with it, why would we trust them?' Trust, transparency and benefit-sharing in Smart Farming", *NJAS: Wageningen Journal of Life Sciences* 90–91(1), (2019): 1–13. https://doi.org/10.1016/j.njas.2018.11.002

23. On fair equality of opportunity, see especially John Rawls, *A Theory of Justice*, Cambridge, MA: Harvard University Press, (1971). On the right to justification, see especially Rainer Forst, *Justice, Democracy and the Right to Justification: Rainer Forst in Dialogue*, London: Bloomsbury Academic, (2014).

24. Benedetta Giovanola and Simona Tiribelli, "Weapons of moral construction? On the value of fairness in algorithmic decision-making", *Ethics Inf Technol* 24, 3 (2022): 7. https://doi.org/10.1007/s10676-022-09622-5

25. Linnet Taylor, "What is data justice? The case for connecting digital rights and freedoms globally", *Big Data & Society* 4, 2 (2017). https://doi.org/10.1177/2053951717736335

26. Marijn Janssen, Paul Brous, Elsa Estevez, *et al.*, "Data Governance: Organizing Data for Trustworthy Artificial Intelligence", *Government Information Quarterly* 37, 3 (2020): 101493. https://doi.org/10.1016/j.giq.2020.101493

27. On this matter, see especially Colin Armistead and Simon Machin, "Implications of business process management for operations management", *International Journal of Operations & Production Management* 17, 9 (1997): 886–898. https://doi.org/10.1108/01443579710171217; Mohamed Zairi, "Business process management: a boundaryless approach to modern competitiveness", *Business Process Management Journal* 3, 1 (1997): 64–80. https://doi.org/10.1108/14637159710161585; Peter Checkland, *Systems Thinking, Systems Practice: Includes a 30 Year Retrospective*, New Ed. edition, Wiley, (1999)

28. Joakim Kembro, Dag Näslund, and Jan Olhager, "Information Sharing across Multiple Supply Chain Tiers: A Delphi Study on Antecedents", *International Journal of Production Economics* 193, (2017): 77–86. https://doi.org/10.1016/j.ijpe.2017.06.032

29. Leanne Wiseman, Jay Sanderson, Airong Zhang, Emma Jakku, "Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming", *NJAS: Wageningen Journal of Life Sciences* 90–91(1), (2019): 1–10. https://doi.org/10.1016/j.njas.2019.04.007

30. Mark Andrejevic, "Big Data, Big Questions | The Big Data Divide", *International Journal of Communication* 8, (2014): 1673–1689

31. Leanne Wiseman, Jay Sanderson, Airong Zhang, Emma Jakku, "Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming", *NJAS: Wageningen Journal of Life Sciences* 90–91(1), (2019): 1–10. https://doi.org/10.1016/j.njas.2019.04.007

32. Emma Jakku, Bruce Taylor, Aysha Fleming, *et al.*, "'If they don't tell us what they do with it, why would we trust them?' Trust, transparency and benefit-sharing in Smart Farming", *NJAS: Wageningen Journal of Life Sciences* 90–91(1), (2019): 1–13. https://doi.org/10.1016/j.njas.2018.11.002

33. Jason A. Colquitt, "On the dimensionality of organizational justice: A

construct validation of a measure," *Journal of Applied Psychology* 86, 3 (2001): 386–400. https://doi.org/10.1037/0021-9010.86.3.386

34. Christine Unterhitzenberger, Kate Lawrence, "Fairness matters: organisational justice in project contexts", *Production Planning & Control* 36, 1 (2023): 45–60. https://doi.org/10.1080/09537287.2023.2251424

35. Felix Ritchie, "The 'Five Safes': a framework for planning, designing and evaluating data access solutions", *Data for Policy 2017: Government by Algorithm? (Data for Policy)*, London, Zenodo, (2017). https://doi.org/10.5281/zenodo.897821

36. Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge-New York: Cambridge University Press, (1990), p. 34.

37. William Edwards Deming, *Out of the Crisis*. Cambridge, MA: MIT Press, (1986).

38. Michael Hammer and James Champy, *Re-engineering the Corporation: A Manifesto for Business Revolution*, London: Nicholas Brealey, (1993).

39. *Inter alia*, from Deming Improvement Cycles: "Plan, Do, Check, Act (PDCA)"; from SixSigma: "Define, Measure, Analyse, Improve, Control (DMAIC)"; from process management: "Select, Record, Examine, Develop, Install/Implement, Maintain (SREDIM)"

40. See implementations of Semi-Automated Checking of Research Outputs (SACRO) in NHS East of England Secure Data Environment, UK Data Service Secure Lab, UK ONS, HDR UK, Instituto Nacional de Estadística (National Institute of Statistics of Uruguay).

41. Marco Stocchi, "Automatic checking of research outputs". Conference of European Statisticians: Expert Meeting on Statistical Data Confidentiality (December 2021), Poland. https://unece.org/sites/default/files/2021-12/SDC2021_Day2_Stocchi_AD.pdf [Accessed September 19, 2025].