Data & Knowledge Engineering 151 (2024) 102301

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

The unresolved need for dependable guarantees on security, sovereignty, and trust in data ecosystems

Johannes Lohmöller [a,*], Jan Pennekamp [a], Roman Matzutt [a], Carolin Victoria Schneider [b], Eduard Vlad [a], Christian Trautwein [b], Klaus Wehrle [a]

[a] Communication and Distributed Systems, RWTH Aachen University, Aachen, Germany
[b] Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany

ARTICLE INFO

Keywords:
Data sharing
Confidentiality
Integrity protection
Data Markets
Distributed databases

ABSTRACT

Data ecosystems emerged as a new paradigm to facilitate the automated and massive exchange of data from heterogeneous information sources between different stakeholders. However, the corresponding benefits come with unforeseen risks as sensitive information is potentially exposed, questioning data ecosystem reliability. Consequently, data security is of utmost importance and, thus, a central requirement for successfully realizing data ecosystems. Academia has recognized this requirement, and current initiatives foster sovereign participation via a federated infrastructure where participants retain local control over what data they offer to whom. However, recent proposals place significant trust in remote infrastructure by implementing organizational security measures such as certification processes before the admission of a participant. At the same time, the data sensitivity incentivizes participants to bypass the organizational security measures to maximize their benefit. This issue significantly weakens security, sovereignty, and trust guarantees and highlights that organizational security measures are insufficient in this context. In this paper, we argue that data ecosystems must be extended with technical means to (re)establish dependable guarantees. We underpin this need with three representative use cases for data ecosystems, which cover personal, economic, and governmental data, and systematically map the lack of dependable guarantees in related work. To this end, we identify three enablers of dependable guarantees, namely trusted remote policy enforcement, verifiable data tracking, and integration of resource-constrained participants. These enablers are critical for securely implementing data ecosystems in data-sensitive contexts.

1. Introduction

Data-driven analyses and business models are invaluable pillars for modern industries and societies alike. Their importance will increase with growing demands, requiring more complex and globally distributed operations and sophisticated collaborations to improve the status quo [1]. *Data ecosystems* provide the foundation for such data-driven analyses and business models as they center around automating data exchanges and value creation based on vast and heterogeneous data sources from various stakeholders [2]. Data ecosystems also create value beyond business, e.g., as a platform for research data lineage and exchange or to organize health records. Added value can be created by, for instance, improving algorithms underlying existing analytics or extracting

* Corresponding author.
E-mail addresses: lohmoeller@comsys.rwth-aachen.de (J. Lohmöller), pennekamp@comsys.rwth-aachen.de (J. Pennekamp), matzutt@comsys.rwth-aachen.de (R. Matzutt), cschneider@ukaachen.de (C.V. Schneider), vlad@comsys.rwth-aachen.de (E. Vlad), ctrautwein@ukaachen.de (C. Trautwein), wehrle@comsys.rwth-aachen.de (K. Wehrle).

https://doi.org/10.1016/j.datak.2024.102301
Received 15 March 2023; Received in revised form 5 January 2024; Accepted 11 March 2024
Available online 19 March 2024

new insights of previously recorded data [3]. Crucially, this process involves the integration of distributed data sources owned by different stakeholders. Here, data ecosystem initiatives such as International Data Spaces (IDS) [4] and GAIA-X [5] aim to provide a trustworthy environment for the discovery, sharing, and processing of available data, irrespective of specific domains. Various related work [6–8] complements these efforts, and selected deployed systems, such as distributed health data platforms [9,10], can already be considered early instantiations of data ecosystems.

However, the data prosperity promised by transitioning toward data ecosystems and their implied large-scale data exchanges among different stakeholders increases data confidentiality and privacy risks. Aggravatingly, current efforts to establish the necessary trust among these stakeholders heavily rely on organizational agreements and processes instead of technical or cryptographic measures to provide security guarantees [4,11]. These agreements frequently include a certification process for stakeholders prior to admission [12,13]. For instance, the IDS certification process asserts that participants use audited software and develop so-called "defense-in-depth" protection strategies that comprise redundant layers of (organizational) security measures for protection [4]. Participants receive no tangible additional security guarantees beyond this ahead-of-time certification. Most notably, they cannot verify that other participants handle their data as intended. Here, the lack of stronger guarantees effectively undermines data sovereignty and trust between participating stakeholders as soon as they start sharing their data.

We thus argue that one needs to make a fine distinction between *organizational security measures* that depend on trust in other stakeholders, and *dependable guarantees* that instead rely on technical or organizational means and require trust in these means, but not in another stakeholder's honesty. Within this paper, we establish that current data ecosystems frequently focus on the former, which might be sufficient for certain scenarios. However, we also show that there are valid use cases, such that data ecosystems should *additionally* consider dependable guarantees that are as of now, not widely implemented.

The notion of dependability is commonly used in systems engineering to express the absence of failures; however, according to Anderson [14] there is also a security aspect: He considers "assurance" a critical component of dependability, which refers to the amount of reliance one can put in specific (security) mechanisms. According to that notion, we understand a *dependable* guarantee in the context of data ecosystems as a guarantee not building upon trust in another participant's or the platform's honesty regarding data handling. Thus, dependable guarantees enable the enforcement of policies between otherwise loosely-coupled and distrustful stakeholders, thereby enhancing data sovereignty.

We acknowledge that besides data security multiple aspects contribute to data sovereignty, including legal and architectural decisions [15]. However, we argue that the notion of dependable guarantees for data security, as established in this work, is a prerequisite for dependable data sovereignty in data ecosystems. Specifically, remote participants must be assured that inside attackers pose no threat to their data sovereignty. Thus, in this paper, we extend our preliminary analysis of the need for dependable security guarantees to enhance data sovereignty in data ecosystems [16]. Specifically, we discuss implementation costs and limitations dependable guarantees would necessarily introduce. Moreover, we further underpin our claim by systematically mapping and analyzing previous works on data ecosystems and related use cases. Namely, this work makes three contributions:

- We provide an understanding of the trade-off between data ecosystems as a tool to facilitate collaboration on and the exchange of data and the impact dependable data sovereignty guarantees would imply.
- We analyze and discuss three important and diverse use cases for data ecosystems: medical data records, supply chains, and a smart city application. Our analysis highlights further privacy and security concerns that are unresolved by current proposals for data ecosystems.
- We systematically map 2135 academic works focusing on data ecosystems concerning their implemented security principles and find 262 considering security or trust, but only 12 works make technical or cryptographic contributions toward improving dependable data security, data sovereignty, or trust. Our security analysis of these works reveals that only three of these works provide dependable guarantees that withstand inside attackers.

The remainder of this paper is structured as follows: First, in Section 2, we establish a common notion of current data ecosystem approaches, their architecture, and recent efforts of related work in securing data ecosystems. Based on our analysis, we derive three central design questions guiding data ecosystem security in Section 3. We complement our analysis with an outlook of possible building blocks for fixing these issues in Section 4. In Section 5, we introduce three use cases concerning personal, economic, and governmental data security. Next, in Section 6, we map academic efforts to improve data system security with a focus toward dependable guarantees. We discuss potential future research directions in Section 7, before concluding in Section 8.

2. A primer on current data ecosystems and their architectures

To ensure a common understanding of the security and trust issues with today's *data ecosystems*, we first briefly introduce data ecosystems and recapitulate the notion of data sovereignty and the role of typical participants in this context. Moreover, we give a short overview of data ecosystem initiatives focusing on their currently implemented security measures.

2.1. Overview

The need to share data with collaborators within specific sectors has been recognized in a variety of domains, including supply chains [17,18], public health [10,19], and mobility [20]. Here, on the one hand, data ecosystems aim to provide *multi-sided platforms* [2] that facilitate an automated data exchange following the *FAIR principle* [21], i.e., the offered data needs to be *findable*, *accessible*, *interoperable*, and *reusable*. On the other hand, today's data ecosystems aim to equip data owners with fine-grained

control over their data, including with whom and under what terms data may be shared. Enabling this fine-grained control is the foundation of *data sovereignty* [3]. The realization of fine-grained control over shareable FAIR date requires solving issues regarding organization [2], semantics and data quality [22], and interfacing [23], all of which are currently under active research.

So far, data ecosystems have primarily been seen as a means for exchanging data as required in emerging data markets and other use cases [3]. The paradigm of data ecosystems gradually emerged without a standard definition in mind, i.e., different accelerators of this paradigm came up with their own working definition of data ecosystems. Identifying this fragmented landscape of data ecosystem developments, Oliveira and Lóscio [24] reviewed concurrent definitions. They proposed a unified definition for data ecosystems based on their review. As a result, they define a data ecosystem as a *combination of independently operated networks* that *produce and provide data*, but also *other assets*, such as software or services [24]. Moreover, the authors highlight that such data ecosystems are *self-regulated* and *driven by collaboration and competition* between actors [24]. Extending this definition, we further emphasize that data ecosystems form platforms that have to define *standard interfaces and rules* to enable collaboration across independent networks. Accordingly, we consider entities that implement the interfaces and accept the rules defined by a given ecosystem data ecosystem as *participants*.

Likewise, the notion of *data sovereignty*, i.e., one of the fundamental motivations behind federated data ecosystems, currently lacks a clear and common definition [15]. If used in the context of data ecosystems, researchers generally agree that data sovereignty relates to the *control and ownership* of data items and the ability to specify, enforce, and trace specific claims on data [25–28]. Data sovereignty must not be confused with other types of notions of sovereignty, such as digital sovereignty or technical sovereignty [29]. Data ecosystems aim to fulfill these demands. Hence, this paper will focus on this aspect of data sovereignty as well as security and trust considerations connected to the enforcement of data sovereignty. Other aspects of data sovereignty, such as inclusive deliberation and the rights of data subjects [15], are as important as those mentioned above but require a societal solution in the first place.

2.2. Proposed data ecosystems and current initiatives

Several existing proposals shape the structure of data ecosystems. In the following, we introduce relevant approaches, both from theory and practice, that exemplify typical design decisions of data ecosystems.

Theoretical Work. To characterize data ecosystems, Azkan et al. [30] derive a morphology along the dimensions service ecosystem, service platform, and value co-creation. The authors emphasize data as a core resource and state that such systems are only sustainable if providers are intrinsically motivated to share data, typically to their own benefit. Gelhaar et al. [22] pick up the aspect of value creation and additionally emphasize trust between participants, a central problem of data ecosystems. Here, Zrenner et al. [7] derive an architecture for distributed usage control in business ecosystems from a use case in car manufacturing. Based on requirements such as fast load time, compatibility with existing systems, and transparency, the authors discuss an architectural solution for usage control [7]. Several other works consider data ecosystems tailored to specific needs, such as open data [31], financial data [32], or to describe the necessary data acquisition framework for machine learning tasks [33], that conceptually have a different focus, i.e., do not emphasize the distributed platform aspect, but consider data quality, quantity, and availability.

Practical Initiatives. Superseding a previously rather tedious bilateral exchange, the goal of initiatives like the International Data Spaces (IDS) [2,3,34], GAIA-X [5,23], Data Sharing Coalition [35], IHAN [36], FIWARE [37], CEF [38], or BDVA [39] is to establish a universal platform to regulate transactions regarding that exchange. The EU or federal offices fund such initiatives, facilitating a top-down approach toward establishing a common data platform [40]. This support recognizes the benefits of shared data, such as demonstrated by several recent efforts to combine previously distributed data silos [41].

Selected initiatives rather bundle forces toward adopting data ecosystems in general (Data Sharing Coalition, CEF, BDVA), while IHAN, for instance, is in an early stage, without publicly released technical documentation so far. Out of the named initiatives, IDS [4], GAIA-X [5], and FIWARE [42] have released technical documentation that enables a more profound analysis concerning what data security and trust measures are implemented in the respective proposals. Specifically, IDS and GAIA-X work toward a standard interface to locate and access data and provide an organizational context, including identification, admission, and certification of participants [23]. Thus, in the remainder of this paper, we primarily study these general-purpose initiatives. While IDS aims to provide a framework for data spaces to be built quickly, e.g., targeting a specific domain with coherent participants, GAIA-X plans to establish a single central cross-domain platform [23]. Moving toward domain-specific concepts, initial projects such as CATENA-X [43], an initiative inside the automotive domain, are picking up their ideas, while established platforms such as FIWARE [37], a framework to connect smart devices, start to provide compatible interfaces [44].

2.3. Key aspects of data ecosystem architectures

Despite their slightly different scopes, IDS and GAIA-X share a similar architecture. Thus, we analyze both initiatives together as a blueprint for data ecosystem implementations. To organize data exchange, data ecosystems commonly assign different roles to participants. Fig. 1 shows the overall scenario we are considering together with the main participants. Here, a single data exchange is considered bilateral, such that we can suppose the following roles [4]: First, a *data owner* legally owns the data to be shared and is interested in enforcing their rights on the data if it is shared. Second, a *data provider* takes over the technical part of offering a dataset to be exchanged on behalf of the data owner. While a single entity certainly can take over both roles, i.e., host the infrastructure to provide their data, the providing entity does not formally own the data in certain situations. For instance, this situation is the case for electronic health records owned by patients, which typically do not provide the infrastructure on their own. On the receiving

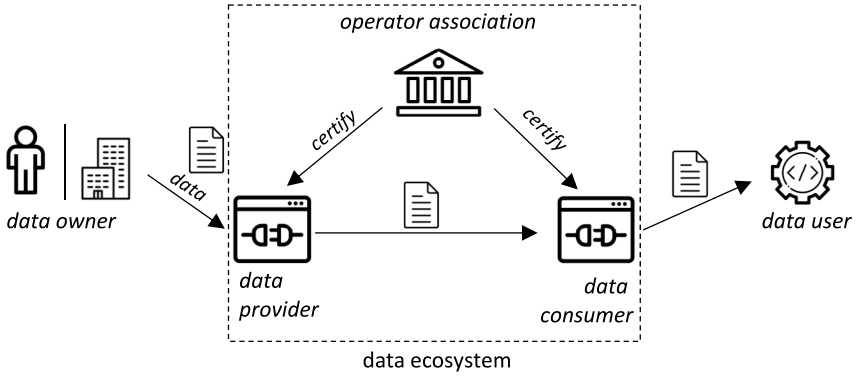


Fig. 1. Participating entities in data ecosystems. Data flows from left to right, with data provider and data consumer implementing a common ecosystem interface. The data ecosystem's operator also handles orthogonal tasks, including admission and discovery of participants and data.

side, a *data consumer* requests and receives the data from the provider and passes it to a *data user*, who processes the exchanged data, e.g., by visualizing it. Again, the consumer might fulfill the data user role if both processes are co-located. For instance, GAIA-X does not separate the data consumer and data user [5]. However, we continue using both terms to separate the logical roles, as described above.

Due to the distribution of providers and consumers, data ecosystems operate as a federation of independent deployments that jointly form a decentralized system. Thereby, data owners can keep their sensitive datasets under their control until they actively decide to share them with selected participants. To this end, data ecosystems enable data sovereignty until a data-sharing decision has been made and data is transferred to the data consumer.

Trust. To keep sovereignty from ending at the point of data exchange, data ecosystems currently require a certification of participants. Hence, they ensure that all entities handling data adhere to a common baseline regarding data protection. Certification includes but is not limited to, defense-in-depth strategies and security event monitoring systems [45,46]. Specifically, the IDS initiative requires prior certification steps. It attests successful certification via a public key infrastructure, establishing a trusted identity layer [4,23]. Contrarily, GAIA-X does not target a specific certification but requires participants to provide a standardized self-description with claims checked before admission [23]. In both cases, the ecosystem equips participants with the means to identify each other and establishes a common ground for mutual trust decisions.

Based on the ecosystem-wide identity layer, data ecosystems can provide fine-granular access control to data and let data owners limit the target audience they are willing to share their data with. However, more than access control is required, as data sovereignty would end once data flow between participants occurred after access was granted. *Usage control* [47] could fill this gap by granting specific rights on data and enforcing certain duties to be adhered to when processing data. Such a policy could be the permission to use a dataset for one week, with the obligation to delete it after that time.

To implement usage control, IDS utilizes and extends ODRL [48], a W3C standardized policy language for digital rights management [4]. Thereby, ODRL enables data owners and platform operators to define and communicate access and usage terms in a standardized fashion. As the only means of enforcement of the specified policies, the data owner has to trust that the consuming party abides by the negotiated terms and can only rely on the certification of the consumer required to join as a participant. However, since the negotiated contracts might also involve monetary compensation, the consuming party has incentives to disobey negotiated terms, e.g., using data more often than requested, sourcing it for other purposes, or sharing it with other systems or third parties.

Legal Context. Providing an environment for data exchange, the IDS builds upon surrounding legal contracts to equip participants with the means to establish credibility with each other [12]. Specifically, such contracts regulate the terms of usage and the overall setting, e.g., regarding a monetary compensation [4] or a penalty for breach of contract. Contracts can be bilateral or multilateral but will typically not cover the entirety of data space participants [4]. Because of this narrow consideration of available participants, current legal constructs centering around data ecosystems severely limit how spontaneous data consumers can access relevant data. Data ecosystems such as IDS then plan to (automatically) negotiate a refined technical contract within negotiated legal agreements. This refined contract translates terms into machine-readable policies that grant specific permissions on the exchanged dataset and potential obligations [4].

2.4. Data ecosystems in related work

Having established a common notion of data ecosystems, we now focus on notable recent research efforts concerning data ecosystems. Specifically, we provide an overview of *fundamental research* regarding the organization of data ecosystems, research efforts investigating the *use cases* that would benefit from data ecosystems, and works that *apply technical security measures* to facilitate data sharing efforts.

Fundamental Data Ecosystem Advancements. Oliveira and Lóscio survey and map the components data ecosystems typically comprise [24,49]. Furthermore, several works discuss requirements and possible ways toward implementing data ecosystems in

general, i.e., independent of specific initiatives [2,3,7,22,23,50]. Another line of research investigates fundamental challenges faced when implementing (distributed) data-sharing systems. Mainly, these challenges include transparency requirements [51], addressing the potential lack of trust between participants [22,52,53], the need for creating a common semantic understanding among all participants [13], and governance as well as legal constraints [12,54–56]. More directly targeted to data ecosystems as they are defined in this work, research considers alternatives to the current IDS and GAIA-X initiatives. For instance, FIWARE [37,44] provides a platform to facilitate data exchange in an Internet of Things context and is related to CEF [38]. Furthermore, special-purpose data ecosystems are being considered, e.g., by the NFDI initiative [57], which focuses on improving the accessibility of research data. Finally, NFDI and FIWARE aim to implement IDS-compatible interfaces, hence working toward ecosystem compatibility.

Use Cases. Another critical aspect of research on data ecosystems revolves around the use cases they are particularly well-suited for. Other works have identified many relevant or desirable use cases in this regard. Among these use cases are the sharing of medical health records [9,10], personal data [58], data emerging in the Industrial Internet of Things [59,60], and data exchange across supply chains, such as in the automotive industry [17,43,61], that have unique requirements concerning data confidentiality, data volume, or long-term persistency. Further data sharing schemes do not specifically target data ecosystems but are conceptually similar, such as applications in medicine [11,19,62,63], for production technology [64,65], along supply chains [17], or in education [66]. We expect that additional domains will also start to investigate the benefits data ecosystems can provide for their use cases as well as for society in general.

Technical Solutions for Data Sharing. Detached from data ecosystems, other research successfully applied technical and especially cryptographic building blocks to tackle the general challenges of data sharing in more narrow scenarios. For instance, Huang et al. [67] propose a data-sharing scheme to later identify sources of data breaches based on oblivious transfers and embedded fingerprints. Moreover, a variety of work considers sharing data with cloud providers [68–73], which can be considered conceptually similar to data ecosystems with multiple stakeholders. Such work includes querying encrypted data [74], attribute- or identity-based encryption for access control [61,63,75,76], and distributed ledgers together with TEEs to enforce accountability and access control [77]. Then again, Bonatti et al. [78] identify correctness and completeness as desirable properties of transparency mechanisms in data sharing. These approaches to strengthen sovereignty guarantees apply to real-world use cases and seem transferable for use in data ecosystems. However, given that these technical solutions originate detached from data ecosystems, the question arises to which end such solutions are already considered.

Security and Trust in Data Ecosystems. Schäfer et al. [79] identify suitable trust-enhancing technologies from expert interviews, including homomorphic encryption and trusted execution technology. Considering data markets in a more general sense, Liang et al. [80] survey digital rights management for privacy-preserving data trading, such as watermarking and software-based methods. Likewise, Garrido et al. [81] review privacy-enhancing technologies in the context of IoT data markets and identify handling copied data and multi-hop enforcement as problems not sufficiently covered in related work. Both of these surveys focus on data markets in general but do not specifically consider federated platforms with their unique (and challenging) threat model. To the best of our knowledge, no other related work systematically reviews the trust assumptions or security of *federated* data ecosystems, platforms, or markets.

2.5. Summary

While still under active research, academia, and large-scale initiatives shape the notion of data ecosystems as a means to simplify data exchange for value co-creation with data sovereignty as a key principle governing design decisions. They provide a common, distributed architecture that specifies interfaces and protocols and often also imply a certain legislative framework to establish trust, irrespective of the use case and domain at hand.

3. Data ecosystems need dependable guarantees

Based on the data ecosystems introduced in Section 2.2 to 2.4, we now critically review the design decisions of security mechanisms implemented in these data ecosystems. As, for instance, the EU facilitates establishing these platforms [40], we consider them relevant and worth studying concerning their dependable guarantees. Specifically, recent regulatory efforts set strict rules on how data may flow across organizational borders, raising, for instance, the need for fine-grained control [82]. To this end, data ecosystems are only sustainable if stakeholders are willing to participate by providing and consuming data actively.

At the same time, the current way of agreeing on terms for data exchange and usage incentivizes data-consuming parties to ignore these terms. Such behavior hurts data owners as they lose adequate compensation for the value of the data they provide and questions whether data ecosystems are adequate to exchange data subject to privacy regulation. Consequently, data owners might restrict their data-sharing efforts or leave the data ecosystem entirely. Hence, data ecosystems that plan to address these use cases need *dependable guarantees*, such as cryptographically enforceable guarantees and verifiable continual security monitoring, to establish trust between remote and potentially mutually unknown participants.

There exist proposals inside and outside academia (cf. Section 2); however, only IDS and GAIA-X currently provide the necessary level of detail in their documentation, allowing a detailed analysis. To this end, we limit our analysis to these proposals. Specifically, we analyze the available technical documentation and reference architecture for their implementation of adequate security and sovereignty measures. Primarily, we identify a lack of technical means to facilitate dependable security guarantees and establish strong participant trust. The current ecosystem initiatives can only partly address the security and trust requirements with their frail certification-based approaches.

Attacker Model. Guiding our position that data ecosystems require stronger data protection mechanisms, we apply the notion of a *malicious-but-cautions* attacker [83]. Specifically, the malicious-but-cautious attacker can misbehave in all possible ways but aims not to leave any verifiable evidence of its misbehavior [83]. Compared to an honest-but-curious (or semi-honest) attacker, this definition includes explicitly local deviation from protocols unless they are verifiable by externals. With data ecosystems exchanging data within established legal contracts, we argue that participants aim to avoid being sued for their misbehavior, and, hence, have incentives not to leave any evidence. To this end, a malicious-but-cautions attacker reflects the typical power and incentives of data ecosystem participants who source, process and utilize somebody else's data.

Data Security. Current notions of data security include security *at-rest*, *in-transit*, and *in-use* [84]. At-rest security and in-transit security are considered solved problems in data ecosystems as they can use widely available building blocks such as storage encryption and transport layer security (TLS), respectively [4]. Contrarily, in-use data security targets data at the moment of processing, e.g., when the decrypted data is loaded into memory, and is more difficult to ensure and implement. Technical or cryptographic measures to protect data by providing in-use security include, for instance, hardware-assisted security or homomorphic encryption [77,85]. However, despite these measures, today's data ecosystems build their guarantees regarding data in-use security upon remote participants' honesty to enforce certain rights on shared data. Unfortunately, incentives to evade enforcement clearly exist with monetary compensation handled as part of data exchange and transfers entrusted for a specific purpose.

Versatility and Cost. A central argument of recent initiatives beyond improved data sovereignty is their versatility regarding data types and participants, as well as a reasonably low implementation cost. To this end, there are a few restrictions and trade-offs to consider when implementing dependable guarantees. First, they try remaining agnostic regarding the data type and format, as long as it can be automatically exchanged [5]. Hence, data protection mechanisms should also be agnostic of the data type and must reflect different options for data usage, depending on its type. Second, they try to keep entry barriers low and, among others, explicitly target participants who have limited (computational) resources or economically cannot afford certified and trained IT professionals. To this end, the overhead of cryptographic means or the need for special hardware should be limited. Overall, mechanisms for dependable guarantees should reflect these efforts by maintaining versatility at reasonable costs.

Having established the need for considering inside attackers across different notions of data security, including data in-use, while maintaining the versatility of data ecosystems, we argue that the following questions are critical to the adoption of data ecosystem initiatives in data-sensitive domains:

I1: How can data owners trust remote infrastructure to enforce their granted rights once data has been shared?
I2: How can data owners track their data in a trusted way if processed by remote facilities?
I3: How can participants with little resources maintain sovereignty without requiring them to host their own infrastructure?

In the following, we elaborate on these high-level design questions regarding strong data sovereignty when implemented in practice.

I1**: Trust in Remote Rights Enforcement.** A first cornerstone of end-to-end data sovereignty is the guaranteed enforcement of digital rights on remote systems, i.e., *usage control*. However, suppose a privileged user on the consuming side, e.g., a system administrator, copies exchanged data without leaving traces in audit-relevant logging systems. This unintended behavior renders usage control enforcement ineffective. While we anticipate that such an action would violate negotiated terms, the data owner depends on fortunate coincidence to notice malicious behavior retrospectively. Consequently, we argue that data owners will refrain from using data ecosystems in their current form to share sensitive data. With such datasets covering manufacturing plans [17], the identity of suppliers [61], or privacy-sensitive health records [62] the lack of enforcement guarantees severely limits the kind of data exchangeable. Hence, such scenarios require stronger data sovereignty guarantees than the currently envisioned (weak) organizational measures.

Partly addressing this issue, IDS can utilize trusted platform modules (TPMs) as a trust anchor on remote systems [4]. However, merely providing verification of the running software, but essentially lacking memory encryption, TPMs still contribute little to effective protection against malicious-but-cautious attackers.

I2**: Trusted Data Usage Reporting.** Besides effective usage control, usage transparency is a second cornerstone to strong data sovereignty and essential to increase the participation of data owners. To this end, data owners who grant permissive access to their data shall still be able to track the usages of their data in remote systems transparently. Within IDS, a clearing house entity is designated to address part of this problem by enabling billing-relevant usage logging [4]. However, similarly to I1, IDS currently lacks a technically or cryptographically enforced guarantee that data usage must be logged. Hence, data users can easily circumvent the implemented logging features of today's data ecosystems and thereby exceed granted usage terms without being caught, such as evading downstream payments for data usage.

I3**: Sovereign Participation without Own Infrastructure.** A third cornerstone of strong data sovereignty is the free choice of data owners with whom to exchange data under which conditions. Within the currently proposed architecture (cf. Fig. 1), data owners entirely rely on and trust data providers to serve their data within the ecosystem. However, if both roles are distributed between separate entities, similar trust issues as between the providing and consuming parties also apply here. Specifically, the owner needs to trust the provider to serve the agreed policies and not misuse data locally. Moreover, usage reporting systems must not assume the provider to be trusted in this case. Hence, the providing side of a data exchange requires the same measures to implement reliable trust as the consumer side.

Takeaway. Today's data ecosystems only provide data protection via organizational means, such that they lack protection against malicious-but-cautious inside attackers on remote systems. At the same time, monetary data usage compensation and usage restrictions create incentives to evade enforcement mechanisms. Currently, these shortcomings limit the applicability of data ecosystems to share sensitive datasets irrespective of the domain or use case, and thus, need a remedy.

4. Toward dependable guarantees on security, sovereignty, and trust

The current data ecosystem initiatives strive to seamlessly interconnect businesses and facilitate the automation of valuable data exchanges. However, in the last section, we identified severe open issues (I1–I3) that impede each participant's data sovereignty in situations where organizational trust mechanisms, such as required certification prior to admission to the ecosystem, are insufficient. Given the competitive advantage a participant can gain by acting in a malicious-but-cautious manner, these open issues only become more pressing. Hence, with the data sovereignty of their participants in mind, data ecosystems must deploy additional means to allow them to establish trust in that new market.

In this paper, we argue that *only technical means providing strong cryptographic guarantees* are suitable to reach the goal of trustworthy and dependable data ecosystems that retain participants' data sovereignty in the presence of inside attackers. To this end, we now discuss how available building blocks can be integrated into data ecosystems to address each of the open issues I1–I3 while maintaining versatility and considering implementation costs.

4.1. Trusted remote policy enforcement (I1)

The foundation of strong data sovereignty in data ecosystems is providing data owners with an assurance that the data ecosystem will enforce terms and conditions on their behalf. Although today's data ecosystems lack trustworthy remote enforcement of data usage terms (I1), promising building blocks for addressing this issue are already available and used in other contexts. Examples of related building blocks are distributed usage control, trusted execution environments, and different cryptographic schemes. In the following, we discuss these building blocks, their application areas, and their relation to data ecosystems.

Distributed usage control [86–91] is an established field of research that focuses on modeling and technically enforcing usage terms, so-called *policies* for data usage. Data ecosystems have already adopted the notion of policies in their organizational architecture [4,92]. To this end, formulating and enforcing policies in providing and consuming connectors, while not trivial, mostly is a design and engineering challenge and basic implementations are available [93]. We expect a reasonably low cost for these mechanisms, and they only reduce versatility by limiting the data usage to applications or environments that can handle policy enforcement (e.g., via Windows group policies). However, *enforcing* these policies proves difficult as soon as the data owner cannot trust a data user and cannot directly observe the misconduct of a data user or the consequences thereof [94]. Hilty and Pretschner [87] hence propose to provide data owners with evidence of policy enforcement and restrict possible computations. Both approaches are hard to realize within a data ecosystem as they require some technical trust anchors on remote systems. Specifically, data ecosystems currently do not offer such trust anchors as the data user gains full control over the exchanged data once it has been obtained from the data owner. This situation is insufficient when considering, for instance, a malicious-but-cautious adversary who does not provide a trustworthy environment for storing or processing the exchanged data.

Hardware-based Trusted Execution Environments (TEEs), such as Intel SGX, AMD SEV-SNP, or ARM TrustZone, are promising candidates for closing this gap in the future [95]. The goal of TEEs is to provide a trustworthy computing environment that can be established even on untrusted remote infrastructure. To this end, a TEE provides an isolated (i.e., memory-encrypted) environment for running applications with the ability to verify the integrity of the executed program code remotely. A CPU-embedded cryptographic key provides the required trust anchor that allows the data owner to verify correct execution independently of the remote host's operating system [95]. Recent TEEs, such as AMD SEV-SNP allow protecting complete virtual machines (VMs) with reasonably low overhead [96]. Compared to earlier TEEs, such as Intel SGX, VMs even enable trusted graphical applications, such that they maintain versatility if applied within data ecosystems. Consequently, TEEs can help trustworthy remote execution by hiding the program's execution state and hardening it against hampering.

Implementing policy enforcement and data processing inside such environments has the potential to resolve the trust issues data ecosystems are currently facing. However, TEE technology is an active field of research, and current implementations still experience security issues [97]. For example, today's TEE implementations are prone to side-channel attacks that allow for limited data extraction [98]. Countermeasures such as oblivious RAM [99] are being investigated to fix these vulnerabilities, and we expect that future enclave designs will provide further remedies against other technical issues as they are being discovered. Hence, TEEs are a promising building block for improving data sovereignty in data ecosystems via technically enforceable data policies. However, further research into hardening TEEs against unintended security breaches is required to improve their applicability to data ecosystems. In fact, in a related context, a first work [77] demonstrates the applicability of TEEs in a trusted data-sharing setting.

We thus call for the established initiatives and researchers to further investigate the utility of TEE technology for data ecosystems to reliably address the lack of trustworthy and technically backed policy enforcement.

4.2. Verifiable data tracking (I2)

Besides policy enforcement, establishing transparency in data usage is equally important to gain data owners' trust. For instance, a data owner might consider granting generous accessibility to their data but require proper attribution by any data user. In such a case, the data owner would profit from technically guaranteed notifications whenever a data user accessed the data.

Currently, IDS implements a clearing house instance, which can log data usage if mandated in a policy, making it transparent to data owners [4]. However, data users have neither a strict technical constraint to log data usage, nor can the system enforce it by

some means. Consequently, IDS cannot currently provide trusted monitoring unless data usage can be observed externally. Hence, the current clearing house instance does not solve the problem of verifiable data tracking (I2).

Instead, technical or cryptographic means would help to incentivize logging. To this end, we consider transparency logging, data-flow tracking, and distributed ledger technology promising for establishing verifiable data tracking in data ecosystems. For instance, certificate transparency logging allows modern web browsers to reject digital certificates that are not tracked in a public log for auditors to verify [83]. It has shown to be a scalable and cost-effective tool for public verifiability in the context of PKI. A similar approach might improve data usage transparency as well [100]. Namely, cryptographically tying the decryption of exchanged data or the transfer of results to a publicly verifiable log entry would force data users to log their actions accurately. As long as log entries are small, only a few distributed log instances could handle the workload of millions of transactions, such as shown by the PKI example [101]. Similar approaches are also being researched in the field of verifiable computing [102,103] and data ecosystems could profit by utilizing corresponding building blocks.

Besides logging, related work also proposes data flow tracking [104] and data fingerprinting [105] to allow for identifying the source of identified data breaches after the fact. However, the cryptographic data fingerprints required to apply these techniques necessitate knowledge of the exact data representation and a sufficient tolerance for minor statistical noise in the monitored data [105]. Thereby, these methods are limited in their versatility, which depends on the concrete data to be shared. Unfortunately, these fingerprints typically cannot survive intermediate processing steps [105], rendering them inapplicable in some situations. Hence, more research on resilient data flow tracking or fingerprinting techniques is required to determine and improve their applicability in the context of data ecosystems.

Finally, distributed ledger technology has emerged in recent years with the explicit goal of facilitating digital interactions among participants who do not fully trust each other. While Bitcoin started by establishing a decentralized and publicly accessible digital currency based on a blockchain [106], it spawned more versatile distributed ledgers for any information using smart contracts [107]. Ultimately, business-focused ledger systems emerged, such as Hyperledger Fabric or Quorum. These architectures can facilitate the event-logging within data ecosystems and provide a medium for the automated billing of data accesses.

To avoid additional privacy or data confidentiality problems, such transparency mechanisms need to take privacy into account, e.g., by encrypting log entries [108]. Overall, technical building blocks for verifiable data tracking are already available. However, they still need to be tailored to the specific verifiable data tracking requirements for utilization in data ecosystems regarding performance, scalability, flexibility, and privacy.

4.3. Integration of resource-constrained participants (I3)

With the separation between the data provider and data owner, data ecosystems also address scenarios that involve particularly resource-constrained or especially privacy-aware data owners who are unable or unwilling to run the complete infrastructure themselves. However, infrastructure control is the foundation of self-sovereign participation in distributed environments [4]. Hence, this approach is not viable for resource-constrained participants. Such participants could be, for instance, small to mid-sized enterprises (SMEs) in a supply chain context, which have no technical expertise to provide the infrastructure to participate in a data ecosystem. In this case, their customers may be capable of assuming the role of a data provider collecting data from their contracted SMEs and offering that data on their behalf within the ecosystem. For instance, large automotive manufacturers can assume the role of a data provider on behalf of their, typically numerous, suppliers [17]. In this case, however, data owners lose their sovereignty and depend on trust in their customers. Thus, appropriate dependable guarantees for such situations are desirable.

A scenario that would assure data owners that their data is treated as intended would be considering the data provider as a different party than the data owner; however, current ecosystem initiatives do not rigorously satisfy this demand [4]. Under this assumption, however, one could implement the same measures discussed in Section 4.1 also on the provider side, i.e., realize a trusted data provider. Moreover, concerning usage transparency, this scenario requires logs, as discussed in Section 4.2, to be accessible with no own infrastructure. Hence, not only the consumer-side aspect of logging must be trusted, but also the instance that provides logging on behalf of data owners.

Resource constraints of SMEs not only affect computational resources, such as for hosting a connector acting as a data provider or data consumer. Another constraint can be human resources or their required training and level of certification, e.g., to obtain ISO 27001 certification [46]. To this end, dependable guarantees might be able to render the need for such certification unnecessary. Given that, for instance, TEEs are widely available on modern CPUs by default, we argue that they can be a viable alternative to certification.

4.4. Summary

Cryptographic building blocks that have been successfully applied in the past are also promising to address the core issues (I1–I3) currently impeding the data sovereignty of data owners in today's data ecosystems. For instance, TEEs have the potential to provide the currently missing trust anchor during remote processing (I1). Similarly, concepts currently applied in the context of certificate transparency logging or distributed ledger technology may help satisfy the requirement for verifiable tracking in data ecosystems (I2) once they are adapted to the scalability demands of envisioned deployments. Finally, these measures can also potentially be applied when data providers operate on behalf of the original data owner to incorporate resource-constrained participants in the process (I3).

Table 1
Summary of main data sovereignty challenges within the presented use cases.

Use case	Personal data	Economic data	Governmental data
Main incentive	Broader data basis	Simplify data logistics	Better utilize existing data
Key concern	Patient privacy	IP protection	Personal privacy
Key requirements	Consent, revocation, traceability	Confidentiality, authenticity, multi-hop dependencies	Integration of constrained participants, unclear data owners
Trusted enforcement (I1)	◑	●	◑
Trusted traceability (I2)	●	●	◑
Participation (I3)	●	◑	●

●: fully applies, ◑: considered, ○: not relevant

5. Exemplary case studies highlight the need for dependable guarantees

In the following, we highlight the findings of our analysis in the context of three real-world applications concerning personal data (), economic data (), and governmental data (). Despite their diverse origin, each application accentuates our proclaimed need for dependable security measures. To this end, we give an overview of three unique use cases, namely data in health research (Section 5.1), supply chains (Section 5.2), and smart cities (Section 5.3). All these use cases have already been approached using data ecosystems. We then compare their unique requirements concerning data security and sovereignty in Table 1 and discuss how each of them relates to the core issues that we identified in Section 4.

5.1. Sharing health data for research

A frequent application of data ecosystems is to share health data for research purposes. In the following, we discuss the example of the UK Biobank (UKB) [109] as one data ecosystem that is already established. The UKB stores and processes personal information of significant sensitivity, and has clear governance principles concerning data security. The UKB is a large-scale model cohort study and biomedical database that provides comprehensive data access for health-related research in the public interest. It contains detailed health and genetic information of more than 500 000 participants aged between 40 and 69, enabling research based on this unique data set. Between 2006 and 2010, participants gave informed consent for long-term health monitoring, genotyping, and linkage to medical records. They provided blood, urine, and saliva samples for future analysis [110] and detailed information about themselves in 22 assessment centers throughout the UK.

Since 2012, researchers can apply to use the UKB for their studies, and the UKB is openly accessible to any bona fide researcher who wishes to use it to conduct health-related research for the benefit of the public. More than 80 000 research groups from industry and academia worldwide can access the data ecosystem or have applied for access. The development of this resource has involved extensive consultation, input from scientific, managerial, legal, and ethical partners, and centralized processes on an industrial scale [111]. The UKB database is expected to grow to 15 petabytes over the next five years [112]. However, while the data infrastructure's impact on the environment is widely recognized, data sovereignty is a topic yet to be discussed [113].

Still, current workflows to secure data when shared include an ethical and governance framework and access procedures for the scientific community, the general public, and other stakeholders [114]. Precisely, the UKB removes personally identifiable information before sharing data with others for research purposes. Data users must follow a reviewed application process, ensuring research legitimacy using the data in public interest [114]. In addition, data users must sign a legal declaration not to try identifying individuals. Here, reidentification is a major problem. For instance, predicate singling out attacks can allow to re-identify individuals in pseudonymized datasets with few traits, such as a rough estimation of location and age [115]. Many other cohort studies have followed UKB, using the same methodology (e.g., "All of us cohort" [116]). As such, the UKB is a prime example of the importance of sovereignty in data ecosystems.

Moreover, the UKB has attempted to address access to data in a practical and non-bureaucratic manner for researchers. The UKB has implemented a fair, transparent, and efficient online access process [117]. The aim is to ensure that the resource and its access arrangements are widely communicated to the scientific community and to provide open access to its data and samples for health-related research in the public interest. Only if the proposed research is in the public interest and the required data are or will be available will requests for data be approved. Additionally, researchers will only be provided with de-identified data and must sign a material transfer agreement. In this agreement, they confirm not to attempt to identify participants, to keep the data secure, and to use it only for the approved research. However, UKB does not employ measures that would render, e.g., participant identification, impossible, and past research has shown that solely relying on the concept of k-anonymity is insufficient [118,119]. Thus, despite the need for technical concepts, currently, there are no technical limitations that prohibit copying, sharing, or using data for other than approved purposes.

The UKB's robust data protection framework sets out the fundamental principles, rights, and obligations for processing personal data [114]. As a responsible custodian of participant data and samples, UKB is legally obligated to ensure that they are stored,

retrieved, and used securely with appropriate technical and organizational measures [120]. To this end, UKB must also take reasonable steps to protect any data or samples it shares with others. Thus, researchers who access UKB data must implement security measures to ensure that the data is processed and used securely and in a compliant manner. These measures should restrict access to authorized users and protect against unauthorized access by internal or external parties [109]. However, UKB has no means to verify or restrict data usage once shared, nor are there viable technical means, e.g., if patients want to withdraw given consent. Most UKB data are available for download, i.e., without control. Only enormous datasets (e.g., whole genome sequencing) are accessible through the research platform itself only, enabling some means to assess data usage [121].

Participants can withdraw their consent or exercise a right to erasure at any point in time [122]. The participant withdrawal form, however, only covers future data usage. Here, UKB argues that researchers only receive a de-identified copy of data, such that data used in ongoing and completed studies are not subject to the GDPR's right to erasure [122]. However, it remains unclear whether the de-identification process is dependable, as research on its reliability is explicitly forbidden by the data usage agreement. To this end, mechanisms for privacy-preserving tracing of data usage (I2) could at least yield a list of pointers to studies relying on data for which consent shall be withdrawn.

Another often-faced criticism is that the UKB does not generally provide feedback on individual-level results derived from data obtained during assessments [109] to respect the informed consent given by participants, i.e., lacks privacy-preserving data provenance and data lineage mechanisms that would allow selective feedback channels upon the patients wish. Solving this problem, e.g., by giving participants a more active role in such a data ecosystem would automatically allow feedback. Currently, participants receive a summary of measures at the end of each assessment visit. They are encouraged to see a doctor if their results are outside the normal range. Direct feedback creates an incentive for participation, eventually increasing the database size. A broader data basis would allow for more input data, e.g., to classifiers trained on that database, but also solve the potential genetic bias UKB might face, as UKB primarily focuses on upper-class white individuals. In the long term, this would enable fairer and more accurate models.

At the same time, stricter guarantees on how the covered personal data is shared and can be analyzed increase the sovereignty of participants and, hence, might reduce the threshold for participation or even enable future studies to utilize health data on a broad scale. Still, establishing the UKB provided valuable insights into the essential components needed to create large-scale studies, including efficient governance, centralized infrastructure, extensive research data, and widespread public support [123]. As such, UKB shows how broad data access in a permissive, voluntary setting and with explicit patient consent enables medical research.

Overall, the UKB highlights the specific need for personal data in medical contexts, including consent and traceability. The UKB consent scheme, i.e., giving broad consent, on the one hand, enables data availability for research purposes. On the other hand, it limits the data basis to those patients willing to give such broad consent despite considerable privacy impacts and the current lack of feedback options or the personal benefit of participation. Concerning our analysis, we argue that traceability (I2) is inherently lacking, and participation on the patient level (I3) is currently limited to giving consent once without a continual sovereign decision beyond the capability to revoke that consent. Patients then entrust UKB with handling their data accordingly and UKB-authorized researchers with not evading differential privacy applied to shared datasets. Hence, dependable guarantees concerning the remote enforcement of granted access and usage rights (I1) is relevant as well.

5.2. Supply-chain data management in digital production

Today's digitized production and globalized supply chains pose another stimulating use case for which data ecosystems are being discussed. In fact, IDS has origins in this domain as several requirements and architectural components originate in a preceding project named industrial data space [124]. The key motivation for data protection here is protecting intellectual property and maintaining competitive advantages instead of personal data protection [125,126]. With physically distributed and organizationally separated suppliers and consumers, supply chain applications match the scenario of federated data ecosystems [61]. Moreover, data exchange and the ability to authentically track and trace goods along supply chains today is considered a cornerstone of modern industry [17]. Relevant data in this context includes, but is not limited to, operations, performance optimization, billing, and logistics [17]. For instance, financial data, trade secrets, or construction plans are subject to intellectual property claims. Hence, data security, especially authenticity and confidentiality, is essential to secure modern supply chains to prevent unauthorized access, data breaches, and cyber-attacks.

The data exchange process in supply chains involves multiple parties, such as suppliers, manufacturers, distributors, retailers, and customers. At the same time, relationships between these parties can be short-lived and not necessarily built upon prior mutual trust [17]. Furthermore, supply chains are inherently intransparent. For instance, involved parties typically only know their direct trade partners, i.e., whom they buy parts from and to whom they sell their goods to, as suppliers and prices are a well-kept secret [17]. Still, specific information needs trustworthy mechanisms to propagate along supply chains, such as resource and environmental certificates [127].

Relationships within supply chains can also build on general distrust, such as between direct competitors that share common production facilities. In this context, companies are reluctant to share data and prefer to protect their intellectual property, also coined as "company privacy" [126]. Proposed data ecosystems, including IDS [4], currently try to solve this problem. However, the concerns regarding the organizational security approach fully apply here as the current design requires companies to entrust competitors with the enforcement of data access terms to their own disadvantage (cf. Issue I1).

Without standardized data ecosystems, data exchange happens via various methods, such as email and customer relationship or supply chain management systems. Initiatives, such as CATENA-X, prepare to supersede these methods [43]. In addition to the

above-mentioned problems, these initiatives face some specific additional challenges: For instance, ownership of data is already a complicated endeavor as it is left unclear by law [15]. However, within supply chains, it must also be trackable who owns derivative data that builds on a prior data exchange [128]. Such requirements are not limited to data ownership. Authenticity and authorization must also cover derivate data, e.g., the German supply chain law aims to enforce human rights in global supply chains and imposes significant fines if necessary certificates are missing [129]. Hence, companies are interested that data passed along supply chains remains authentic.

In conclusion, data security is critical in supply chain applications that involve exchanging sensitive and confidential information. Efforts centering around current initiatives (cf. Section 2) reflect the need to simplify data logistics within supply chains and also to secure them accordingly. Ensuring the security of data exchanged in supply chains protects data's confidentiality, availability, and authenticity. To this end, we emphasize that trust in remote rights enforcement (I1) and trusted traceability (I2) require a solution for data ecosystems to be viable in this context. Given that companies of different sizes, including small and midsized enterprises, are often an essential part of supply chains, we argue that their sovereign participation (I3) must also be considered.

5.3. Smart city applications

Besides personal and economic data, governmental data poses another application area for data ecosystems. Several unique challenges require a solution to implement data sovereignty in the context of governmental data and governmental data use cases often fit the scenario of data ecosystems. Applications range from open data as a relatively permissive scenario that, however, still can have traceability and revocation requirements [31], to rather sensitive use cases, such as demographic data on an individual level [31]. At the same time, especially to combat climate change and deal with the change in demographics, the analysis of rich data sources promises the potential for significant insights [130]. In the following, we emphasize two use cases of data ecosystems centering around smart city applications, which despite their public origin face several severe challenges in implementing data sovereignty. To underline that dependable guarantees are not only a problem of sectors involving personal or economic data, we highlight two of these directions, namely smart grids and smart public transport, in more detail.

Generally, Kirimtat et al. [131] characterize smart cities as consisting of several themes, including, for instance, smart heating, smart transport, and smart grids. We refer to prior work for a more detailed analysis [131] but emphasize that all these applications build upon derivates from personal data, e.g., collected at municipal utilities, public transport, or in private homes. Often, the data itself was not collected for applications in smart cities in the first place but was intended for other purposes, such as billing or operational monitoring. Such secondary use then poses privacy and data sovereignty questions. For instance, local electrical power or district heating grids gain significant relevance and are in the process of digitization [132]. Given that data collection happens at the level of households, i.e., individuals, data sovereignty becomes relevant here as well: Existing grid meters, which first serve billing purposes, now also provide presence detection, i.e., are a proxy for personal data [133].

Mapping such a use case to the scenario that we emphasized in Section 2 introduces a few unique challenges: The data provider often is the local grid operator. However, whether the data owner should also be the grid operator or the data subject, i.e., the residents of the household being monitored, remains unclear. Customers to the grid operator should ideally be capable of specifying with whom and under what granularity they want to share data. However, grid operators are also a stakeholder in this decision. As the central goal is to gain more insights into the grid, allowing better utilization, service delivery, and eventually a reduction of costs, grid operators have the incentive to pursue integration. Solving these problems with dependable guarantees concerning the (sensitive) data provided by residents might resolve some of their concerns against rolling out monitoring systems on a broad scale [134].

Another frequent use case within smart cities covers public transport systems [135]. These systems heavily rely on collecting, processing, and exchanging vast amounts of data, frequently including personal information about passengers, such as their travel patterns, payment details, and geolocation data, i.e., data that passengers aim to keep under their control [135]. At the same time, public transport systems are often a consortium of different actors, e.g., partitioned by means of transportation [136] and affected by mutual trust issues (cf. I1). These actors already exchange data but keep passengers out of the loop, who hence cannot control or audit the usage of their data (cf. I2). Additionally, using sensors and other data collection technologies can raise concerns about data sovereignty, particularly if this data is being shared with third-party providers or used for commercial purposes without passenger consent.

Dependable guarantees to solve the discussed issues with governmental data are manifold. For instance, grid meter readings are reasonably small to be protected by cryptographic means, such as HE, and we anticipate most processing to happen automatically, such that an implementation of dependable guarantees would not limit the versatility for that use case. TEEs, on the other hand, could help consortia of public transport providers to reduce their interdependencies when working with personal data. With their general availability on (at least) server hardware (cf. Section 4), limited overhead can be expected. Efforts toward embedding TEEs in mobile devices, e.g., via ARM TrustZone, might complement these directions and enable dependable guarantees even for shared ticket machines.

In conclusion, implementing data ecosystems in governmental applications, such as power grids and public transport, can provide several benefits, including increased efficiency, cost reduction, improved service delivery, and sustainability. At the same time, re-purposing already collected data, such as in the power grid, raises new data sovereignty requirements, which a data ecosystem must handle. Concerning the identified issues, we deem the sovereign integration of smart grid customers and public transport passengers an important issue (cf. I3), but we further note that trusted enforcement (I1) and traceability (I2) also apply here.

5.4. Summary

Table 1 summarizes the key findings from each presented use case. While incentives for applying data ecosystems to the introduced use cases and the key requirements regarding data security and sovereignty are diverse, our identified issues are relevant to all of them. In Section 4, we already introduced several building blocks to mitigate I1–I3. Given the specific requirements e.g., regarding data volume and computational effort, research on personal health data might rather implement a performant technical solution to I1, such as via TEEs, whereas data in the context of smart grids, such as meter readings, can be aggregated via cryptographic means. Similarly, the design of traceability (I2) and participation (I3) building blocks depend on the use case, and hence, need a tailored solution. In the following, we relate the findings from our use case description to the issues we identified in this paper and set out future directions for research to mitigate this situation.

6. Systematically mapping security features in federated data ecosystems

In this paper, we have detailed our position that current data ecosystems are not well-prepared to address the needs of sovereignty in federated environments, such as those targeted by the discussed data ecosystems. At the same time, such dependable guarantees would help to increase data ecosystems' applicability to common use cases, as we discussed in Section 5. We now underpin our position by systematically mapping academic works in the context of data ecosystems concerning the issues I1 to I3.

We follow the guidelines of Petersen et al. [137] to conduct a systematic mapping study on the security assumptions and technical building blocks for security, sovereignty, and trust employed by academia. Compared to an in-depth literature review, a mapping study employs broader inclusion criteria and intends to map security considerations rather than synthesize study results. For a more general overview of data ecosystems, we refer to the study from Oliveira et al. [49], who have mapped related work on data ecosystems in a general sense, but did not review any security aspects.

To quantify the proclaimed lack of strong security approaches (cf. Section 3), we assess prior work in the context of federated data ecosystems and data markets concerning their contributions toward secure and trustworthy data exchange and processing. We base our analysis on the Issues I1 (trusted enforcement), I2 (usage transparency) and I3 (sovereign participation) that we have established in Section 3. Dependable guarantees for these issues exist, as outlined in Section 4, but it remains unclear whether academia implements any of them. Hence, we study the adoption of those dependable guarantees in the following and also map academia's notion of security and trust in data ecosystems in a more general sense based on the following research questions:

RQ1 To which end does prior work related to data ecosystems consider security and trust issues generally?

RQ2 What kind of security contributions are made?

RQ3 Which building block, technology, or concept provides dependable security features, if any?

These research questions provide an orthogonal view on the issues we identified in Section 3: While I1 to I3 each highlight needed security features, RQ1 to RQ3 focus on different levels of security considerations in related work, i.e., RQ1 considers the general notion of security, RQ2 focuses on the specific security features, and RQ3 on the technical realization of these features. This gradation allows us to differentiate between different levels of technical readiness regarding the stated issues. To answer these research questions, we first describe the methodology of our systematic mapping study with its methodology regarding the inclusion and exclusion of related work, before discussing RQ1–RQ3 in detail.

Methodology and Research Corpus. Focusing on a broad coverage with our study, we include works that contain at least one of the terms "data ecosystem", "data sovereignty", "data market" or "data marketplace" in their title, abstract, or keywords. While the latter two search terms are not directly related to data ecosystems, we consider them closely related to the idea of what has more recently been coined as a data ecosystem. Thus, these works might contribute dependable guarantees in a scenario applicable to our study. We query Scopus and Web of Science for results matching either of our keywords in June 2023. Overall, our search yields 2993 papers containing at least one of our search terms.

From the query result set, we excluded studies that match at least one of the following exclusion criteria (in that order, Table 2(b) details each criterium):

EC1 The work is duplicated (based on title and authors) or redundant (e.g., full conference proceedings).

EC2 The work is not written in English language.

EC3 The work is not peer-reviewed (e.g., keynote abstracts, editorials, presentation slides, etc.).

EC4 The work does not mention at least one of the keywords listed in Table 2(b) in its title, abstract, or keywords.

EC5 The study does not consider security or trust in the context of data ecosystems.

For EC1 to EC3, we rely on the output from Web of Science and Scopus to exclude studies automatically. EC4 to EC5 help to narrow the scope to papers relevant to our research questions. To this end, EC4 filters studies based on additional keywords relevant to our research questions. We validated the chosen set of keywords by randomly selecting a subset of 1/4 of the studies excluded by EC4 where we manually reviewed abstracts. This step did not reveal any additional relevant studies that would have been falsely excluded.

EC5 was manually assessed by a single author, based on title, abstract, and keywords, and if in doubt, also based on the full text. Here, we do not assess works qualitatively, but categorize them based on their contribution and their applicability to data ecosystems. We argue that this categorization is bias-free, but want to briefly cover the most frequent causes of exclusions to support

Table 2
Quantitative overview of the works discovered during our literature review, together with their categorization.

(a) Evaluation of exclusion criteria

	Category	# Papers
	Total	*1047 (WoS) + 1946 (Scopus)*
EC1	Unique papers	2135
EC2	English Language	2117
EC3	Peer Reviewed	1958
EC4	Matched keywords (b)	773
EC5	Security contribution to DE	262

(b) Keyword matches

Keyword	# Matches (after EC4)	# Included (after EC5)
privacy	355	110
secur{e, ity}	326	142
trust	218	97
blockchain	211	124
transparency	92	23
sensitive	78	34
cryptographic	42	25
confidential	42	23
gdpr	31	10
certif*	18	7
verifiable	7	4

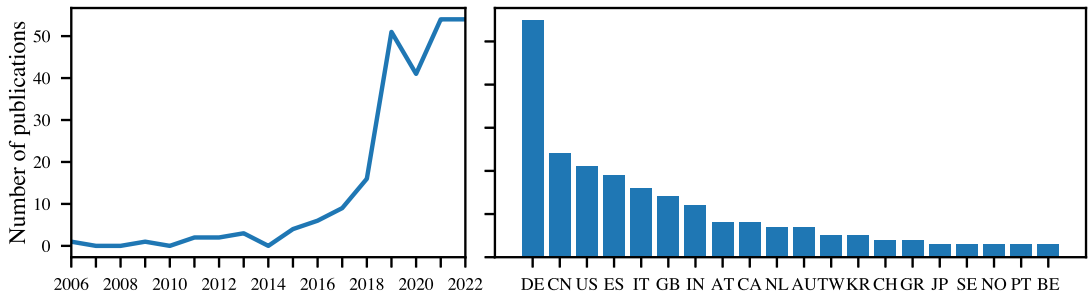


Fig. 2. Distribution of the 262 publications in scope of our study, based on the year of publication and affiliation of the first author.

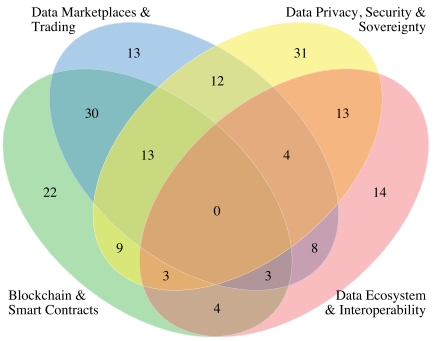
reproducibility. Specifically, we exclude works that, despite mentioning at least one of the chosen keywords, do not contribute to the scope of our review. Frequent excluded topics concern the governance and (structural) organization of data ecosystems or markets (64 articles), ethics and legal concepts related to data sovereignty (57 articles), or pricing models for private data and economic considerations of data sharing (41 articles). Additionally, 20 articles focus on the data sovereignty of indigenous people. We argue that a central aspect of data ecosystems is to differentiate between data owners and data users, and thus exclude works that do not differentiate between these roles. Besides that, EC5 does not exclude architectures not considering a platform for initiation of data exchanges.

RQ1: **To which end does prior work related to data ecosystems consider security and trust issues generally?** Considering the exclusion criteria gives a first quantitative impression of the role security-relevant research plays in the context of data ecosystems. Out of the 1958 peer-reviewed studies after deduplication, 773 match at least one of the keywords specified in Table 2(b). However, our manual review of these 773 studies to filter for EC5 shows that a relatively large fraction of these works (66 %), despite mentioning a keyword, do not further focus on a contribution to enhance data ecosystem security. A frequent storyline in these excluded studies is that federated platforms enhance security, privacy, or trust through their architecture. Moreover, studies mention the advancements in the context of blockchain as a (valid) motivation to research data platforms. Another focal point of investigation are works that outline requirements, challenges, and future research directions, most often quite generally and not specifically geared toward the scope of our study. Overall, after filtering EC5, we consider 262 studies with a title or abstract that claims to make a relevant contribution relating to data platforms, such as federated data ecosystems as defined in Section 2.

Fig. 2 shows the distribution of the included studies based on their publication year and the geographic location of the first author's first affiliation. The topics have gained momentum from 2017 to 2018, with a significant increase and a relatively stable publication rate since then. The introduction of stricter privacy legislation around that time, e.g., GDPR, likely contributed to this increase. The geographic distribution shows that research is conducted worldwide, although with an emphasis on EU member states, as 50 % of studies originate there. Further, the authors that have contributed to the included studies constitute a diverse group: 889 unique authors have on average authored 1.16 of the 262 considered articles. No author has published more than 6 of the included articles.

Blockchain & Smart Contracts	#	Data Markets & Trading	#	Data Priv., Sec. & Sovereignty	#	Data Ecosystem & Interop.	#
blockchain	77	data marketplace	30	data sovereignty	30	cloud	6
smart contracts	17	data sharing	17	security	15	data ecosystems	5
smart contract	15	data market	11	privacy	13	int. data spaces	5
dist. ledger tech.	5	data markets	7	data protection	6	trust	5
ethereum	4	data trading	7	access control	5	data governance	4
reputation	4	data marketp.	6	cloud computing	5	data space	4
		data exchange	5	gdpr	4	ind. data space	4
		open data	5	usage control	4	interoperability	4
		data monet.	4	data security	3	architecture	3

(a) Most frequent author keywords per category with the number of papers mentioning a keyword.



(b) Topical combination of author keywords.

Fig. 3. Topical distribution of included papers based on the provided author keywords.

To get a first intuition of topics considered by the authors of the studies we include for our analysis, we source the list of author-provided keywords. Here, we identified four general directions, as shown in Fig. 3(a) , to which we attribute the most frequent keywords. Fig. 3(b) shows the topical combination of these general directions. Here, we find that works centering around data ecosystems (red category) seem to consider data security and privacy more frequently (59 %) than works in the data market category (33 %), although the latter covers more works in total numbers (49 vs. 83). However, given that the most frequent keywords shown in Fig. 3(a) remain at a rather superficial level, we instead rely on manual assessment to investigate contributions, as detailed in the following.

RQ2**: What kind of security contributions are made?** Having quantified security considerations within research related to data ecosystems, we now assess what kind of contributions are made and if any of them relate to dependable guarantees. To understand what directions studies take, Fig. 4 categorizes studies by their contributions into five major categories, which we define as follows:

- **Conceptual.** These studies focus on contributing an architecture, (network) model, or governance strategies geared toward building secure data markets. They do not directly propose dependable mechanisms but contribute to their conceptual foundation.
- **Empirical.** These works perform an empirical approach to research security within data ecosystems via surveys, literature reviews, or expert interviews. They contribute summaries, derive security requirements, or outline future research directions.
- **Organizational.** These studies propose mechanisms and concepts that aim at enforcing certain data sovereignty properties, such as via access or usage control, or distributed computing concepts, such as federated learning. These mechanisms found their guarantees upon the infrastructure's properties, such as their distributed architecture, or the presence of a standard middleware, such as policy enforcement proxies, that aim at enforcing data sovereignty.
- **Technical.** Similar to the organizational category, these works contribute mechanisms to enforce certain data sovereignty properties. However, these studies do not settle their guarantees in the infrastructure's properties but employ cryptographic or technical means to enforce data sovereignty (cf. Section 4).
- **Blockchain.** We dedicate a separate category to blockchain-based contributions, which center around transparency, reputation, billing, or access control. Technically, blockchain technology can provide dependable guarantees for these objectives. However, unless processing happens entirely on-chain or within smart contracts (with limited expressiveness and a poor fit to data ecosystem architectures), we observe these works still relying on some (organizational) means for usage control (cf. Issue I1). Hence, there are fewer guarantees than for the dependable category.

With the diversity of contributions shown in Fig. 4, we note that data sovereignty and mechanisms to establish data sovereignty are topics academia is well aware of. However, the share of organizational contributions compared to technical ones highlights that problems centering around trusted infrastructure, as it is required in the organizational setting, and problems related to inside attackers have not been widely picked up. The application of blockchains for data markets and data ecosystems, on the other hand, seems to be widely researched. Here, frequent notions investigate the trading of data (63 articles), verification of data ownership (16 articles), or access control (13 articles). However, as argued above, these mechanisms contribute little, for instance, to Issue I1 and hence cannot provide dependable data sovereignty.

Given that the utilized building blocks incur certain implementation costs and overheads (cf. Section 4), we will now focus on the 12 studies making dependable contributions and review their considered threat model, limitations and applicability.

RQ3**: Which building blocks, technologies, or concepts provide the claimed guarantees?** So far, we have advocated for employing technical guarantees for federated data ecosystems and have shown that only a small fraction of 4.6 % of included works already provides such guarantees. We will now detail how these guarantees are implemented, at which cost, and in which threat model. To this end, Table 3 shows the 12 works that contribute dependable guarantees according to our mapping study. They differ in the applied building blocks, implemented guarantees, and their versatility concerning data processing and its cost, which we detail in the following.

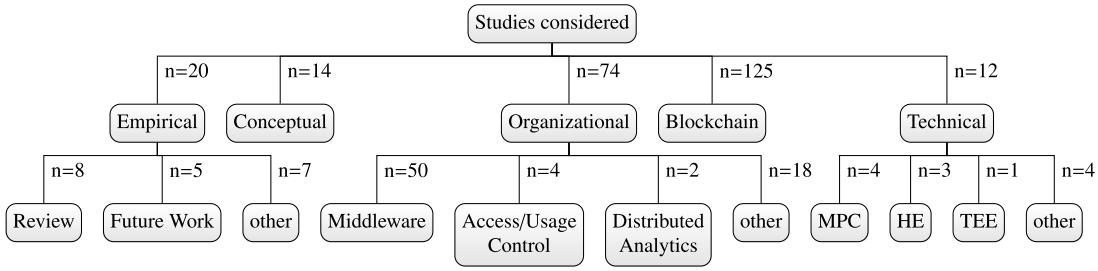


Fig. 4. Categorization of included papers based on their contributions.

Table 3
Overview of studies with dependable guarantees from our systematic mapping study with their mechanism, attacker model, and the complexity and costs that data access, processing, and sharing incur. The use technologies are secure multiparty computation (MPC), homomorphic encryption (HE), trusted execution environment (TEE), smart contracts (SC), zero-knowledge proofs (ZKP), Attribute-based encryption (ABE), and proxy re-encryption (PRE).

Reference	Mechanism	Considered scenario & attacker model			Versatility and cost
		Data Owner	Platform	Data User	
T1 [138]	MPC	trusted	malicious	malicious	costly computation, shares computation result only, requires computation nodes
T2 [139]	HE	trusted	malicious	honest but curious	limited applicability (quadratic polynomials)
T3 [140]	TEE, SC, HE	malicious	trusted	malicious	quite general computability, HW dependency and 128MB memory limit
T4 [141]	MPC, ZKP	trusted	at least one honest node	malicious	limited versatility, requires computation nodes
T5 [142]	MPC	trusted	at least one honest node	malicious	simple statistics, requires computation nodes
T6 [143]	collusion detection	trusted, can verify platform	malicious but cautious	malicious but cautious	limit matching of offerings
T7 [144]	HE	trusted	malicious	honest but curious	limited precision and complexity
T8 [145]	secret sharing	trusted	at least one honest node	trusted	sharing fragmented images over cloud
T9 [146]	trusted platform modules	trusted	(partly) trusted	(partly) trusted	cheap, but boot-time attestation with limited guarantees
T10 [147]	HE	malicious (authentication and integrity guaranteed)	semi-honest, non-collusion	not considered	evaluation of polynomial functions
T11 [148]	ABE, PRE	trusted	no data access, entrusted with access control	trusted once access granted	versatile (plaintext access), costly asym. data encryption
T12 [149]	MPC, HE	trusted	malicious	malicious	simple statistics (Pearson correlation), requires computation nodes

First, the utilized building blocks are diverse, i.e., all building blocks we identified in Section 4 found application in the context of federated data ecosystems. Homomorphic encryption (HE) and secure multiparty computation (MPC) have been applied most frequently. The shown works implementing HE-based solutions let the data owner encrypt some dataset that is subsequently shared with the platform (T2, T3, T7, T10, T12). The following design decision regarding where to compute with the shared data decides whether data users can be malicious: If the platform conducts computation and the data user only receives a result (T1, T3, T4, T5, T12), they can be untrusted or even malicious. Otherwise, the data owner needs to trust the data user with adhering to the agreed policy concerning the handling of the shared data, as the latter technically can conduct arbitrary calculations (T2, T7). While practical limitations of the applied HE schemes limit their utility (e.g., T2 can only evaluate quadratic polynomials) this setting still restricts the data owner in its sovereignty.

The shown MPC-based solutions are less affected by the problems with HE, as calculations can only be conducted jointly. In this case, the data owner's sovereignty is assured if one of the computing parties is honest and does not collaborate (e.g., T4–T5 set this threshold to three parties). As for HE, the utility is limited (e.g., T12 computes a simple correlation function) and costly, as multiple parties must jointly compute on ciphertext, which introduces additional communication overhead [149]. Additionally, MPC is not a perfect fit for the scenario of federated data ecosystems, as it remains open to which parties should jointly compute during a

typically bilateral data exchange (cf. Fig. 1). Thus, MPC would likely incur additional costs to compensate others for contributing computing capacities—a necessity federated data ecosystems typically want to avoid.

T3 circumvents the downsides of the above approaches by using a TEE as a trusted third party to conduct computations. However, in their scheme, the data owner and user must trust the platform with the attestation of enclaves that execute smart contracts. We argue that this is not a conceptual limitation, as one could build multiple hops of attestation from enclave to enclave, but future work will need to further investigate this issue.

T1–T12 not only differ in their used building blocks but also in their considered threat model. Most works assume a trusted data owner that aims to enforce their own sovereignty demands, such as data protection or monetary compensation, in the presence of a malicious platform and/or an honest but curious or malicious data owner. T3 and T10 deviate from this scheme by assuming a trusted data user or platform that aims at protecting from malicious data owners.

In T9, Oliver et al. [150] evaluate the usage of Trusted Platform Modules (TPMs) to ensure trust in remote clouds. However, the authors conclude that more than a continual enforcement boot-time attestation is needed to ensure trust. Likewise, in T11, Esposito et al. [148] restrict data flows in a smart city application via location-dependent cryptography. However, their work is limited to spatial access control without enforcing certain guarantees during data usage.

When relating T1–T12 to the issues described in Section 3, we find these works provide some initial dependable guarantees. For instance, all works that consider a malicious data user essentially solve I1 as they do not rely on the data user's trustworthiness. However, based on the discussed cost and limitations, these solutions do not yet provide a practical solution to the problem, due to limited utility. They also require additional infrastructure, such as computation nodes, which might not be available in all scenarios. Concerning transparency (I2), T4 and T5 build transparent verification of computed results on top of MPC, such that the data owner can remain informed about future data usage. Still, none of the surveyed works consider long-term data availability or allow transparency for offline data users. Lastly, we were unable to identify any explicit consideration of constrained data owners (I3) in T1–T12. Based on our assessment, we argue that T9 might provide an applicable solution to this issue, as trusted platform modules are only required for the platform and data user, whereas the data owner only needs to perform attestation. The other approaches at least require the data owner to prepare data in a certain format, e.g., for MPC. Overall, these limitations underline that the state of the art does not yet provide a practical solution to the important problem of dependable guarantees in federated data ecosystems, significantly hindering their practical utility and widespread use.

Summary. Comparing the quantitative results from our mapping study with the research direction in the area of privacy-enhancing technologies (cf. Section 2.4) yields a torn picture. On the one hand, dependable technical and cryptographic methods that can improve sovereignty in data ecosystems by contributing to I1–I3 exist, but on the other hand, most current works in this context do not (yet) utilize these methods. Given the relatively high number of papers contributing to organizational measures as well as papers with empirical or conceptual contributions, we argue that there is a lack of dependable guarantees. Related work frequently implements corresponding concepts, indicating that organizational measures are favored over cryptographic or technical methods that withhold inside attackers. Our analysis of those works that contribute technical methods yields, that only half of the identified approaches withhold inside attackers, whereas the rest still requires some trust in other participants. In the following, we elaborate on why research centering around data ecosystems should instead focus on dependable guarantees, i.e., consider inside attacks a valid threat and work toward more dependable security, sovereignty, and trust measures.

7. Discussion and future research directions

As highlighted in Section 3, today's data ecosystems mostly rely on organizational means to implement data protection. However, the analyzed use cases (cf. Section 5) and also ongoing research efforts (cf. Section 2.4) indicate that suitable applications of data ecosystems include the handling of privacy-sensitive data, such as patient records in medical contexts, but also confidentiality demands of critical business data require those guarantees. Technical building blocks are already available to address the remaining challenges for data sovereignty in data ecosystems by providing stronger guarantees for participants (cf. Section 4), but our mapping study (cf. Section 6) shows that they have not yet been widely picked up in the context of data ecosystems. Additionally, only six of twelve works with technical contributions withstand inside attackers. Yet, we argue that data ecosystems must provide a framework that allows users contributing sensitive data to trust the system in enforcing their rights at any time, including processing in remote systems after access was granted and data was shared.

Based on our analysis of the status quo as well as ongoing research efforts so far, we discuss in the following that overcoming current *shortcomings of usage control* is a crucial research direction to sustainably strengthen the data sovereignty for participants of data ecosystems. We recommend looking into *hardware-based security* measures as one potential mechanism to adopt dependable guarantees in data ecosystems, before concluding with a summary of our lessons learned from the systematic mapping study.

Shortcomings of Usage Control. With (distributed) usage control, prior work already addresses the issues I1–I3 today's data ecosystems are facing. However, the enforcement has not (yet) been thoroughly picked up by recent initiatives, possibly due to the current lack of technical guarantees [94]. Most work in this area either targets rights modeling (e.g., [86,151,152]) or assumes operation on trusted infrastructure (e.g., [153,154]), which we argue does not withstand malicious-but-cautions attackers, as applicable to data ecosystems. Given that guaranteed policy enforcement is crucial for sharing sensitive datasets within data ecosystems, this question still needs to be addressed to allow for widespread adoption of data ecosystems.

With cryptographic and technical solutions, the ways toward stronger guarantees are at least two-fold and not straightforward. The discussed cryptographic approaches toward stronger guarantees, i.e., providing usage control and transparency via cryptographic means, implement the strongest protection among the discussed techniques but currently either allow only limited expressiveness

or suffer from a severe performance penalty. Hence, we argue that they are currently not suited for general application in data ecosystems but should be selectively applied for the most sensitive datasets, where the named limitations and overheads are acceptable [62].

Need for Hardware-based Security. Hardware-based solutions provide a trust anchor under the malicious-but-cautious attacker model. Moreover, they are less affected by performance penalties and eventually allow the same operations as standard hardware. However, TPMs, as currently envisaged by the IDS [4], cannot provide adequate protection of sensitive data due to the lacking memory encryption. Hence, TEEs, despite current known side-channel attacks and related weaknesses, seem to be a better choice for strong guarantees regarding data sovereignty expanding to remote systems.

With hardware-based TEEs being available for a few years, the question arises as to why today's data ecosystems do not yet implement TEE-based security. One reason might be known weaknesses, which need to be addressed in future designs. However, these weaknesses do not seem to hinder deployment in further applications, as, for instance, Microsoft Azure offers commercial support for TEEs in its cloud service [155]. Hence, we argue that data ecosystems should consider employing TEEs as a measure to enforce data owners' rights on remote infrastructure, which would fill the current gap toward implementing end-to-end data sovereignty.

Future Research Directions. These required research efforts motivate our call for future work in the domain of data ecosystems. Regarding the reliable enforcement of usage terms (I1), future work must address tailoring existing data protection schemes to data ecosystems. Here, a promising idea seems to employ TEEs as a trust anchor on remote infrastructure. However, further research must clarify to which degree current limitations, such as performance penalties, affect application within data ecosystems. Subsequently, this can be integrated with transparency mechanisms (I2) where current work demonstrates the applicability of cryptographic mechanisms, e.g., in certificate transparency. To this end, further research must investigate how these concepts can support transparency in data ecosystems, while not creating new privacy issues. The combination of technically enforceable usage control with usage transparency might also be the first step toward sovereign integration of resource-constrained participants (I3).

Impact on Data Ecosystem Architectures. Extended usage control policies and their enforcement, e.g., via hardware could provide one way of revising current data ecosystem architectures to provide dependable guarantees of data security and sovereignty. Besides the above-outlined technical challenges, future work thus also needs to investigate how these mechanisms align with current data ecosystem architectures and their environments, e.g., concerning stakeholders, requirements, and business models. Finally, the legislative aspect of sharing data in data ecosystems demands investigation. The current contractual overhead conflicts with efforts to automate and simplify data exchange. The building blocks introduced in this work render several parts of contractual agreements redundant from a technical viewpoint, but therefore, they first need legal recognition.

Lessons learned from our Systematic Mapping Study. While our mapping study confirmed our initial intuition that dependable security guarantees for federated data ecosystems are still in their infancy, we also learned several lessons from the study, which we summarize in the following: First, we identify a lack of a common understanding of the term data ecosystem, despite existing definitions [24]. In some cases, we discover relevant approaches under the keywords "data market" or "data marketplace", but not under "data ecosystem". We further observe that all three terms are frequently used interchangeably in other contexts, e.g., smart contracts in blockchains [139]. Second, related work seems to convey that the notion of federated architectures generally providing security and data sovereignty is widely accepted, as we find it across several diverse publications [3,7,12,23]. If any, the justification mostly refers to the fact that participants retain control over their infrastructure. However, we have shown that federated architectures do not yet provide dependable guarantees, but need to be complemented by technical means. Third, we find research within the scope of our study scattered across several communities, including information systems, distributed systems, and cryptography. We would argue that due to the diversity of the research area, such a distribution is not surprising. Mixing expertise from different areas and domains is likely to be beneficial for future research.

8. Conclusion

Today's data ecosystems facilitate an automated exchange of data in a standardized manner while simultaneously providing access to massive, heterogeneous, and previously isolated data sources. Given that these data exchanges and corresponding higher-level applications across domains (e.g., in the automotive industry or for governmental services) also frequently deal with sensitive information, including business secrets and data subject to privacy regulations, data ecosystems must implement reliable measures to prevent any undesirable exposure of sensitive data. Currently, this need is recognized and claimed to be addressed but primarily based on organizational means. We argue that such means fail to provide sufficient guarantees in settings with malicious-but-cautious participants, i.e., participants who aim to remain unnoticed while still trying to infer all possible information from the data ecosystem and associated data exchanges.

At the same time, today's data ecosystems lack appropriate guarantees regarding confidential processing on systems operated by third parties, transparency of data access and usage, and the participation of parties with no infrastructure under their control (I1–I3). Our systematic mapping study quantifies this lack, i.e., only 0.6 % of works matching our keywords provide dependable guarantees. At the same time, 39 % of works related to data ecosystems mention security or trust aspects (RQ1), but only 262 works (13 %) focus on contributions related to these aspects. Here, contributions related to distributed ledger technologies (125) and organizational security (74) are most frequent (RQ2), while we count only 12 works with technical security contributions (RQ3). These numbers underline the above-criticized lack of dependable guarantees: The corresponding building blocks are available [138–140,145] but are seldomly applied to implement dependable guarantees for data ecosystems.

Based on our analysis of use cases concerning personal, economic, and governmental data, we conclude that our identified issues generally affect all of them, although there are distinct incentives for dependable guarantees. To this end, the success of data ecosystems directly depends on their ability to address the present need for dependable data sovereignty of participants. As such, modern technical solutions, such as TEEs, promise to provide data owners with trustworthy guarantees of correct data handling, increasing their willingness to participate in available data ecosystems.

CRediT authorship contribution statement

Johannes Lohmöller: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jan Pennekamp:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Roman Matzutt:** Methodology, Supervision, Writing – original draft, Writing – review & editing. **Carolin Victoria Schneider:** Conceptualization, Writing – original draft, Writing – review & editing. **Eduard Vlad:** Investigation, Validation. **Christian Trautwein:** Supervision. **Klaus Wehrle:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612 and the Excellence Strategy of the Federal Government and the Länder.

References

[1] J. Pennekamp, R. Glebke, M. Henze, T. Meisen, C. Quix, R. Hai, L. Gleim, P. Niemietz, M. Rudack, S. Knape, A. Epple, D. Trauth, U. Vroomen, T. Bergs, C. Brecher, A. Buhrig-Polaczek, M. Jarke, K. Wehrle, Towards an infrastructure enabling the internet of production, in: 2019 IEEE International Conference on Industrial Cyber Physical Systems, (ICPS), IEEE, Taipei, Taiwan, 2019, pp. 31–37, http://dx.doi.org/10.1109/ICPHYS.2019.8780276.

[2] B. Otto, M. Jarke, Designing a multi-sided data platform: Findings from the international data spaces case, Electron Markets 29 (4) (2019) 561–580, http://dx.doi.org/10.1007/s12525-019-00362-x.

[3] B. Otto, Interview with reinhold achatz on data sovereignty and data ecosystems, Bus. Inf. Syst. Eng. 61 (5) (2019) 635–636, http://dx.doi.org/10.1007/s12599-019-00609-z.

[4] B. Otto, S. Steinbuss, A. Teuscher, S. Bader, et al., IDS reference architecture model (version 4.0), 2022.

[5] Gaia-X Technical Committee, Gaia-X architecture document, 2021.

[6] J. Gelhaar, T. Groß, B. Otto, A taxonomy for data ecosystems, in: Hawaii International Conference on System Sciences, 2021, http://dx.doi.org/10.24251/HICSS.2021.739.

[7] J. Zrenner, F.O. Möller, C. Jung, A. Eitel, B. Otto, Usage control architecture options for data sovereignty in business ecosystems, JEIM 32 (3) (2019) 477–495, http://dx.doi.org/10.1108/JEIM-03-2018-0058.

[8] A. Ibrahim, T. Dimitrakos, Towards collaborative security approaches based on the European digital sovereignty ecosystem, in: Collaborative Approaches for Cyber Security in Cyber-Physical Systems, Springer International Publishing, Cham, 2023, pp. 123–144, http://dx.doi.org/10.1007/978-3-031-16088-2_6.

[9] J. Scheibner, J.L. Raisaro, J.R. Troncoso-Pastoriza, M. Ienca, J. Fellay, E. Vayena, J.-P. Hubaux, Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis, J. Med. Internet Res. 23 (2) (2021) e25120, http://dx.doi.org/10.2196/25120.

[10] A. Appenzeller, S. Bartholomaus, R. Breitschwerdt, C. Claussen, S. Geisler, T. Hartz, P. Kachel, E. Krempel, S. Robert, S.R. Zeissig, Towards distributed healthcare systems – virtual data pooling between cancer registries as backbone of care and research, in: 2021 IEEE/ACS 18th International Conference on Computer Systems and Applications, (AICCSA), IEEE, Tangier, Morocco, 2021, pp. 1–8, http://dx.doi.org/10.1109/AICCSA53542.2021.9686918.

[11] D. Froelicher, P. Egger, J.S. Sousa, J.L. Raisaro, Z. Huang, C. Mouchet, B. Ford, J.-P. Hubaux, Unlynx: A decentralized system for privacy-conscious data sharing, Proc. Privacy Enhancing Technol. 2017 (4) (2017) 232–250, http://dx.doi.org/10.1515/popets-2017-0047.

[12] A. Duisberg, Legal aspects of IDS: data sovereignty - what does it imply? in: Designing Data Spaces, Springer, 2022.

[13] S. Bader, J. Pullmann, C. Mader, S. Tramp, C. Quix, A.W. Müller, H. Akyürek, M. Böckmann, B.T. Imbusch, J. Lipp, S. Geisler, C. Lange, The international data spaces information model – an ontology for sovereign exchange of digital content, in: J.Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, L. Kagal (Eds.), The Semantic Web – ISWC 2020, vol. 12507, Springer International Publishing, Cham, 2020, pp. 176–192, http://dx.doi.org/10.1007/978-3-030-62466-8_12.

[14] R. Anderson, Security Engineering: A Guide to Building Dependable Distributed Systems, 1st Edition, Wiley, 2020, http://dx.doi.org/10.1002/9781119644682.

[15] P. Hummel, M. Braun, M. Tretter, P. Dabrock, Data sovereignty: A review, Big Data Soc. 8 (1) (2021) 205395172098201, http://dx.doi.org/10.1177/2053951720982012.

[16] J. Lohmöller, J. Pennekamp, R. Matzutt, K. Wehrle, On the need for strong sovereignty in data ecosystems, in: Proceedings of the 1st International Workshop on Data Ecosystems, (DEco '22), vol. 3306, CEUR-WS, Sydney, Australia, 2022, pp. 51–63.

[17] L. Bader, J. Pennekamp, R. Matzutt, D. Hedderich, M. Kowalski, V. Lücken, K. Wehrle, Blockchain-based privacy preservation for supply chains supporting lightweight multi-hop information accountability, Inf. Process. Manage. 58 (3) (2021) 102529, http://dx.doi.org/10.1016/j.ipm.2021.102529.

[18] J. Pennekamp, R. Matzutt, C. Klinkmüller, L. Bader, M. Serror, E. Wagner, S. Malik, M. Spiß, J. Rahn, T. Gürpinar, E. Vlad, S.J.J. Leemans, S.S. Kanhere, V. Stich, K. Wehrle, An interdisciplinary survey on information flows in supply chains, ACM Comput. Surv. 56 (2) (2024) 1–38, http://dx.doi.org/10.1145/3606693.

[19] H. Ma, R. Zhang, G. Yang, Z. Song, K. He, Y. Xiao, Efficient fine-grained data sharing mechanism for electronic medical record systems with mobile devices, IEEE Trans. Dependable Secure Comput. 17 (5) (2020) 1026–1038, http://dx.doi.org/10.1109/TDSC.2018.2844814.
[20] Z. Du, C. Wu, T. Yoshinaga, K.-L.A. Yau, Y. Ji, J. Li, Federated learning for vehicular internet of things: recent advances and open issues, IEEE Open J. Comput. Soc. 1 (2020) 45–61, http://dx.doi.org/10.1109/OJCS.2020.2992630.
[21] M.D. Wilkinson, et al., The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 160018, http://dx.doi.org/10.1038/sdata.2016.18.
[22] J. Gelhaar, B. Otto, Challenges in the emergence of data ecosystems, in: Pacific Asia Conference on Information Systems, (PACIS), Dubai, 2020.
[23] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to European digital sovereignty with Gaia-X and IDSA, IEEE Netw. 35 (2) (2021) 4–5, http://dx.doi.org/10.1109/MNET.2021.9387709.
[24] M.I.S. Oliveira, B.F. Lóscio, What is a data ecosystem? in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, ACM, Delft the Netherlands, 2018, pp. 1–9, http://dx.doi.org/10.1145/3209281.3209335.
[25] M. Schanzenbach, Towards Self-sovereign, Decentralized Personal Data Sharing and Identity Management (Ph.D. thesis), 2020.
[26] V. Pedreira, D. Barros, P. Pinto, A review of attacks, vulnerabilities, and defenses in industry 4.0 with new challenges on data sovereignty ahead, Sensors 21 (15) (2021) 5189, http://dx.doi.org/10.3390/s21155189.
[27] S. Couture, S. Toupin, What does the notion of sovereignty mean when referring to the digital? New Media Soc. 21 (10) (2019) 2305–2322, http://dx.doi.org/10.1177/1461444819865984.
[28] K. Irion, Government cloud computing and national data sovereignty: government cloud computing and national data sovereignty, POI 4 (3–4) (2012) 40–71, http://dx.doi.org/10.1002/poi3.10.
[29] M. Hellmeier, F. von Scherenberg, A delimitation of data sovereignty from digital and technological sovereignty, in: ECIS 2023 Research Papers, vol. 306, 2023.
[30] C. Azkan, F. Möller, L. Meisel, B. Otto, Service dominant logic perspective on data ecosystems-a case study based morphology., in: Proceedings of the 28th European Conference on Information Systems, (ECIS), 2020.
[31] D. Lee, Building an open data ecosystem: An Irish experience, in: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, ACM, Guimaraes Portugal, 2014, pp. 351–360, http://dx.doi.org/10.1145/2691195.2691258.
[32] C. Moiso, R. Minerva, Towards a user-centric personal data ecosystem the role of the bank of individuals' data, in: 2012 16th International Conference on Intelligence in Next Generation Networks, IEEE, Berlin, Germany, 2012, pp. 202–209, http://dx.doi.org/10.1109/ICIN.2012.6376027.
[33] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, Y. Liu, A global manufacturing big data ecosystem for fault detection in predictive maintenance, IEEE Trans. Ind. Inform. 16 (1) (2020) 183–192, http://dx.doi.org/10.1109/TII.2019.2915846.
[34] S.R. Bader, M. Maleshkova, SOLIOT—decentralized data control and interactions for IoT, Future Internet 12 (6) (2020) 105, http://dx.doi.org/10.3390/fi12060105.
[35] https://datasharingcoalition.eu/about-the-data-sharing-coalition/. Accessed 9 August 2022, 2022.
[36] https://ihan.fi/. Accessed 9 August 2022, 2022.
[37] F. Cirillo, G. Solmaz, E.L. Berz, M. Bauer, B. Cheng, E. Kovacs, A standard-based open source IoT Platform: FIWARE, IEEE Internet Things M. 2 (3) (2019) 12–18, http://dx.doi.org/10.1109/IOTM.0001.1800022.
[38] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/CEF+Digital+Home. Accessed 9 August 2022, 2022.
[39] https://www.bdva.eu/. Accessed 9 August 2022, 2022.
[40] M. Palviainen, J. Suksi, Data marketplace research: A review of the state-of-the-art with a focus on smart cities and on edge data exchange and trade, in: 2023 Smart City Symposium Prague, (SCSP), IEEE, Prague, Czech Republic, 2023, pp. 1–7, http://dx.doi.org/10.1109/SCSP58044.2023.10146227.
[41] T. Kariotis, M.P. Ball, B. Greshake Tzovaras, S. Dennis, T. Sahama, C. Johnston, H. Almond, A. Borda, Emerging health data platforms: from individual control to collective data governance, Data Policy 2 (2020) e13, http://dx.doi.org/10.1017/dap.2020.14.
[42] ETSI GR CIM 007 V1.1.1: Security and Privacy, Tech. rep., France, 2022.
[43] O. Voß, Catena-X: datenstandards für die autobranche, tagesspiegel background digitalisierung & KI, 2021.
[44] Á. Alonso, A. Pozo, J. Cantera, F. de la Vega, J. Hierro, Industrial data space architecture implementation using FIWARE, Sensors 18 (7) (2018) 2226, http://dx.doi.org/10.3390/s18072226.
[45] N. Menz, A. Resetko, B. Otto, Framework for the IDS Certification Scheme 2.0, Tech. rep., IDSA, 2019, http://dx.doi.org/10.5281/ZENODO.5244858.
[46] https://www.iso.org/isoiec-27001-information-security.html. Accessed 16 February 2023, 2023.
[47] A. Pretschner, M. Hilty, F. Schütz, C. Schaefer, T. Walter, Usage control enforcement: present and future, IEEE Secur. Privacy Mag. 6 (4) (2008) 44–53, http://dx.doi.org/10.1109/MSP.2008.101.
[48] R. Ianella, Open digital rights language (ODRL), in: Open Content Licensing: Cultivating the Creative Commons, 2007.
[49] M.I.S. Oliveira, G.D.F. Barros Lima, B. Farias Lóscio, Investigations into data ecosystems: A systematic mapping study, Knowl. Inf. Syst. 61 (2) (2019) 589–630, http://dx.doi.org/10.1007/s10115-018-1323-6.
[50] M. Henze, M. Grossfengels, M. Koprowski, K. Wehrle, Towards data handling requirements-aware cloud computing, in: 2013 IEEE 5th International Conference on Cloud Computing Technology and Science, IEEE, Bristol, United Kingdom, 2013, pp. 266–269, http://dx.doi.org/10.1109/CloudCom.2013.145.
[51] S. Geisler, M.-E. Vidal, C. Cappiello, B.F. Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto, E. Paja, B. Pernici, J. Rehof, Knowledge-driven data ecosystems toward data transparency, J. Data Inf. Qual. 14 (1) (2022) 1–12, http://dx.doi.org/10.1145/3467022.
[52] A. Munoz-Arcentales, S. López-Pernas, A. Pozo, Á. Alonso, J. Salvachúa, G. Huecas, An architecture for providing data usage and access control in data sharing ecosystems, Procedia Comput. Sci. 160 (2019) 590–597, http://dx.doi.org/10.1016/j.procs.2019.11.042.
[53] M. Huber, S. Wessel, G. Brost, N. Menz, Building trust in data spaces, Designing Data Spaces, Springer, 2022, http://dx.doi.org/10.1007/978-3-030-93975-5_9.
[54] C. Ducuing, Data as infrastructure? A study of data sharing legal regimes, Compet. Regul. Netw. Ind. 21 (2) (2020) 124–142, http://dx.doi.org/10.1177/1783591719895390.
[55] D. Wu, S.G. Verhulst, A. Pentland, T. Avila, K. Finch, A. Gupta, How data governance technologies can democratize data sharing for community well-being, Data Policy 3 (2021) e14, http://dx.doi.org/10.1017/dap.2021.13.
[56] L. Helminger, C. Rechberger, Multi-party computation in the GDPR, in: Privacy Symposium 2022 - Data Protection Law International Convergence and Compliance with Innovative Technologies, (DPLICIT), 2022, http://dx.doi.org/10.1007/978-3-031-09901-4_2.
[57] N.L. Weisweiler, R. Bertelmann, P. Braesicke, T. Bronger, C. Curdt, F.O. Glöckner, S. Rank, O. Stegle, Y. Sure-Vetter, N. Villacorta, Helmholtz Open Science Briefing: Helmholtz in der Nationalen Forschungsdateninfrastruktur (NFDI): Report des Helmholtz Open Science Forums, Tech. rep., Helmholtz Open Science Office, 2021, http://dx.doi.org/10.48440/OS.HELMHOLTZ.030.
[58] R. Matzutt, D. Müllmann, E.-M. Zeissig, C. Horst, K. Kasugai, S. Lidynia, S. Wieninger, J.H. Ziegeldorf, G. Gudergan, I.S. gen. Döhmann, K. Wehrle, M. Ziefle, myneData: towards a trusted and user-controlled ecosystem for sharing personal data, 2017, http://dx.doi.org/10.18420/IN2017_109.
[59] H. Baars, A. Tank, P. Weber, H.-G. Kemper, H. Lasi, B. Pedell, Cooperative approaches to data sharing and analysis for industrial internet of things ecosystems, Appl. Sci. 11 (16) (2021) 7547, http://dx.doi.org/10.3390/app11167547.

[60] A.L. Marra, F. Martinelli, P. Mori, A. Saracino, A distributed usage control framework for industrial internet of things, in: C. Alcaraz (Ed.), Security and Privacy Trends in the Industrial Internet of Things, Springer International Publishing, Cham, 2019, pp. 115–135, http://dx.doi.org/10.1007/978-3-030-12330-7_6.

[61] S. Malik, N. Gupta, V. Dedeoglu, S.S. Kanhere, R. Jurdak, TradeChain: decoupling traceability and identity in blockchain enabled supply chains, in: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications, (TrustCom), 2021, pp. 1141–1152, http://dx.doi.org/10.1109/TrustCom53373.2021.00155.

[62] D. Froelicher, J.R. Troncoso-Pastoriza, J.L. Raisaro, M.A. Cuendet, J.S. Sousa, H. Cho, B. Berger, J. Fellay, J.-P. Hubaux, Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption, Bioinformatics (2021) http://dx.doi.org/10.1101/2021.02.24.432489, Preprint.

[63] X. Lu, X. Cheng, A secure and lightweight data sharing scheme for internet of medical things, IEEE Access 8 (2020) 5022–5030, http://dx.doi.org/10.1109/ACCESS.2019.2962729.

[64] J. Pennekamp, E. Buchholz, Y. Lockner, M. Dahlmanns, T. Xi, M. Fey, C. Brecher, C. Hopmann, K. Wehrle, Privacy-preserving production process parameter exchange, in: Annual Computer Security Applications Conference, ACM, Austin USA, 2020, pp. 510–525, http://dx.doi.org/10.1145/3427228.3427248.

[65] S. Mangel, L. Gleim, J. Pennekamp, K. Wehrle, S. Decker, Data reliability and trustworthiness through digital transmission contracts, in: The Semantic Web, vol. 12731, Springer International Publishing, Cham, 2021, pp. 265–283, http://dx.doi.org/10.1007/978-3-030-77385-4_16.

[66] R. Matzutt, J. Pennekamp, K. Wehrle, A secure and practical decentralized ecosystem for shareable education material, in: 2020 International Conference on Information Networking, (ICOIN), IEEE, Barcelona, Spain, 2020, pp. 529–534, http://dx.doi.org/10.1109/ICOIN48656.2020.9016478.

[67] C. Huang, D. Liu, J. Ni, R. Lu, X. Shen, Achieving accountable and efficient data sharing in industrial internet of things, IEEE Trans. Ind. Inform. 17 (2) (2021) 1416–1427, http://dx.doi.org/10.1109/TII.2020.2982942.

[68] J. Shen, T. Zhou, D. He, Y. Zhang, X. Sun, Y. Xiang, Block design-based key agreement for group data sharing in cloud computing, IEEE Trans. Dependable Secure Comput. 16 (6) (2019) 996–1010, http://dx.doi.org/10.1109/TDSC.2017.2725953.

[69] A. Fromm, V. Stepa, HDFT++ hybrid data flow tracking for saas cloud services, in: 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing, (CSCloud), IEEE, New York, NY, USA, 2017, pp. 333–338, http://dx.doi.org/10.1109/CSCloud.2017.9.

[70] Z. Qin, H. Xiong, S. Wu, J. Batamuliza, A survey of proxy re-encryption for secure data sharing in cloud computing, IEEE Trans. Serv. Comput. (2016) 1, http://dx.doi.org/10.1109/TSC.2016.2551238.

[71] T. Pasquier, J. Bacon, J. Singh, D. Eyers, Data-centric access control for cloud computing, in: Proceedings of the 21st ACM on Symposium on Access Control Models and Technologies, ACM, Shanghai China, 2016, pp. 81–88, http://dx.doi.org/10.1145/2914642.2914662.

[72] A. Bessani, M. Correia, B. Quaresma, F. André, P. Sousa, DepSky: dependable and secure storage in a cloud-of-clouds, ACM Trans. Storage 9 (4) (2013) 1–33, http://dx.doi.org/10.1145/2535929.

[73] S. Sundareswaran, A. Squicciarini, D. Lin, Ensuring distributed accountability for data sharing in the cloud, IEEE Trans. Dependable Secure Comput. 9 (4) (2012) 556–568, http://dx.doi.org/10.1109/TDSC.2012.26.

[74] A. Rafique, D. Van Landuyt, E. Heydari Beni, B. Lagaisse, W. Joosen, Cryptdice: distributed data protection system for secure cloud data storage and computation, Inf. Syst. 96 (2021) 101671, http://dx.doi.org/10.1016/j.is.2020.101671.

[75] K. Edemacu, B. Jang, J.W. Kim, CESCR: CP-ABE for efficient and secure sharing of data in collaborative ehealth with revocation and no dummy attribute, PLoS One 16 (5) (2021) e0250992, http://dx.doi.org/10.1371/journal.pone.0250992.

[76] B. Waters, Ciphertext-policy attribute-based encryption:?An expressive, efficient, and provably secure realization, in: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, D. Catalano, N. Fazio, R. Gennaro, A. Nicolosi (Eds.), Public Key Cryptography – PKC 2011, vol. 6571, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 53–70, http://dx.doi.org/10.1007/978-3-642-19379-8_4.

[77] H. Lei, Y. Yan, Z. Bao, Q. Wang, Y. Zhang, W. Shi, SDSBT: A secure multi-party data sharing platform based on blockchain and TEE, in: Cyberspace Safety and Security, vol. 12653, Springer International Publishing, Cham, 2021, pp. 184–196, http://dx.doi.org/10.1007/978-3-030-73671-2_17.

[78] P. Bonatti, S. Kirrane, A. Polleres, R. Wenning, Transparent personal data processing: the road ahead, in: F. Bitsch S. Tonetta (Ed.), in: Computer Safety, Reliability, and Security, vol. 10489, Springer International Publishing, Cham, 2017, pp. 337–349, http://dx.doi.org/10.1007/978-3-319-66284-8_28.

[79] F. Schäfer, J. Rosen, C. Zimmermann, F. Wortmann, Unleashing the potential of data ecosystems: establishing digital trust through trust-enhancing technologies, in: ECIS 2023 Research Papers, 2023.

[80] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, W. Zhao, A survey on big data market: pricing, trading and protection, IEEE Access 6 (2018) 15132–15154, http://dx.doi.org/10.1109/ACCESS.2018.2806881.

[81] G.M. Garrido, J. Sedlmeir, Ö. Uludağ, I.S. Alaoui, A. Luckow, F. Matthes, Revealing the landscape of privacy-enhancing technologies in the context of data markets for the IoT: A systematic literature review, J. Netw. Comput. Appl. 207 (2022) 103465, http://dx.doi.org/10.1016/j.jnca.2022.103465.

[82] D. McCabe, A. Satariano, The Era of Borderless Data Is Ending, New York Times, 2022.

[83] M.D. Ryan, Enhanced certificate transparency and end-to-end encrypted mail, in: Proceedings 2014 Network and Distributed System Security Symposium, Internet Society, San Diego, CA, 2014, http://dx.doi.org/10.14722/ndss.2014.23379.

[84] L. kacha, A. Zitouni, An overview on data security in cloud computing, in: Cybernetics Approaches in Intelligent Systems, vol. 661, Springer International Publishing, 2018, pp. 250–261, http://dx.doi.org/10.1007/978-3-319-67618-0_23.

[85] F. Boemer, A. Costache, R. Cammarota, C. Wierzynski, nGraph-HE2: A high-throughput framework for neural network inference on encrypted data, in: Proceedings of the 7th ACM Workshop on Encrypted Computing & Applied Homomorphic Cryptography - WAHC'19, ACM Press, London, United Kingdom, 2019, pp. 45–56, http://dx.doi.org/10.1145/3338469.3358944.

[86] J. Park, R. Sandhu, The UCON_{ABC} usage control model, ACM Trans. Inf. Syst. Secur. 7 (1) (2004) 128–174, http://dx.doi.org/10.1145/984334.984339.

[87] M. Hilty, D. Basin, A. Pretschner, On obligations, in: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, S.d.C. di Vimercati, P. Syverson, D. Gollmann (Eds.), Computer Security – ESORICS 2005, vol. 3679, Springer, Berlin, Heidelberg, 2005, pp. 98–117, http://dx.doi.org/10.1007/11555827_7.

[88] M. Hilty, A. Pretschner, D. Basin, C. Schaefer, T. Walter, A policy language for distributed usage control, in: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, J. Biskup, J. López (Eds.), Computer Security – ESORICS 2007, vol. 4734, Springer, Berlin, Heidelberg, 2007, pp. 531–546, http://dx.doi.org/10.1007/978-3-540-74835-9_35.

[89] F. Kelbert, A. Pretschner, Data usage control enforcement in distributed systems, in: Proceedings of the Third ACM Conference on Data and Application Security and Privacy - CODASPY '13, ACM Press, San Antonio, Texas, USA, 2013, p. 71, http://dx.doi.org/10.1145/2435349.2435358.

[90] F. Kelbert, A. Pretschner, A fully decentralized data usage control enforcement infrastructure, in: T. Malkin, V. Kolesnikov, A.B. Lewko, M. Polychronakis (Eds.), Applied Cryptography and Network Security, vol. 9092, Springer International Publishing, Cham, 2015, pp. 409–430, http://dx.doi.org/10.1007/978-3-319-28166-7_20.

[91] I. Akaichi, S. Kirrane, Usage control specification, enforcement, and robustness: A survey, 2022, cs, arXiv:2203.04800.

[92] S. Steinbuss, et al., Usage control in the international data spaces, 2021.

[93] J. Pampus, B.-F. Jahnke, R. Quensel, Evolving data space technologies: lessons learned from an IDS connector reference implementation, in: T. Margaria, B. Steffen (Eds.), Leveraging Applications of Formal Methods, Verification and Validation. Practice, vol. 13704, Springer Nature, Switzerland, Cham, 2022, pp. 366–381, http://dx.doi.org/10.1007/978-3-031-19762-8_27.

[94] A. Hosseinzadeh, A. Eitel, C. Jung, A systematic approach toward extracting technically enforceable policies from data usage control requirements, in: Proceedings of the 6th International Conference on Information Systems Security and Privacy, SCITEPRESS - Science and Technology Publications, Valletta, Malta, 2020, pp. 397–405, http://dx.doi.org/10.5220/0008936003970405.
[95] M. Schneider, R.J. Masti, S. Shinde, S. Capkun, R. Perez, SoK: hardware-supported trusted execution environments, 2022, arXiv:2205.12742.
[96] X. Ge, H.-C. Kuo, W. Cui, Hecate: lifting and shifting on-premises workloads to an untrusted cloud, in: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, ACM, Los Angeles CA USA, 2022, pp. 1231–1242, http://dx.doi.org/10.1145/3548606.3560592.
[97] A. Nilsson, P.N. Bideh, J. Brorsson, A survey of published attacks on intel SGX, 2020, cs, arXiv:2006.13598.
[98] M.-W. Shih, S. Lee, T. Kim, M. Peinado, T-SGX: eradicating controlled-channel attacks against enclave programs, in: Proceedings 2017 Network and Distributed System Security Symposium, Internet Society, San Diego, CA, 2017, http://dx.doi.org/10.14722/ndss.2017.23193.
[99] S. Sasy, S. Gorbunov, C.W. Fletcher, ZeroTrace : oblivious memory primitives from intel SGX, in: Proceedings 2018 Network and Distributed System Security Symposium, Internet Society, San Diego, CA, 2018, http://dx.doi.org/10.14722/ndss.2018.23239.
[100] J. Lohmöller, E. Vlad, M. Dahlmanns, K. Wehrle, Poster: bridging trust gaps: data usage transparency in federated data ecosystems, in: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, ACM, Copenhagen, Denmark, 2023, pp. 3582–3584, http://dx.doi.org/10.1145/3576915.3624371.
[101] E. Stark, J. DeBlasio, D. O'Brien, Certificate transparency in google chrome: past, present, and future, IEEE Secur. Privacy 19 (6) (2021) 112–118, http://dx.doi.org/10.1109/MSEC.2021.3103461.
[102] R. Gennaro, C. Gentry, B. Parno, Non-interactive verifiable computing: outsourcing computation to untrusted workers, in: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, T. Rabin (Eds.), Advances in Cryptology – CRYPTO 2010, vol. 6223, Springer, Berlin, Heidelberg, 2010, pp. 465–482, http://dx.doi.org/10.1007/978-3-642-14623-7_25.
[103] B. Parno, J. Howell, C. Gentry, M. Raykova, Pinocchio: nearly practical verifiable computation, in: 2013 IEEE Symposium on Security and Privacy, IEEE, Berkeley, CA, 2013, pp. 238–252, http://dx.doi.org/10.1109/SP.2013.47.
[104] I. Kunz, V. Casola, A. Schneider, C. Banse, J. Schütte, Towards tracking data flows in cloud architectures, in: 2020 IEEE 13th International Conference on Cloud Computing, (CLOUD), 2020, pp. 445–452.
[105] M. Backes, N. Grimm, A. Kate, Data lineage in malicious environments, IEEE Trans. Dependable Secure Comput. 13 (2) (2016) 178–191, http://dx.doi.org/10.1109/TDSC.2015.2399296.
[106] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, Decentralized Bus. Rev. (2008) 21260.
[107] V. Buterin, et al., A next-generation smart contract and decentralized application platform, White Pap. 3 (37) (2014) 2–1.
[108] T. Pulls, R. Peeters, K. Wouters, Distributed privacy-preserving transparency logging, in: Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society, ACM, Berlin, Germany, 2013, pp. 83–94, http://dx.doi.org/10.1145/2517840.2517847.
[109] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, PLoS Med. 12 (3) (2015) e1001779, http://dx.doi.org/10.1371/journal.pmed.1001779.
[110] P. Elliott, T.C. Peakman, on behalf of UK Biobank, The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine, Int. J. Epidemiol. 37 (2) (2008) 234–244, http://dx.doi.org/10.1093/ije/dym276.
[111] H. Busby, P. Martin, Biobanks, national identity and imagined communities: the case of UK biobank, Sci. Cult. 15 (3) (2006) 237–251, http://dx.doi.org/10.1080/09505430600890693.
[112] https://www.bio-itworld.com/news/2020/08/26/uk-biobank-contracts-with-dnanexus-aws-to-build-data-analysis-platform. Accessed 16 February 2023, 2023.
[113] M. Doucet, M. Yuille, L. Georghiou, G. Dagher, Biobank sustainability: Current status and future prospects, BSAM 5 (2017) 1–7, http://dx.doi.org/10.2147/BSAM.S100899.
[114] https://www.ukbiobank.ac.uk/media/llupxihh/20210309-access-procedures-v2-0-final.pdf. Accessed 16 February 2023, 2023.
[115] A. Cohen, K. Nissim, Towards formalizing the GDPR's notion of singling out, Proc. Natl. Acad. Sci. USA 117 (15) (2020) 8344–8352, http://dx.doi.org/10.1073/pnas.1914598117.
[116] The All of Us Research Program Investigators, The all of us research program, N. Engl. J. Med. 381 (7) (2019) 668–676, http://dx.doi.org/10.1056/NEJMsr1809937.
[117] https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. Accessed 16 February 2023, 2023.
[118] A. Anjum, S.U.R. Malik, K.-K.R. Choo, A. Khan, A. Haroon, S. Khan, S.U. Khan, N. Ahmad, B. Raza, An efficient privacy mechanism for electronic health records, Comput. Secur. 72 (2018) 196–211, http://dx.doi.org/10.1016/j.cose.2017.09.014.
[119] J.J. Panackal, A.S. Pillai, V.N. Krishnachandran, Disclosure risk of individuals: A k-anonymity study on health care data related to Indian population, in: 2014 International Conference on Data Science & Engineering, (ICDSE), IEEE, Kochi, India, 2014, pp. 200–205, http://dx.doi.org/10.1109/ICDSE.2014.6974637.
[120] T.A. Manolio, R. Collins, Enhancing the feasibility of large cohort studies, JAMA 304 (20) (2010) 2290, http://dx.doi.org/10.1001/jama.2010.1686.
[121] https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-creates-cloud-based-health-data-analysis-platform-to-unleash-the-imaginations-of-the-world-s-best-scientific-mind. Accessed 16 February 2023, 2023.
[122] https://www.ukbiobank.ac.uk/media/nt0p5s1k/gdpr.pdf. Accessed 26 July 2023, 2018.
[123] R. Collins, What makes UK Biobank special? Lancet 379 (9822) (2012) 1173–1174, http://dx.doi.org/10.1016/S0140-6736(12)60404-8.
[124] G.S. Brost, M. Huber, M. Weiß, M. Protsenko, J. Schütte, S. Wessel, An ecosystem and IoT device architecture for building trust in the industrial data space, in: Proceedings of the 4th ACM Workshop on Cyber-Physical System Security, ACM, Incheon Republic of Korea, 2018, pp. 39–50, http://dx.doi.org/10.1145/3198458.3198459.
[125] J. Pennekamp, M. Henze, S. Schmidt, P. Niemietz, M. Fey, D. Trauth, T. Bergs, C. Brecher, K. Wehrle, Dataflow challenges in an internet of production: A security & privacy perspective, in: Proceedings of the ACM Workshop on Cyber-Physical Systems Security & Privacy - CPS-SPC'19, ACM Press, London, United Kingdom, 2019, pp. 27–38, http://dx.doi.org/10.1145/3338499.3357357.
[126] J. Pennekamp, J. Lohmöller, E. Vlad, J. Loos, N. Rodemann, P. Sapel, I.B. Fink, S. Schmitz, C. Hopmann, M. Jarke, G. Schuh, K. Wehrle, M. Henze, Designing secure and privacy-preserving information systems for industry benchmarking, in: Proceedings of the 35th International Conference on Advanced Information Systems Engineering, (CAiSE '23), 2023, http://dx.doi.org/10.1007/978-3-031-34560-9_29.
[127] I. Ali, M.G.S. Aboelmaged, Implementation of supply chain 4.0 in the food and beverage industry: Perceived drivers and barriers, IJPPM 71 (4) (2022) 1426–1443, http://dx.doi.org/10.1108/IJPPM-07-2020-0393.
[128] L. Gleim, J. Pennekamp, M. Liebenberg, M. Buchsbaum, P. Niemietz, S. Knape, A. Epple, S. Storms, D. Trauth, T. Bergs, C. Brecher, S. Decker, G. Lakemeyer, K. Wehrle, FactDAG: formalizing data interoperability in an internet of production, IEEE Internet Things J. 7 (4) (2020) 3243–3253, http://dx.doi.org/10.1109/JIOT.2020.2966402.
[129] A. Rühmkorf, Article: the german supply chain law: A first step towards more corporate sustainability, EUCL 20 (1) (2023) 6–14, http://dx.doi.org/10.54648/EUCL2023003.

[130] J. Cowls, A. Tsamados, M. Taddeo, L. Floridi, The AI gambit: Leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations, AI Soc. 38 (1) (2023) 283–307, http://dx.doi.org/10.1007/s00146-021-01294-x.
[131] A. Kirimtat, O. Krejcar, A. Kertesz, M.F. Tasgetiren, Future trends and current state of smart city concepts: A survey, IEEE Access 8 (2020) 86448–86467, http://dx.doi.org/10.1109/ACCESS.2020.2992441.
[132] M. Zhou, J. Yan, D. Feng, Digital twin and its application to power grid online analysis, CSEE JPES (2019) 391–398, http://dx.doi.org/10.17775/CSEEJPES.2018.01460.
[133] M.R. Asghar, G. Dan, D. Miorandi, I. Chlamtac, Smart meter data privacy: A survey, IEEE Commun. Surv. Tutor. 19 (4) (2017) 2820–2835, http://dx.doi.org/10.1109/COMST.2017.2720195.
[134] E. Hossain, I. Khan, F. Un-Noor, S.S. Sikander, M.S.H. Sunny, Application of big data and machine learning in smart grid, and associated security concerns: A review, IEEE Access 7 (2019) 13960–13988, http://dx.doi.org/10.1109/ACCESS.2019.2894819.
[135] T. Li, D. Sun, P. Jing, K. Yang, Smart card data mining of public transport destination: A literature review, Information 9 (1) (2018) 18, http://dx.doi.org/10.3390/info9010018.
[136] S. Porru, F.E. Misso, F.E. Pani, C. Repetto, Smart mobility and public transport: opportunities and challenges in rural and urban areas, J. Traffic Transp. Eng. (Engl. Ed.) 7 (1) (2020) 88–97, http://dx.doi.org/10.1016/j.jtte.2019.10.002.
[137] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: 12th International Conference on Evaluation and Assessment in Software Engineering, (EASE), 2008, http://dx.doi.org/10.14236/ewic/EASE2008.8.
[138] S. More, L. Alber, You shall not compute on my data: access policies for privacy-preserving data marketplaces and an implementation for a distributed market using MPC, in: Proceedings of the 17th International Conference on Availability, Reliability and Security, ACM, Vienna, Austria, 2022, pp. 1–8, http://dx.doi.org/10.1145/3538969.3544445.
[139] C. Niu, Z. Zheng, F. Wu, X. Gao, G. Chen, Trading data in good faith: integrating truthfulness and privacy preservation in data markets, in: 2017 IEEE 33rd International Conference on Data Engineering, (ICDE), IEEE, San Diego, CA, USA, 2017, pp. 223–226, http://dx.doi.org/10.1109/ICDE.2017.80.
[140] P. Chen, P. Shi, J. Xu, X. Fu, L. Li, T. Zhong, L. Xiang, J. Kong, TeeSwap: private data exchange using smart contract and trusted execution environment, in: 2021 IEEE 23rd Int Conf on High Performance Computing & Communications; 7th Int Conf on Data Science & Systems; 19th Int Conf on Smart City; 7th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application, (HPCC/DSS/SmartCity/DependSys), IEEE, Haikou, Hainan, China, 2021, pp. 237–244, http://dx.doi.org/10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00057.
[141] K. Koch, S. Krenn, D. Pellegrino, S. Ramacher, Privacy-preserving analytics for data markets using MPC, in: M. Friedewald, S. Schiffner, S. Krenn (Eds.), in: Privacy and Identity Management, vol. 619, Springer International Publishing, Cham, 2021, pp. 226–246, http://dx.doi.org/10.1007/978-3-030-72465-8_13.
[142] K. Koch, S. Krenn, T. Marc, S. More, S. Ramacher, KRAKEN: A privacy-preserving data market for authentic data, in: Proceedings of the 1st International Workshop on Data Economy, ACM, Rome, Italy, 2022, pp. 15–20, http://dx.doi.org/10.1145/3565011.3569057.
[143] K. Kayaba, H. Oguri, Y. Yamaoka, Evaluation of secure remote offering service for information bank, in: Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy, ACM, New Orleans la USA, 2020, pp. 144–146, http://dx.doi.org/10.1145/3374664.3379526.
[144] N. Serrano, F. Cuenca, A peer-to-peer ownership-preserving data marketplace, in: 2021 IEEE International Conference on Blockchain (Blockchain), IEEE, Melbourne, Australia, 2021, pp. 394–400, http://dx.doi.org/10.1109/Blockchain53845.2021.00062.
[145] A. Andreas, C.X. Mavromoustakis, G. Mastorakis, D.-T. Do, J.M. Batalla, E. Pallis, E.K. Markakis, Towards an optimized security approach to IoT devices with confidential healthcare data exchange, Multimed. Tools Appl. 80 (20) (2021) 31435–31449, http://dx.doi.org/10.1007/s11042-021-10827-x.
[146] I. Oliver, S. Holtmanns, Y. Miche, S. Lal, L. Hippeläinen, A. Kalliola, S. Ravidas, Experiences in trusted cloud computing, in: Z. Yan, R. Molva, W. Mazurczyk, R. Kantola (Eds.), in: Network and System Security, vol. 10394, Springer International Publishing, Cham, 2017, pp. 19–30, http://dx.doi.org/10.1007/978-3-319-64701-2_2.
[147] C. Niu, Z. Zheng, F. Wu, X. Gao, G. Chen, Achieving data truthfulness and privacy preservation in data markets, IEEE Trans. Knowl. Data Eng. 31 (1) (2019) 105–119, http://dx.doi.org/10.1109/TKDE.2018.2822727.
[148] C. Esposito, A. Castiglione, F. Frattini, M. Cinque, Y. Yang, K.-K.R. Choo, On data sovereignty in cloud-based computation offloading for smart cities applications, IEEE Internet Things J. 6 (3) (2019) 4521–4535, http://dx.doi.org/10.1109/JIOT.2018.2886410.
[149] A. Kiayias, B. Yener, M. Yung, Privacy-preserving information markets for computing statistical data, in: R. Dingledine, P. Golle (Eds.), vol. 5628, Springer, Berlin, Heidelberg, 2009, pp. 32–50, http://dx.doi.org/10.1007/978-3-642-03549-4_3.
[150] I. Oliver, S. Holtmanns, S. Lal, Experiences in trusted cloud computing, JICTS 6 (3) (2018) 263–278, http://dx.doi.org/10.13052/jicts2245-800X.635.
[151] M. Colombo, A. Lazouski, F. Martinelli, P. Mori, A proposal on enhancing XACML with continuous usage control features, in: F. Desprez, V. Getov, T. Priol, R. Yahyapour (Eds.), Grids, P2P and Services Computing, Springer US, Boston, MA, 2010, pp. 133–146, http://dx.doi.org/10.1007/978-1-4419-6794-7_11.
[152] Q.H. Cao, M. Giyyarpuram, R. Farahbakhsh, N. Crespi, Policy-based usage control for a trustworthy data sharing platform in smart cities, Future Gener. Comput. Syst. 107 (2020) 998–1010, http://dx.doi.org/10.1016/j.future.2017.05.039.
[153] F. Cirillo, B. Cheng, R. Porcellana, M. Russo, G. Solmaz, H. Sakamoto, S.P. Romano, IntentKeeper: intent-oriented data usage control for federated data analytics, in: 2020 IEEE 45th Conference on Local Computer Networks, (LCN), IEEE, Sydney, NSW, Australia, 2020, pp. 204–215, http://dx.doi.org/10.1109/LCN48667.2020.9314823.
[154] F. Kelbert, A. Pretschner, Data usage control for distributed systems, ACM Trans. Priv. Secur. 21 (3) (2018) 1–32, http://dx.doi.org/10.1145/3183342.
[155] F.Y. Rashid, The rise of confidential computing: big tech companies are adopting a new security model to protect data while it's in use-[news], IEEE Spectr. 57 (6) (2020) 8–9, http://dx.doi.org/10.1109/MSPEC.2020.9099920.

Johannes Lohmöller received the B.Sc. and M.Sc. degrees in Computer Science from RWTH Aachen University. He is a researcher at the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University. His research centers around sovereignty, security and trust across distributed data ecosystems. In particular, his special interests include privacy-, authenticity-, and trust-enhancing technologies, and secure computations together with their application to data ecosystems.

Jan Pennekamp received the B.Sc. and M.Sc. degrees in Computer Science from RWTH Aachen University with honors. He is a researcher at the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University. His research focuses on security & privacy aspects in the Industrial Internet of Things (IIoT). In particular, his special interests include privacy-enhancing technologies, the design of privacy-preserving protocols, and secure computations as well as their application.

Roman Matzutt received a PhD in Computer Science on the data-management implications for permissionless blockchains in 2023 from RWTH Aachen University. He has since joined Fraunhofer FIT as a postdoctoral researcher. His research focuses on the challenges and opportunities of accountable and distributed data ledgers, privacy-enhancing technologies, and security and privacy aspects and applications of artificial intelligence.

Carolin Victoria Schneider studied medicine in Aachen, Germany, and completed a postdoctoral fellowship at the Center for Genetics and Translational Therapeutics at the University of Pennsylvania. Following her postdoctoral fellowship, she started as a W1 professor in Aachen and as an adjunct professor at University of Pennsylvania, focusing on integrating big data to prevent metabolic diseases at a population level.

Eduard Vlad received the B.Sc. degree in Computer Science from RWTH Aachen University. He is currently pursuing his M.Sc. in Computer Science and works as a student assistant at the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University.

Christian Trautwein studied medicine in Mainz and Tübingen. He worked as a research assistant at the I Medical Clinic Mainz (1988–1992) and performed research stays in Birmingham, U.K., and at the University of California, San Diego, USA. In 1993, he moved to Hannover Medical School (MHH) and founded his own research group. At MHH, he was appointed as a consultant (1995), assistant Professor (C2, 1996), senior consultant (2000), and associate Professor (C3, 2002) at the Center for Internal Medicine. From 2005 to 2024 he was full professor and Director of Medicine III at the University Hospital, RWTH Aachen.

Klaus Wehrle received the Diploma (equiv. M.Sc.) and Ph.D. degrees from University of Karlsruhe (now KIT), both with honors. Since 2010, he is a full professor and the head of the Chair of Communication and Distributed Systems (COMSYS) at RWTH Aachen University. His research interests include (but are not limited to) engineering of networking protocols, (formal) methods for protocol engineering and network analysis, reliable communication software, and all operating system issues of networking. He is a member of IEEE, ACM, Sigcomm, GI, VDE, GI/ITG-Fachgruppe KuVS, and ACATECH. Before joining RWTH Aachen University in 2006, he was a postdoctoral researcher at the International Computer Science Institute (ICSI) in 2002 and 2003. He co-coordinates the DFG Priority Programme on Cyber–Physical Networking. Furthermore, he serves as a representative for EE and CS in the main evaluation board of the Alexander-von-Humboldt Foundation.