On the Need for Strong Sovereignty in Data Ecosystems

Johannes Lohmöller
RWTH Aachen University
Aachen, Germany
lohmoeller@comsys.rwth-aachen.de

Jan Pennekamp
RWTH Aachen University
Aachen, Germany
pennekamp@comsys.rwth-aachen.de

Roman Matzutt
RWTH Aachen University
Aachen, Germany
matzutt@comsys.rwth-aachen.de

Klaus Wehrle
RWTH Aachen University
Aachen, Germany
wehrle@comsys.rwth-aachen.de

ABSTRACT

Data ecosystems are the foundation of emerging data-driven business models as they (i) enable an automated exchange between their participants and (ii) provide them with access to huge and heterogeneous data sources. However, the corresponding benefits come with unforeseen risks as also sensitive information is potentially exposed. Consequently, data security is of utmost importance and, thus, a central requirement for the successful implementation of these ecosystems. Current initiatives, such as IDS and GAIA-X, hence foster sovereign participation via a federated infrastructure where participants retain local control. However, these designs place significant trust in remote infrastructure by mostly implementing organizational security measures such as certification processes prior to admission of a participant. At the same time, due to the sensitive nature of involved data, participants are incentivized to bypass security measures to maximize their own benefit: In practice, this issue significantly weakens sovereignty guarantees. In this paper, we hence claim that data ecosystems must be extended with technical means to reestablish such guarantees. To underpin our position, we analyze promising building blocks and identify three core research directions toward stronger data sovereignty, namely trusted remote policy enforcement, verifiable data tracking, and integration of resource-constrained participants. We conclude that these directions are critical to securely implement data ecosystems in data-sensitive contexts.

Reference Format:
Johannes Lohmöller, Jan Pennekamp, Roman Matzutt, and Klaus Wehrle. On the Need for Strong Sovereignty in Data Ecosystems. PVLDB, 15(1): XXX-XXX, 2022.
doi:XX.XX/XXX.XX

1 INTRODUCTION

Data-driven business models are an invaluable pillar for modern industries, and their importance will increase with growing demands requiring more complex and globally distributed operation, as well as sophisticated collaborations to improve the status quo [73]. *Data ecosystems* provide the foundation for such data-driven business models as they center around automating data exchanges and value creation based on huge and heterogeneous data sources from various stakeholders [66]. Added value can be created by, for instance, improving algorithms underlying existing analytics or extracting new insights of previously recorded data [65]. Crucially, this process involves the integration of distributed data sources owned by different stakeholders. Here, data ecosystem initiatives such as International Data Spaces (IDS) [67] and GAIA-X [31] aim to provide a trustworthy environment for the discovery, sharing, and processing of available data, irrespective of specific domains.

However, current efforts to establish the necessary trust among stakeholders heavily rely on organizational agreements and processes [28, 67]. For instance, the IDS certification process asserts that participants use audited software and develop defense-in-depth strategies for protection [67]. Participants receive no additional security guarantees beyond this ahead-of-time certification and have no means to verify that other participants handle their data as intended (and required). Here, the lack of stronger guarantees effectively ends sovereignty of participants in the moment of sharing.

In this paper, we argue that data ecosystems need to provide their participants with *strong and continual guarantees* about the security of their provided data to maintain each participant's *data sovereignty*. Moreover, driven by privacy and security concerns, recent regulatory efforts set strict rules on how data may flow across organizational borders, raising the need for fine-grained control [59]. To this end, data ecosystems are only sustainable if stakeholders are willing to participate by providing and consuming data actively. However, we argue that data-consuming parties are currently incentivized to ignore previously agreed terms for data usage. Such behavior hurts data owners as they are not adequately compensated 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 require solid *technical measures*, such as cryptographically enforceable guarantees and verifiable continual security monitoring, to facilitate the establishment of trust between remote and potentially mutually unknown participants. In this paper, we provide more background on the current state of data ecosystems, identify shortcomings of ongoing data ecosystem initiatives, and derive and discuss future research directions steered toward improving the sovereignty and trust of participants in data ecosystems.

Proceedings of the VLDB Endowment, Vol. 15, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX

2 A PRIMER ON CURRENT DATA ECOSYSTEM INITIATIVES AND THEIR ARCHITECTURES

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

Ecosystem Goals. The need to share data with collaborators within specific sectors has been recognized in a variety of domains, including supply chains [7], public health [4, 53], and mobility [24]. Here, on the one hand, data ecosystems aim to provide *multi-sided platforms* [66] that facilitate an automated data exchange following the *FAIR principle* [91], 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 it will be shared and under what terms. This fine-grained control is the foundation of *data sovereignty* [65]. Achieving these goals requires solving issues w.r.t. organization [66], semantics and data quality [33], and interfacing [14], all of which are currently under active research.

Definitions. So far, we have seen data ecosystems only as a means for exchanging data as required in emerging data markets and other use cases [65]. In fact, data ecosystems emerged without a standard definition in mind. Oliveira and Lóscio [64] address this gap by reviewing and merging concurring data ecosystem definitions; as a result, they define a data ecosystem as a combination of independently operated networks that produce and provide data, but also other assets like software or services. Furthermore, the authors highlight that such data ecosystems are self-regulated and driven by collaboration and competition between actors [64]. Additionally, we emphasize that data ecosystems form platforms that have to define common interfaces and rules to enable collaboration across independent networks. Accordingly, we refer to data ecosystem *participants* as networks that implement the interfaces and accept the rules defined by a given ecosystem.

Similarly, the notion of *data sovereignty*, i.e., one of the critical concepts of data ecosystems, currently lacks a clear and common definition [42]. If used in the context of data ecosystems, researchers generally agree that data sovereignty relates to control and ownership of data items, together with specific claims and obligations made by involved parties [22, 45, 71, 81]. Hence, within this paper, we will focus on this aspect of data sovereignty. To set this into a broader context, the review by Hummel et al. [42] describes data sovereignty as covering multiple contexts and values ranging from legislation to clinical practice and control and power to recognition, respectively.

Initiatives. Superseding a previously rather tedious bilateral exchange, the goal of initiatives like the International Data Spaces (IDS) [9, 65, 66], GAIA-X [14, 31], Data Sharing Coalition [23], IHAN [44], FIWARE [20], CEF [17], or BDVA [11] 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. Some initiatives rather bundle forces toward the adoption of 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 [67], GAIA-X [31], and FIWARE [1] have released technical documentation that permit a deeper analysis with regard to implemented data security and trust measures. Specifically, IDS and GAIA-X both work toward a standard interface to locate and access data and provide an organizational context, including identification, admission, and certification of participants [14]. Thus, in the remainder of this paper, we primarily study these general-purpose initiatives. While IDS aims to provide a framework under which data spaces can be built quickly, e.g., targeting a specific domain with coherent participants, GAIA-X plans to establish a single central cross-domain platform [14]. Moving toward domain-specific concepts, initial projects such as CATENA-X [88], an initiative inside the automotive domain, are picking up their ideas, while established platforms such as FIWARE [20], a framework to connect smart devices, start to provide compatible interfaces [3].

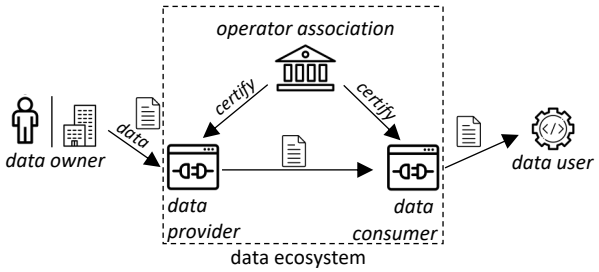

Figure 1: Participating entities in data ecosystems. Data flows from left to right, with data provider and data consumer implementing a common ecosystem interface. The data ecosystem's operator also handles orthogonal tasks, including admission and discovery of participants and data.

Architecture. Despite their slightly different scopes, IDS and GAIA-X share a similar architecture, so we analyze both initiatives together as data ecosystem implementations. Organizing the data exchange, data ecosystems commonly assign different roles to participants. Figure 1 shows the overall scenario we are considering together with the main participants. A single data exchange can be considered bilateral, such that we can suppose the following roles [67]: 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, in certain situations, the providing entity does not formally own the data. 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 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 also fulfill the data user role if both processes are co-located. Noteworthy, GAIA-X does not separate the data consumer and data user [31], but 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 up to the point where a data sharing decision has been made and data is actually transferred to the data consumer.

Trust. To not let sovereignty end 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 w.r.t. data protection. Certification includes, but is not limited to, defense-in-depth strategies and security event monitoring systems [18, 60]. Specifically, the IDS requires prior certification steps and attests successful certification via a public key infrastructure, establishing a trusted identity layer [14, 67]. Contrarily, GAIA-X does not target a specific certification but requires participants to provide a standardized self-description with claims that are checked before a participant's admission [14]. 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, access control alone is insufficient, as data sovereignty would end once the flow of data between participants took place after access has been legitimately granted. *Usage control* [74] could possibly 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, for instance, 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 [43], a policy language for digital rights management that allows fine-grained modeling of usage terms [67]. For enforcement, the data owner has to trust that the consuming party abides by the negotiated terms. To this end, he can only rely on the certification of the consumer required to join as a participant, but can neither monitor the process himself, nor receive a credible proof that usage terms were enforced. 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 [26]. Specifically, such contracts regulate the terms of usage and the overall setting, e.g., regarding a monetary compensation [67] or a penalty for breach of contract. Contracts can be bilateral or multilateral but will typically not cover the entirety of data space participants [67], thereby limiting spontaneous data access. Within negotiated legal contracts, data ecosystems such as IDS then plan to (automatically) negotiate a refined technical contract. This refined contract translates terms into machine-readable policies that grant specific permissions on the exchanged dataset and potential obligations [67].

3 DATA ECOSYSTEMS NEED TECHNICAL SECURITY GUARANTEES

Having outlined the fundamental ideas of sovereign data exchange and the technical and organizational framework data ecosystems provide, we now critically review the design decisions of security mechanisms implemented in state-of-the-art data ecosystems. To this end, we analyze the available technical documentation and reference architecture for IDS and GAIA-X. Primarily, we identify a lack of technical means to facilitate strong security guarantees and establish strong trust between participants. Namely, 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 [79]. Specifically, the malicious-but-cautious attacker can misbehave in all possible ways but aims not to leave any verifiable evidence of its misbehavior [79]. 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* [46]. At-rest security and in-transit security are considered solved problems in the context of data ecosystems as they can use widely available building blocks such as storage encryption and transport layer security (TLS), respectively [67]. Contrarily, in-use data security targets data at the moment of processing, e.g., when the decrypted data is loaded into memory and is hence 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 [12, 51]. 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, with monetary compensation handled as part of data exchange and transfers entrusted for a specific purpose, incentives to evade enforcement clearly exist.

Hence, 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 ever sharing sensitive data. With such datasets covering manufacturing plans [7], the identity of suppliers [54], or privacy-sensitive health records [29] 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 [67]. However, merely providing verification of the running software, but essentially lacking memory encryption, TPMs still contribute little to an 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 that grant permissive access to their data shall still be able to track 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 [67]. However, similarly to **I1**, there is currently no 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. Figure 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 there is no 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 and thus need a remedy.

4 TOWARD STRONGER DATA SOVEREIGNTY

The current data ecosystem initiatives strive for seamlessly interconnecting businesses and facilitating 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 (cf. Section 3), 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 data ecosystems that retain participants' data sovereignty. Next, we discuss how available building blocks can be integrated into data ecosystems to address each of the open issues **I1**–**I3**.

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 [2, 37, 38, 47, 48, 68] 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 [67, 86]. However, *enforcing* these policies proves difficult as the data owner cannot directly observe the misconduct of a data user or the consequences thereof [39]. Hilty and Pretschner [37] hence propose to provide data owners with evidence of policy enforcement and limit possible computations. Both approaches are hard to realize within a data ecosystem as they require some technical trust anchor 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, or ARM TrustZone, are promising candidates for closing this gap in the future [83]. 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 [83]. Consequently, TEEs allow for 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 [63]. For example, today's TEE implementations are prone to side-channel attacks that allow for limited data extraction [85]. Countermeasures such as oblivious RAM [80] 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, first work [51] 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 [67]. 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 [79]. A similar approach might improve data usage transparency as well. 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. Such approaches are being researched in the field of verifiable computing [34, 69] and data ecosystems could profit by utilizing corresponding building blocks.

Besides logging, related work also proposes data flow tracking [50] and data fingerprinting [6] 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 [6]. Unfortunately, these fingerprints typically cannot survive intermediate processing steps [6], rendering them inapplicable in some situations. Hence, more research maturing 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 [62], it spawned more versatile distributed ledgers for any information using smart contracts [15]. 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 [75]. 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 [67]. 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 [7]. In this case, however, data owners lose their sovereignty and depend on trust in their customers. Thus, appropriate (technical) guarantees for such situations are desirable.

A scenario that would give data owners assurance 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 [67]. 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.

4.4 Summary

Cryptographic building blocks that have been successfully applied in the past are promising also 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**).

5 ONGOING AND PAST RESEARCH EFFORTS

The potential to improve data ecosystems and the need to address their current issues has also been recognized in previous work. All in all, data ecosystems are subject to past and active research alike, especially due to ongoing large-scale initiatives. In this section, we present notable recent research efforts in 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 [64] survey the components data ecosystems typically comprise. Furthermore, several works discuss requirements and possible ways toward implementing data ecosystems in general, i.e., independent of specific initiatives [14, 33, 36, 65, 66, 93]. Another line of research investigates fundamental challenges faced when implementing (distributed) data sharing systems. Mainly, these challenges engulf transparency requirements [32], addressing the potential lack of trust between participants [33, 41, 61], the need for creating a common semantic understanding among all participants [8], and governance as well as legal constraints [25, 26, 35, 92]. 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 [3, 20] provides a platform to facilitate data exchange in an Internet of Things context and is related to CEF [17]. Furthermore, special-purpose data ecosystems are being considered, e.g., by the NFDI initiative [90], 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 [4, 82], personal data [57], data emerging in the Industrial Internet of Things [5, 56], and data exchange across supply chains, such as in the automotive industry [7, 54, 88], 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 [28, 29, 52, 53], for production technology [55, 72], along supply chains [7], or in education [58]. 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. Besides identifying novel use cases for sharing data via 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. [40] 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 [10, 30, 70, 76, 84, 87], which can be considered conceptually similar to data ecosystems with multiple stakeholders. Such work includes querying encrypted data [77], attribute- or identity-based encryption for access control [27, 52, 54, 89], and distributed ledgers together with TEEs to enforce accountability and access control [51]. Then again, Bonatti et al. [13] 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 might even be translatable for use in data ecosystems.

6 DISCUSSION AND FUTURE WORK

As we have highlighted in Section 3, today's data ecosystems mostly rely on organizational means to implement data protection. However, 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). Finally, ongoing research efforts (cf. Section 5) have envisioned 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. To this end, data ecosystems must provide a framework that allows users to trust the overall system w.r.t. 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* and stronger *hardware-based security* measures are crucial research directions to sustainably strengthen the data sovereignty for participants of data ecosystems.

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 [39]. Most work in this area either targets rights modeling (e.g., [16, 21, 68]) or assumes operation on trusted infrastructure (e.g., [19, 49]), 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 a wide-spread adoption of data ecosystems.

With cryptographic and technical solutions, the ways toward stronger guarantees are 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 [29].

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 [67], cannot provide adequate protection of sensitive data due to the lacking memory encryption. Hence, Trusted Execution Environments (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 [78]. Hence, we argue that data ecosystems should consider employing TEEs as a measure to enforce data owner's rights on remote infrastructure, which would fill the current gap toward implementing end-to-end data sovereignty.

Future Work. 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. Finally, the combination of technically enforceable usage control with usage transparency might also be the first step toward sovereign integration of resource-constrained participants (**I3**).

7 CONCLUSION

Today's data ecosystems facilitate an automated exchange of data in a standardized manner while simultaneously providing access to huge and heterogeneous data sources. Given that these data exchanges and corresponding higher-level applications across domains (e.g., in the automotive industry) 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, these measures are mostly based on organizational means, which we argue, 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.

We raise the crucial issue that today's data ecosystems lack appropriate guarantees w.r.t. 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**). We have further surveyed corresponding technical solutions to these issues and highlight that they are available but have not yet been adopted in practice. To this end, we argue that the success of data ecosystems directly depends on their ability to address the present need for strong data sovereignty of participants. As such, especially modern technical solutions, such as TEEs, promise to provide data owners with strong guarantees of correct data handling, increasing their willingness to participate in available data ecosystems.

ACKNOWLEDGMENTS

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

REFERENCES

[1] 2022. *ETSI GR CIM 007 V1.1.1: Security and Privacy.* Technical Report. France.
[2] Ines Akaichi and Sabrina Kirrane. 2022. Usage Control Specification, Enforcement, and Robustness: A Survey. *arXiv:2203.04800 [cs]* (2022). arXiv:2203.04800 [cs]
[3] Álvaro Alonso, Alejandro Pozo, José Cantera, Francisco de la Vega, and Juan Hierro. 2018. Industrial Data Space Architecture Implementation Using FIWARE. *Sensors* 18, 7 (2018), 2226. https://doi.org/10.3390/s18072226
[4] Arno Appenzeller, Sebastian Bartholomaus, Rudiger Breitschwerdt, Carsten Claussen, Sandra Geisler, Tobias Hartz, Philipp Kachel, Erik Krempel, Sebastian Robert, and Sylke Ruth Zeissig. 2021. 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, 1–8. https://doi.org/10.1109/AICCSA53542.2021.9686918
[5] Henning Baars, Ann Tank, Patrick Weber, Hans-Georg Kemper, Heiner Lasi, and Burkhard Pedell. 2021. Cooperative Approaches to Data Sharing and Analysis for Industrial Internet of Things Ecosystems. *Applied Sciences* 11, 16 (2021), 7547. https://doi.org/10.3390/app11167547
[6] Michael Backes, Niklas Grimm, and Aniket Kate. 2016. Data Lineage in Malicious Environments. *IEEE Transactions on Dependable and Secure Computing* 13, 2 (2016), 178–191. https://doi.org/10.1109/TDSC.2015.2399296
[7] Lennart Bader, Jan Pennekamp, Roman Matzutt, David Hedderich, Markus Kowalski, Volker Lücken, and Klaus Wehrle. 2021. Blockchain-Based Privacy Preservation for Supply Chains Supporting Lightweight Multi-Hop Information Accountability. *Information Processing & Management* 58, 3 (2021), 102529. https://doi.org/10.1016/j.ipm.2021.102529
[8] Sebastian Bader, Jaroslav Pullmann, Christian Mader, Sebastian Tramp, Christoph Quix, Andreas W. Müller, Haydar Akyürek, Matthias Böckmann, Benedikt T. Imbusch, Johannes Lipp, Sandra Geisler, and Christoph Lange. 2020. The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content. In *The Semantic Web – ISWC 2020,* Jeff Z. Pan, Valentina Tamma, Claudia d'Amato, Krzysztof Janowicz, Bo Fu, Axel Polleres, Oshani Seneviratne, and Lalana Kagal (Eds.). Vol. 12507. Springer International Publishing, Cham, 176–192. https://doi.org/10.1007/978-3-030-62466-8_12
[9] Sebastian R. Bader and Maria Maleshkova. 2020. SOLIOT—Decentralized Data Control and Interactions for IoT. *Future Internet* 12, 6 (2020), 105. https://doi.org/10.3390/fi12060105
[10] Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando André, and Paulo Sousa. 2013. DepSky: Dependable and Secure Storage in a Cloud-of-Clouds. *ACM Transactions on Storage* 9, 4 (2013), 1–33. https://doi.org/10.1145/2535929
[11] Big Data Value Association 2022. https://www.bdva.eu/. Accessed 2022-08-09.
[12] Fabian Boemer, Anamaria Costache, Rosario Cammarota, and Casimir Wierzynski. 2019. 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, 45–56. https://doi.org/10.1145/3338469.3358944
[13] Piero Bonatti, Sabrina Kirrane, Axel Polleres, and Rigo Wenning. 2017. Transparent Personal Data Processing: The Road Ahead. In *Computer Safety, Reliability, and Security,* Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch

(Eds.). Vol. 10489. Springer International Publishing, Cham, 337–349. https://doi.org/10.1007/978-3-319-66284-8_28

[14] Arnaud Braud, Gael Fromentoux, Benoit Radier, and Olivier Le Grand. 2021. The Road to European Digital Sovereignty with Gaia-X and IDSA. *IEEE Network* 35, 2 (2021), 4–5. https://doi.org/10.1109/MNET.2021.9387709

[15] Vitalik Buterin et al. 2014. A Next-Generation Smart Contract and Decentralized Application Platform. *white paper* 3, 37 (2014), 2–1.

[16] Quyet H. Cao, Madhusudan Giyyarpuram, Reza Farahbakhsh, and Noel Crespi. 2020. Policy-Based Usage Control for a Trustworthy Data Sharing Platform in Smart Cities. *Future Generation Computer Systems* 107 (2020), 998–1010. https://doi.org/10.1016/j.future.2017.05.039

[17] CEF Digital 2022. https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/CEF+Digital+Home. Accessed 2022-08-09.

[18] CEN European Committee for Standardization. 2017. Information Technology - Security Techniques - Information Security Management Systems - Requirements (ISO/IEC 27001:2013 Including Cor 1:2014 and Cor 2:2015).

[19] Flavio Cirillo, Bin Cheng, Raffaele Porcellana, Marco Russo, Gurkan Solmaz, Hisashi Sakamoto, and Simon Pietro Romano. 2020. IntentKeeper: Intent-oriented Data Usage Control for Federated Data Analytics. In *2020 IEEE 45th Conference on Local Computer Networks (LCN)*. IEEE, Sydney, NSW, Australia, 204–215. https://doi.org/10.1109/LCN48667.2020.9314823

[20] Flavio Cirillo, Gurkan Solmaz, Everton Luis Berz, Martin Bauer, Bin Cheng, and Erno Kovacs. 2019. A Standard-Based Open Source IoT Platform: FIWARE. *IEEE Internet of Things Magazine* 2, 3 (2019), 12–18. https://doi.org/10.1109/IOTM.0001.1800022

[21] Maurizio Colombo, Aliaksandr Lazouski, Fabio Martinelli, and Paolo Mori. 2010. A Proposal on Enhancing XACML with Continuous Usage Control Features. In *Grids, P2P and Services Computing*, Frédéric Desprez, Vladimir Getov, Thierry Priol, and Ramin Yahyapour (Eds.). Springer US, Boston, MA, 133–146. https://doi.org/10.1007/978-1-4419-6794-7_11

[22] Stephane Couture and Sophie Toupin. 2019. What Does the Notion of "Sovereignty" Mean When Referring to the Digital? *New Media & Society* 21, 10 (2019), 2305–2322. https://doi.org/10.1177/1461444819865984

[23] Data Sharing Coalition 2022. https://datasharingcoalition.eu/about-the-data-sharing-coalition/. Accessed 2022-08-09.

[24] Zhaoyang Du, Celimuge Wu, Tsutomu Yoshinaga, Kok-Lim Alvin Yau, Yusheng Ji, and Jie Li. 2020. Federated Learning for Vehicular Internet of Things: Recent Advances and Open Issues. *IEEE Open Journal of the Computer Society* 1 (2020), 45–61. https://doi.org/10.1109/OJCS.2020.2992630

[25] Charlotte Ducuing. 2020. Data as Infrastructure? A Study of Data Sharing Legal Regimes. *Competition and Regulation in Network Industries* 21, 2 (2020), 124–142. https://doi.org/10.1177/1783591719895390

[26] Alexander Duisberg. 2022. Legal Aspects of IDS: Data Sovereignty - What Does It Imply? In *Designing Data Spaces*. Springer.

[27] Kennedy Edemacu, Beakcheol Jang, and Jong Wook Kim. 2021. 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. https://doi.org/10.1371/journal.pone.0250992

[28] David Froelicher, Patricia Egger, João Sá Sousa, Jean Louis Raisaro, Zhicong Huang, Christian Mouchet, Bryan Ford, and Jean-Pierre Hubaux. 2017. UnLynx: A Decentralized System for Privacy-Conscious Data Sharing. *Proceedings on Privacy Enhancing Technologies* 2017, 4 (2017), 232–250. https://doi.org/10.1515/popets-2017-0047

[29] David Froelicher, Juan R. Troncoso-Pastoriza, Jean Louis Raisaro, Michel A. Cuendet, Joao Sa Sousa, Hyunghoon Cho, Bonnie Berger, Jacques Fellay, and Jean-Pierre Hubaux. 2021. *Truly Privacy-Preserving Federated Analytics for Precision Medicine with Multiparty Homomorphic Encryption.* Preprint. Bioinformatics. https://doi.org/10.1101/2021.02.24.432489

[30] Alexander Fromm and Vladislav Stepa. 2017. 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, 333–338. https://doi.org/10.1109/CSCloud.2017.9

[31] Gaia-X Technical Committee. 2021. Gaia-X Architecture Document.

[32] Sandra Geisler, Maria-Esther Vidal, Cinzia Cappiello, Bernadette Farias Lóscio, Avigdor Gal, Matthias Jarke, Maurizio Lenzerini, Paolo Missier, Boris Otto, Elda Paja, Barbara Pernici, and Jakob Rehof. 2022. Knowledge-Driven Data Ecosystems Toward Data Transparency. *Journal of Data and Information Quality* 14, 1 (2022), 1–12. https://doi.org/10.1145/3467022

[33] Joshua Gelhaar and Boris Otto. 2020. Challenges in the Emergence of Data Ecosystems. In *Pacific Asia Conference on Information Systems (PACIS)*. Dubai.

[34] Rosario Gennaro, Craig Gentry, and Bryan Parno. 2010. Non-Interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers. In *Advances in Cryptology – CRYPTO 2010*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, and Tal Rabin (Eds.). Vol. 6223. Springer Berlin Heidelberg, Berlin, Heidelberg, 465–482. https://doi.org/10.1007/978-3-642-14623-7_25

[35] Lukas Helminger and Christian Rechberger. 2022. Multi-Party Computation in the GDPR. In *Privacy Symposium 2022 - Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT)*.

[36] Martin Henze, Marcel Grossfengels, Maik Koprowski, and Klaus Wehrle. 2013. Towards Data Handling Requirements-Aware Cloud Computing. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*. IEEE, Bristol, United Kingdom, 266–269. https://doi.org/10.1109/CloudCom.2013.145

[37] Manuel Hilty, David Basin, and Alexander Pretschner. 2005. On Obligations. In *Computer Security – ESORICS 2005*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Sabrina de Capitani di Vimercati, Paul Syverson, and Dieter Gollmann (Eds.). Vol. 3679. Springer Berlin Heidelberg, Berlin, Heidelberg, 98–117. https://doi.org/10.1007/11555827_7

[38] M. Hilty, A. Pretschner, D. Basin, C. Schaefer, and T. Walter. 2007. A Policy Language for Distributed Usage Control. In *Computer Security – ESORICS 2007*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Joachim Biskup, and Javier López (Eds.). Vol. 4734. Springer Berlin Heidelberg, Berlin, Heidelberg, 531–546. https://doi.org/10.1007/978-3-540-74835-9_35

[39] Arghavan Hosseinzadeh, Andreas Eitel, and Christian Jung. 2020. 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, 397–405. https://doi.org/10.5220/0008936003970405

[40] Cheng Huang, Dongxiao Liu, Jianbing Ni, Rongxing Lu, and Xuemin Shen. 2021. Achieving Accountable and Efficient Data Sharing in Industrial Internet of Things. *IEEE Transactions on Industrial Informatics* 17, 2 (2021), 1416–1427. https://doi.org/10.1109/TII.2020.2982942

[41] Monika Huber, Sascha Wessel, Gerd Brost, and Nadja Menz. 2022. Building Trust in Data Spaces. In *Designing Data Spaces*. Springer.

[42] Patrik Hummel, Matthias Braun, Max Tretter, and Peter Dabrock. 2021. Data Sovereignty: A Review. *Big Data & Society* 8, 1 (2021), 205395172098201. https://doi.org/10.1177/2053951720982012

[43] Renato Ianella. 2007. Open Digital Rights Language (ODRL). *Open Content Licensing: Cultivating the Creative Commons* (2007).

[44] IHAN 2022. https://ihan.fi/. Accessed 2022-08-09.

[45] Kristina Irion. 2012. Government Cloud Computing and National Data Sovereignty: Government Cloud Computing and National Data Sovereignty. *Policy & Internet* 4, 3-4 (2012), 40–71. https://doi.org/10.1002/poi3.10

[46] Lynda kacha and Abdelhafid Zitouni. 2018. An Overview on Data Security in Cloud Computing. Vol. 661. 250–261. https://doi.org/10.1007/978-3-319-67618-0_23 arXiv:1812.09053 [cs]

[47] Florian Kelbert and Alexander Pretschner. 2013. 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, 71. https://doi.org/10.1145/2435349.2435358

[48] Florian Kelbert and Alexander Pretschner. 2015. A Fully Decentralized Data Usage Control Enforcement Infrastructure. In *Applied Cryptography and Network Security*, Tal Malkin, Vladimir Kolesnikov, Allison Bishop Lewko, and Michalis Polychronakis (Eds.). Vol. 9092. Springer International Publishing, Cham, 409–430. https://doi.org/10.1007/978-3-319-28166-7_20

[49] Florian Kelbert and Alexander Pretschner. 2018. Data Usage Control for Distributed Systems. *ACM Transactions on Privacy and Security* 21, 3 (2018), 1–32. https://doi.org/10.1145/3183342

[50] Immanuel Kunz, Valentina Casola, Angelika Schneider, Christian Banse, and Julian Schütte. 2020. Towards Tracking Data Flows in Cloud Architectures. *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)* (2020), 445–452.

[51] Hong Lei, Yun Yan, Zijian Bao, Qinghao Wang, Yongxin Zhang, and Wenbo Shi. 2021. SDSBT: A Secure Multi-party Data Sharing Platform Based on Blockchain and TEE. In *Cyberspace Safety and Security*, Jieren Cheng, Xiangyan Tang, and Xiaozhang Liu (Eds.). Vol. 12653. Springer International Publishing, Cham, 184–196. https://doi.org/10.1007/978-3-030-73671-2_17

[52] Xiuqing Lu and Xiangguo Cheng. 2020. A Secure and Lightweight Data Sharing Scheme for Internet of Medical Things. *IEEE Access* 8 (2020), 5022–5030. https://doi.org/10.1109/ACCESS.2019.2962729

[53] Hui Ma, Rui Zhang, Guomin Yang, Zishuai Song, Kai He, and Yuting Xiao. 2020. Efficient Fine-Grained Data Sharing Mechanism for Electronic Medical Record Systems with Mobile Devices. *IEEE Transactions on Dependable and Secure Computing* 17, 5 (2020), 1026–1038. https://doi.org/10.1109/TDSC.2018.2844814

[54] Sidra Malik, Naman Gupta, Volkan Dedeoglu, Salil S. Kanhere, and Raja Jurdak. 2021. TradeChain: Decoupling Traceability and Identity in Blockchain Enabled Supply Chains. (2021). https://doi.org/10.48550/ARXIV.2105.11217

[55] Simon Mangel, Lars Gleim, Jan Pennekamp, Klaus Wehrle, and Stefan Decker. 2021. Data Reliability and Trustworthiness Through Digital Transmission Contracts. In *The Semantic Web*. Vol. 12731. Springer International Publishing, Cham, 265–283. https://doi.org/10.1007/978-3-030-77385-4_16

[56] Antonio La Marra, Fabio Martinelli, Paolo Mori, and Andrea Saracino. 2019. A Distributed Usage Control Framework for Industrial Internet of Things. In *Security and Privacy Trends in the Industrial Internet of Things*, Cristina Alcaraz (Ed.). Springer International Publishing, Cham, 115–135. https://doi.org/10.1007/978-3-030-12330-7_6

[57] Roman Matzutt, Dirk Müllmann, Eva-Maria Zeissig, Christiane Horst, Kai Kasugai, Sean Lidynia, Simon Wieninger, Jan Henrik Ziegeldorf, Gerhard Gudergan, Indra Spiecker gen. Döhmann, Klaus Wehrle, and Martina Ziefle. 2017. myneData: Towards a Trusted and User-controlled Ecosystem for Sharing Personal Data. (2017). https://doi.org/10.18420/IN2017_109

[58] Roman Matzutt, Jan Pennekamp, and Klaus Wehrle. 2020. A Secure and Practical Decentralized Ecosystem for Shareable Education Material. In *2020 International Conference on Information Networking (ICOIN)*. IEEE, Barcelona, Spain, 529–534. https://doi.org/10.1109/ICOIN48656.2020.9016478

[59] David McCabe and Adam Satariano. 2022. The Era of Borderless Data Is Ending. *New York Times* (2022).

[60] Nadja Menz, Aleksei Resetko, and Boris Otto. 2019. *Framework for the IDS Certification Scheme 2.0.* Technical Report. IDSA. https://doi.org/10.5281/ZENODO.5244858

[61] Andres Munoz-Arcentales, Sonsoles López-Pernas, Alejandro Pozo, Álvaro Alonso, Joaquín Salvachúa, and Gabriel Huecas. 2019. An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. *Procedia Computer Science* 160 (2019), 590–597. https://doi.org/10.1016/j.procs.2019.11.042

[62] Satoshi Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. *Decentralized Business Review* (2008), 21260.

[63] Alexander Nilsson, Pegah Nikbakht Bideh, and Joakim Brorsson. 2020. A Survey of Published Attacks on Intel SGX. *arXiv:2006.13598 [cs]* (2020). arXiv:2006.13598 [cs]

[64] Marcelo Iury S. Oliveira and Bernadette Farias Lóscio. 2018. 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, 1–9. https://doi.org/10.1145/3209281.3209335

[65] Boris Otto. 2019. Interview with Reinhold Achatz on "Data Sovereignty and Data Ecosystems". *Business & Information Systems Engineering* 61, 5 (2019), 635–636. https://doi.org/10.1007/s12599-019-00609-z

[66] Boris Otto and Matthias Jarke. 2019. Designing a Multi-Sided Data Platform: Findings from the International Data Spaces Case. *Electronic Markets* 29, 4 (2019), 561–580. https://doi.org/10.1007/s12525-019-00362-x

[67] Boris Otto, Sebastian Steinbuss, Andreas Teuscher, and Steffen Lohmann et al. 2019. IDS Reference Architecture Model (Version 3.0).

[68] Jaehong Park and Ravi Sandhu. 2004. The UCON $_{ABC}$ Usage Control Model. *ACM Transactions on Information and System Security* 7, 1 (2004), 128–174. https://doi.org/10.1145/984334.984339

[69] B. Parno, J. Howell, C. Gentry, and M. Raykova. 2013. Pinocchio: Nearly Practical Verifiable Computation. In *2013 IEEE Symposium on Security and Privacy*. IEEE, Berkeley, CA, 238–252. https://doi.org/10.1109/SP.2013.47

[70] Thomas Pasquier, Jean Bacon, Jatinder Singh, and David Eyers. 2016. Data-Centric Access Control for Cloud Computing. In *Proceedings of the 21st ACM on Symposium on Access Control Models and Technologies*. ACM, Shanghai China, 81–88. https://doi.org/10.1145/2914642.2914662

[71] Vítor Pedreira, Daniel Barros, and Pedro Pinto. 2021. A Review of Attacks, Vulnerabilities, and Defenses in Industry 4.0 with New Challenges on Data Sovereignty Ahead. *Sensors* 21, 15 (2021), 5189. https://doi.org/10.3390/s21155189

[72] Jan Pennekamp, Erik Buchholz, Yannik Lockner, Markus Dahlmanns, Tiandong Xi, Marcel Fey, Christian Brecher, Christian Hopmann, and Klaus Wehrle. 2020. Privacy-Preserving Production Process Parameter Exchange. In *Annual Computer Security Applications Conference*. ACM, Austin USA, 510–525. https://doi.org/10.1145/3427228.3427248

[73] Jan Pennekamp, Rene Glebke, Martin Henze, Tobias Meisen, Christoph Quix, Rihan Hai, Lars Gleim, Philipp Niemietz, Maximilian Rudack, Simon Knape, Alexander Epple, Daniel Trauth, Uwe Vroomen, Thomas Bergs, Christian Brecher, Andreas Buhrig-Polaczek, Matthias Jarke, and Klaus Wehrle. 2019. Towards an Infrastructure Enabling the Internet of Production. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. IEEE, Taipei, Taiwan, 31–37. https://doi.org/10.1109/ICPHYS.2019.8780276

[74] Alexander Pretschner, Manuel Hilty, Florian Schütz, Christian Schaefer, and Thomas Walter. 2008. Usage Control Enforcement: Present and Future. *IEEE Security & Privacy Magazine* 6, 4 (2008), 44–53. https://doi.org/10.1109/MSP.2008.101

[75] Tobias Pulls, Roel Peeters, and Karel Wouters. 2013. Distributed Privacy-Preserving Transparency Logging. In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*. ACM, Berlin Germany, 83–94. https://doi.org/10.1145/2517840.2517847

[76] Zhiguang Qin, Hu Xiong, Shikun Wu, and Jennifer Batamuliza. 2016. A Survey of Proxy Re-Encryption for Secure Data Sharing in Cloud Computing. *IEEE Transactions on Services Computing* (2016), 1–1. https://doi.org/10.1109/TSC.2016.2551238

[77] Ansar Rafique, Dimitri Van Landuyt, Emad Heydari Beni, Bert Lagaisse, and Wouter Joosen. 2021. CryptDICE: Distributed Data Protection System for Secure Cloud Data Storage and Computation. *Information Systems* 96 (2021), 101671. https://doi.org/10.1016/j.is.2020.101671

[78] Fahmida Y Rashid. 2020. The Rise of Confidential Computing: Big Tech Companies Are Adopting a New Security Model to Protect Data While It's in Use-[News]. *IEEE Spectrum* 57, 6 (2020), 8–9.

[79] Mark D. Ryan. 2014. Enhanced Certificate Transparency and End-to-End Encrypted Mail. In *Proceedings 2014 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2014.23379

[80] Sajin Sasy, Sergey Gorbunov, and Christopher W. Fletcher. 2018. ZeroTrace : Oblivious Memory Primitives from Intel SGX. In *Proceedings 2018 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2018.23239

[81] Martin Schanzenbach. 2020. *Towards Self-sovereign, Decentralized Personal Data Sharing and Identity Management.* Ph.D. Dissertation.

[82] James Scheibner, Jean Louis Raisaro, Juan Ramón Troncoso-Pastoriza, Marcello Ienca, Jacques Fellay, Effy Vayena, and Jean-Pierre Hubaux. 2021. Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis. *Journal of Medical Internet Research* 23, 2 (2021), e25120. https://doi.org/10.2196/25120

[83] Moritz Schneider, Ramya Jayaram Masti, Shweta Shinde, Srdjan Capkun, and Ronald Perez. 2022. SoK: Hardware-supported Trusted Execution Environments. arXiv:2205.12742 [cs]

[84] Jian Shen, Tianqi Zhou, Debiao He, Yuexin Zhang, Xingming Sun, and Yang Xiang. 2019. Block Design-Based Key Agreement for Group Data Sharing in Cloud Computing. *IEEE Transactions on Dependable and Secure Computing* 16, 6 (2019), 996–1010. https://doi.org/10.1109/TDSC.2017.2725953

[85] Ming-Wei Shih, Sangho Lee, Taesoo Kim, and Marcus Peinado. 2017. T-SGX: Eradicating Controlled-Channel Attacks Against Enclave Programs. In *Proceedings 2017 Network and Distributed System Security Symposium*. Internet Society, San Diego, CA. https://doi.org/10.14722/ndss.2017.23193

[86] Sebastian Steinbuss and et. al. 2021. Usage Control in the International Data Spaces.

[87] Smitha Sundareswaran, Anna Squicciarini, and Dan Lin. 2012. Ensuring Distributed Accountability for Data Sharing in the Cloud. *IEEE Transactions on Dependable and Secure Computing* 9, 4 (2012), 556–568. https://doi.org/10.1109/TDSC.2012.26

[88] Oliver Voß. 2021. Catena-X: Datenstandards Für Die Autobranche. *Tagesspiegel Background Digitalisierung & KI* (2021).

[89] Brent Waters. 2011. Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization. In *Public Key Cryptography – PKC 2011*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Dario Catalano, Nelly Fazio, Rosario Gennaro, and Antonio Nicolosi (Eds.). Vol. 6571. Springer Berlin Heidelberg, Berlin, Heidelberg, 53–70. https://doi.org/10.1007/978-3-642-19379-8_4

[90] Nina Leonie Weisweiler, Roland Bertelmann, Peter Braesicke, Torsten Bronger, Constanze Curdt, Frank Oliver Glöckner, Sven Rank, Oliver Stegle, York Sure-Vetter, and Nicolas Villacorta. 2021. *Helmholtz Open Science Briefing: Helmholtz in der Nationalen Forschungsdateninfrastruktur (NFDI): Report des Helmholtz Open Science Forums.* Technical Report. Helmholtz Open Science Office. 184 pages pages. https://doi.org/10.48440/OS.HELMHOLTZ.030

[91] Mark D. Wilkinson et al. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3, 1 (2016), 160018. https://doi.org/10.1038/sdata.2016.18

[92] Dan Wu, Stefaan G. Verhulst, Alex Pentland, Thiago Avila, Kelsey Finch, and Abhishek Gupta. 2021. How Data Governance Technologies Can Democratize Data Sharing for Community Well-Being. *Data & Policy* 3 (2021), e14. https://doi.org/10.1017/dap.2021.13

[93] Johannes Zrenner, Frederik Oliver Möller, Christian Jung, Andreas Eitel, and Boris Otto. 2019. Usage Control Architecture Options for Data Sovereignty in Business Ecosystems. *Journal of Enterprise Information Management* 32, 3 (2019), 477–495. https://doi.org/10.1108/JEIM-03-2018-0058