Let the Computer Say NO! The Neglected Potential of Policy Definition Languages for Data Sovereignty

Jan Bartsch, Tobias Dehling, Florian Lauf, Sven Meister and Ali Sunyaev

Abstract

During interaction with today's internet services and platform ecosystems, consumer data is often harvested and shared without their consent; that is, consumers seized to be the sovereigns of their own data with the proliferation of the internet. Due to the rapid and abundant nature of interactions in today's platform ecosystems, manual consent management is impractical. To support development of semi-automated solutions for reestablishing data sovereignty, we investigate the use of policy definition languages as machine-readable and enforceable mechanisms for fostering data sovereignty. We conducted a realist

J. Bartsch (✉) · T. Dehling · A. Sunyaev
Institute of Applied Informatics and Formal Description Methods (AIFB), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
E-Mail: jan.bartsch@kit.edu

T. Dehling
E-Mail: dehling@kit.edu

A. Sunyaev
E-Mail: sunyaev@kit.edu

F. Lauf
Fraunhofer Institute for Software and System Engineering ISST, Dortmund, Germany
E-Mail: florian.lauf@isst.fraunhofer.de

S. Meister
Witten/Herdecke University Faculty of Health/School of Medicine, Witten, Germany
E-Mail: sven.meister@uni-wh.de

T. Dehling · A. Sunyaev
KASTEL Security Research Labs, Karlsruhe, Germany

M. Friedewald et al. (Hrsg.), *Selbstbestimmung, Privatheit und Datenschutz,*
DuD-Fachbeiträge, https://doi.org/10.1007/978-3-658-33306-5_22

literature review of the capabilities of policy definition languages developed for pertinent application scenarios (e.g., for access control in cloud computing). We consolidate extant literature into a framework of the chances and challenges of leveraging policy definition languages as central building blocks for data sovereignty in platform ecosystems.

Keywords

Policy languages • Data sovereignty • Digital platforms • Policy driven management • Platform ecosystems

1 Introduction

By using a single internet application, consumers usually indirectly connect themselves with a full ecosystem of third-party platforms [1–3]. For consumers, platform applications are usually provided cheap or for free; instead of requiring monetary fees, consumer data is harvested [4]. However, consumers gain little to no insight into the storage, sharing, processing, and use of their data in platform ecosystems [5, 6]. As it is today, consumers cease to be the sovereign of their data once they start using an internet application, since they lose influence over the flow of their information [4, 7]. Until consumers can influence the flow of their information, their privacy demands will not be met [8] and a sustainable data economy cannot be established [9].

The literature on data sovereignty is still evolving, predominantly driven by sociological and politically oriented literature streams [10–12]. On a technical level, little research investigates support for data sovereignty [13]. However, it is important to emphasize the technical side of data sovereignty since, by 2025, six billion consumers are expected to have a data interaction at least every eighteen seconds [14, 15]. By then, it will be impossible for humans to manually decide on consents for data sharing and handling.

To cope with the rapid digitalization, communication of consumer preferences and policies for the use of their data needs to be, at least, semi-automated and machine-enforceable. A far-spread solution for the fast handling of policies of multiple parties is the use of machine-readable languages that specify policies on a technical level to guide system components in their actions [16]. To avoid ambiguity, we use the term policy definition language (PDL) for the language and policy rule for the instantiation of a specific policy. Existing PDLs are quite diverse and address a broad spectrum of different concepts in various domains.

The quantity of available PDLs is not the issue; a sufficiently large portion of PDLs seems to work well in their target domains [16, 17]. As we see it, the more pressing challenge is to realize the chances and master the challenges of combining PDLs in a sensible way to establish data sovereignty. Accordingly, the intriguing question is why data sovereignty has not yet been established with PDLs as central building blocks: PDLs could offer exactly the functionalities consumers need for data sovereignty–they will say 'NO!' if the user wants them to and they can do it fast.

The goal of our research is to understand the key considerations for the use of PDLs as central building blocks for realizing data sovereignty in platform ecosystems. Since there are most likely already too many PDLs, we prioritize the state-of-the-art and extant solutions instead of adding language n + 1 to the ever-expanding list of PDLs. Hence, our research question is: What are the challenges and chances of using PDLs as a central building block for data sovereignty in platform ecosystems?

To answer our research question, we conducted a comprehensive literature review of PDLs [18, 19]. We harmonized extant quality criteria of PDLs and performed a thematic analysis to assess the key considerations for using PDLs to foster data sovereignty [20]. We organize our findings in abstract layers of relevant development and operation stages of PDL-supported systems with data sovereignty.

Our work contributes to the existing literature in two ways. First, we extend extant research on data sovereignty by investigating a key technology. Second, we contribute to the PDL literature by pointing out and synthesizing the desired capabilities of PDLs for establishing data sovereignty. Regardless of the field of application, PDLs are characterized by common concepts, thus, our findings support development and review of PDLs also beyond the data sovereignty context.

2 Information Flow Governance and Data Sovereignty

2.1 Policy-Based Management

Policy-based Management (PBM) emerged as computer networks grew in size and participants. To allow for flexible and dynamic adjustment of system behavior, policies were introduced to specify the conditions under which a set of predefined operations or actions can be invoked to provide desired functionality [21]. Policies allow to change the behavior of system components without changing or

reimplementing the mechanisms that execute the actions, due to the separation of management and mechanisms [22]. This is accomplished through specification of high-level requirements that are refined into low-level policies which are more related to the actual domain or device the policy applies to [16]. Today, PBM is used in multiple domains because it is faster and less error-prone than traditional, manual system management [16]. Dependent on its goals and development status, a PDL might include components beyond the language itself, for instance, a model for the deployment of policy rules, a policy execution engine and a policy decision point that intercept requests and validate them against policy rules, and monitoring and quality ensuring components [23, 24].

2.2 A Brief History of PDLs

For over two decades, PDLs for different application scenarios were developed, leading to a rich body of PDLs for different application scenarios. One of these scenarios is the use of PDLs to influence consumers' privacy perceptions, PDLs can represent a company's privacy practices, inform the company about consumer preferences for data handling, and adjust company practices to consumer preferences [25]. Privacy-focused PDLs did not find widespread adoption because they are either too complicated or too simplistic to cope with consumers' privacy preferences [26, 27]. Overall, the development of privacy PDLs shows a stronger focus on data controllers than consumers [26]. Due to the ability of some privacy PDLs to restrict access to resources, comparisons with security PDLs are common [28]. Security PDLs are specialized on access control rules [16] and less common on usage control rules, due to harder system requirements [29]. The capabilities of access control languages are especially important in large applications [16]. In large applications, a fast retrieval of policy rules, written in a standard format with formal foundations to specify event-based rules, role-based access control, and obligations are desired properties of a PDL [16]. Besides access control in large applications, security PDLs can support network management [16]. Related to security applications is the use of PDLs for trust negotiation between strangers over the internet, where the handling of credentials is a key concern [30–32].

Tim Berners-Lee's vision of a semantic web [33] spurred the search and development of PDLs for the specification of (access control) rules in multi-agent systems on the semantic web [34, 35]. Famous PDLs in this context are KAoS and Rei, which utilize ontologies as a knowledge base to reason about policy rules. No PDL has a clear edge in all situations [34]. PDLs for data protection in the semantic web have shortcomings in addressing the principle of minimal

information disclosure and the control of data after disclosure [35]. PDLs suited for data protection need to be able to specify obligations, time constraints, and to be formalizable [36]. Fulfilling all these criteria still poses a challenge for most PDLs. Other studied PDL application areas are their suitability in service-oriented systems [23] and their capabilities for access control, privacy policies and preferences, transparency, data trading, and service level agreements [17]. However, PDLs are not necessarily bound to an application area; Ponder [37] and A-PPL [38] are, for instance, holistic PDLs. In the technical context of data sovereignty, usage control architectures for business ecosystems, designed to encompass traditional access control, were studied [13]. Albeit PDLs are part of the investigated architectures and data sovereignty is addressed, this work differs from our research. We focus on one part of the architecture because PDLs are a very relevant technology suitable for (re-)establishing data sovereignty for individuals and not only for organizations.

2.3 Data Sovereignty

There are several, interchangeably used, concepts of digital sovereignty [39] that share notions of independence, control, and autonomy [10]. The essential distinction in data sovereignty concepts is who is seen as the sovereign. States are the most commonly mentioned data sovereigns, as they strive to control the data generated or passing through their national internet infrastructures [11]. Other definitions see data sovereignty as self-determined data use by companies and individuals [13, 40] or focus on individuals and society as a whole along with individuals' ability to have comprehensive knowledge on data flows [41]. For individual consumers, this translates into the ability to articulate, monitor, and enforce how to handle their data [7]. Our research builds on a consumer-centered definition of data sovereignty. Consumers should have the freedom to change the platform and migrate their data whenever desired.

However, full control of one's data and full understanding of the involved mechanisms is not practical in modern platform ecosystems. We instead define data sovereignty as consumers' ability to determine what information is and is not shared, influence the flow of their data/information, trace data flows across the involved platforms, and check for compliance with their preferences. PDLs appear to be a feasible technology to address these requirements since the governance of information flows is their key strength.

3 Study Design

To answer our research question, we conducted a realist literature review [19]. To reduce out of scope search results and narrow down the vast literature on PBM, we chose a snowballing approach [18]. A preliminary search was conducted to compile a starting set of surveys, taxonomies, typologies, and reviews targeting PDLs. Relevant literature was identified through iterative and purposive forward and backward search [42, 43]. We included conceptual and empirical scientific papers and technical reports. Preferred snowballing objects were all kinds of literature performing quality appraisals or providing requirements for PDLs. Non-English sources and sources with a sole focus on technical means for connecting platforms, technical mechanisms to enforce policies, or user behavior were excluded. In contrast to extant studies, we rather stay on an abstract level and put a strong emphasis on the capabilities and open issues of PDLs that one could interpret as requirements for building a new PDL for data sovereignty.

To establish an overview of the capabilities of extant PDLs, a thematic analysis was performed [20]. After familiarization with the identified literature, an initial coding of PDL features was performed in parallel with a harmonization of terminology and concepts. After refinement of the codes, themes representing multiple codes were derived, iteratively refined, and named. We based our themes and codes on the full spectrum of PDLs, leading to results that were too abstract to address data sovereignty. Thus, we conducted an extra refinement step to compress and filter our findings by performing another thematic analysis [20]. A code was included if a direct (e.g., a prohibition operation) or indirect (e.g., policy rules leaking information) (dis-)advantage for consumer's data sovereignty was present and labeled as chance or challenge accordingly. We excluded codes with little fit to data sovereignty despite their technical relevance for PDLs in general (e.g., trigger paradigms for policy rules). The second thematic analysis yielded overarching themes forming our model's layers and grouping the remaining themes to inform the key considerations for using PDLs to design and operate PDL-supported systems allowing for data sovereignty.

4 Results

The ability of consumers to use a PDL to say 'NO!' and to regain data sovereignty, requires a system concept that can represent such rules while ensuring the security of the system. Finally, policy rules can be entered and enforced at runtime. Table 1 offers an overview of the key considerations in these three layers.

Tab. 1 Layers where PDLs directly or indirectly support consumer data sovereignty

Layer	Key Considerations	Decisions to be made
Concept	Policy rules for data sovereignty	What policy rules supported by extant PDLs are useful to protect data sovereignty?
	Syntax and semantics	How to represent policy rules in the system?
	Supported operations	What operations should the PDL support to implement data sovereignty rules?
Security	Access control	How to ensure that not everyone can access shared data?
	Usage control	How to ensure appropriate use after access?
	Protecting policies	How to protect the information in policy rules?
Runtime	Usability	How easy is it to specify and assess policy rules?
	Storage and deployment of policy rules	How to associate a policy rule with the corresponding data?

4.1 The Concept Layer

The concept layer deals with conceptual decisions to decide what kind of policy rules are important, how to represent rules in the system, and what operations need to be expressed in the rules.

4.1.1 Policy Rules for Data Sovereignty

PDLs have been used for over two decades to communicate privacy expectations of consumers and match them with company practices [44]. The first privacy policy language was P3P, which was designed to tackle online privacy concerns of consumers that arise due to difficulties in obtaining information on the privacy practices of websites [45]. Consumers could match P3P rules with the related preference PDLs APPEL-P3P and XPref to avoid websites with P3P policy rules that mismatched their privacy preferences, data use consents, or disclosure conditions [27, 46]. Despite P3P's abandonment in 2018, provision of information on the privacy practices of a company [45] is still relevant for modern PDLs to address data protection [17, 36, 47]. Modern PDLs do not only consider purposes

for data collection, data use, and processing, but also rules for data anonymization [47], location of processing [38, 48], and data retention times that specify how long data should reside at a location [36, 47, 49]. Additionally, it is beneficial for data protection when PDLs (e.g., AAL/A-PPL) already support accountability and auditing [17, 38, 48]. Data protection PDLs are not limited to companies, for instance, the preference language YaPPL is used to formulate consumers' data protection requirements in IoT contexts [50]. Ideally, the company and consumer side is addressed in system design to realize a semi-automated exchange of consents [27, 44, 46, 51].

However, a few challenges remain. A rule for retention times must, for instance, go beyond the data itself and delete it out of log files without impeding system performance [49]. Since platform ecosystems are built to facilitate data exchange, secondary data use poses a risk for data sovereignty. Only a few PDLs are designed to support auditing, handle secondary data use, facilitate data trading, or support tracing of data flows [17]. The ability of consumers to trace the provenance of their data within and across platform ecosystems is one prerequisite to achieve data sovereignty, while auditing is necessary to ensure compliance with policy rules. Here, the challenge lies in invoking rights and claims after the data has been anonymized and transferred across multiple parties [47]. So far, LPL is the only proposed PDL that addresses these requirements. Extant literature shows that a few PDL proposals address these challenges and that solutions can be found. Chances arise from further adapting and circulating these solutions and combining them with the years of scientific knowledge published on PDLs and associated possibilities for rule specification in the privacy and data protection context. This can result in PDLs allowing to formulate rules for machine-readable exchange of consent.

4.1.2 Syntax and Semantics

Policy rules predominantly follow a declarative paradigm. They state constraints on operations and boundaries but do not describe how the system has to satisfy these constraints. From the different options to write down policy rules (e.g., via an own syntax or functional languages), standardized data description languages are often used [e.g., XML or JSON, 27, 46] and use of such languages is a commonly found PDL quality criteria in the literature. While XML is predominantly used, more recent PDLs tend to support JSON due to its less verbose notation and lower storage consumption [50, 52]. The semantics of a PDL can be based on mathematical formalisms [e.g., mathematical logic by an ex ante or post hoc formalization, 16] or on ontologies [e.g., the PDL KaOS, 34, 53]. Ontologies are the foundation of the semantic web and provide computers with access to a

structured collection of information, relations between concepts, and can include inference rules to perform automated reasoning [33]. Such semantics allow for a better analysis of the policy rules [16, 54] without needing the PDL [53], since mathematical formalisms [30, 32] and ontologies [34] allow investigation and alteration of rules independent of the particular implementation of a PDL. However, it is more difficult to implement PDLs with a mathematical formalism [54].

Depending on the syntax and semantics, PDLs show design properties that make them ideal candidates for communicating consent to heterogeneous parties, which constitutes a chance for implementing data sovereignty. PDLs were designed to facilitate interplay and dynamic changes to system components through commonly understandable rules that emphasize 'what to do' and not 'how to do' something [21, 22]. PDLs build upon XML or other standard notations that are used beyond the PDL context and allow for readability by machines and humans [16]. Formalisms can help to avoid ambiguities in the reasoning process and yield useful properties based on the underlying logical system [e.g., monotonicity, 30–32]. Understanding data handling is an important part for establishing data sovereignty [7, 41] and can be fostered by ontology-based PDLs [34]. Such PDLs can help consumers in their decision-making as they offer them a chance to understand their data's properties and, more importantly, relationships between their data. Moreover, ontologies are helpful to understand relationships between policy rules, which improves system analysis and error handling capabilities.

4.1.3 Supported Operations

After considering how to represent policy rules in computer systems, we take a deeper look at the of content of rules; in particular, at two operations that are especially useful for implementing data sovereignty: prohibitions and obligations. Prohibitions are negative authorization policies to explicitly state forbidden actions [37]. Despite sounding simple, prohibitions are not part of every PDL as they can create conflicts, rule violations, or wrong attestations of rights and increase the complexity of policy analysis, but this can be mitigated with standard error-detection techniques [24 at 9.1.7, 37]. Obligations represent a more complex operation. In the PDL-context an obligation is an action that must be performed when certain events occur [37, 49]. That is, the ability of a PDL to trigger actions prior to, during, or after data access [16, 55]: Pre-obligations are actions that the requester must perform before access [49, 55]. For instance, handling of payment requirements before enabling access to resources in digital rights management [DRM, 55] or enabling provisional authorizations [56]. Provisional authorizations are feedback-mechanisms that refer to the issued access request. For instance, the binary return message 'access denied' could be changed into the more informative

provisional authorization message 'provide a valid certificate to gain access' [56]. Peri(ongoing)-obligations need to be performed during use; otherwise, access will be revoked [55]. For instance, logging of access (attempts), performed actions, or error messages during the session [55]. Post-obligations are the most common obligations and specify actions that need to be performed after access [55]. Post-obligations can be used to support policy rules that cover legal requirements, for instance, data retention times, usage conditions, and notifications [36, 49, 55].

Considering both operations in PDL design is a chance for (re-)establishing data sovereignty for consumers. Prohibitions are a high-level and intuitive operation to specify actions users disapprove of [37], are included in modern preference languages [e.g., YaPPL, 50], and are the operation that enables consumers to say 'NO'. The different types of obligations constitute multiple chances for implementing data sovereignty or to foster platform economy. Provisional authorizations are beneficial for all users as they offer suggestions, explanations, and decision support in contrast to PDLs that just reply with uninformative binary error messages [25, 28, 32, 48, 56]. Obligations are important for sensitive data [e.g., in health IT, 55] or to implement data protection rules. However, not every PDL supports obligations to protect the data of consumers [36]. The first challenge is to specify legal obligations ('liabilities') as obligations in a policy rule since the gap between convoluted legal language and clear PDLs needs to be bridged [49]. The second challenge is to design and implement a system that can handle and enforce different obligation types [55, 57]. The policy enforcement point must, for instance, understand obligations and act accordingly [28] while being consistent with conclusions of the policy decision point [55]. Sophisticated obligation rules that can be used in practical applications are still an open issue in current research [57].

4.2 The Security-Layer

For unprotected data, all claims to data sovereignty are in vain as they can be simply ignored. To restrict unwanted leakage or manipulation of data, security mechanisms are required. While security is a vast field, we outline the key considerations for PDLs.

4.2.1 Access Control

PDLs for access control are part of access control systems for many years and are predominantly specified and executed on the company side [16]. They are used to express rules to authorize the actions a requester or other participating party can perform with resources [28, 37]. Before authorization, requesters must verify

their claimed identity, for instance, by providing evidence in form of a digital certificate [30, 31, 58]. What evidence is needed and how to submit it should be stated in policy rules. The logic behind the access control process and its policy rules are specified by access control schemes, which are usually based on roles or attributes [16, 52], dependent on application requirements.

For many years, PDLs have been successfully used to specify policy rules in access control systems [e.g., XACML for over 15 years, 24]. This provides a great chance for the implementation of data sovereignty as rules can be stated and communicated to restrict access to consumer data.

4.2.2 Usage Control

To state how sensitive data should be handled by the receiving party while allowing for dynamic changes of permissions, traditional access control is encompassed by usage control [59]. Variations of usage control that make extensive use of obligations in their policy rules are used to protect the full lifecycle of data in different applications [13, 29, 59, 60]. Usage control rules can, for instance, be expressed in the PDL OSL, a formalized language that is translatable into two common DRM-PDLs to use their enforcement mechanisms [61].

Usage control and data provenance tracking work well together [62], thereby, providing a chance with respect to controlling and tracking data flows. Approaches for implementing data sovereignty [13, 60] can build on extant usage control technologies [e.g., LUCON, 63] provided, for instance, by the International Data Spaces (IDS) initiative [40, 64]. Challenges of PDLs for usage control, besides the specification of suitable policy rules, are, especially, conflict resolution between rules [65], harder requirements for the rest of the system [29], and requirements for cryptographically trusted soft- or hardware systems [66]. Usage control is a promising chance for establishing data sovereignty, but recent developments tend to focus on the business-to-business context and need to be modified for the consumer-to-business context.

4.2.3 Sensitive Policies and Credentials

A policy rule, written in a PDL, might contain information, for instance, on credentials or processes, that can be considered sensitive and needs to be handled and released with care [30, 31, 35, 56]. For instance, security policies aim to protect the asset they are written for and do not necessarily protect the private credentials to be inferred out of the policy rules [24 at 9.2.7]. Protection of sensitive rules can be done during their specification or at runtime [30]. PDLs themselves can protect sensitive policies by using policies about policies [metapolicies; 23, 56, 58] or by

simply treating the policy rules like all other protected resources [35]. Additionally, PDLs for 'trust negotiation' can be used when policy rules or credentials cannot be fully disclosed at the beginning of an interaction [30, 32, 56, 67]. In this context, 'negotiation' is a stateful step-by-step exchange of credentials and partial results between two foreign parties that reason together about their distinct policy rules until they reach an agreement or terminate the process [32, 56, 58].

Trust negotiations can help to minimize disclosure, foster interplay, and provide consumers with the ability to demand further evidence from the other party before proceeding with the negotiation. The disadvantage is that the negotiation of trust in large distributed systems is a challenge for most PDLs [67]. Overall, the handling of sensitive policies can be considered a challenge for consumer privacy. Despite the existence of mechanisms to handle sensitive policies [31, 35], the policy must first be considered as sensitive and the mechanisms (e.g., for trust negotiation) must be implemented by administrators [24 at 9.2.7].

4.3 The Runtime-Layer

The focus of the runtime layer are the different stages for handling a policy rule, from its instantiation to execution.

4.3.1 Usability

Policy rules are created and managed by users with different backgrounds and skills. For instance, a logic-based language might require some literacy in mathematical logic. In general, a PDL needs to be readable, writable, comprehensible, and easily manageable for users with different experience levels to enable them to express a policy rule quickly and easily [26, 27, 37, 47, 51, 56]. The users' job in writing policies can be simplified by tools to read, write, modify, visualize, and analyze policy rules. For instance, the PDL KaOS [53] offers a graphical user interface for writing and applying policies and can load ontologies and check them for conflicts [34]. Ponder offers a toolkit for administrative tasks and the analysis of policy specifications and conflicts [34]. Especially for preference languages, graphical tools should be available [17]. An early tool to support consumers was Privacy Bird, a P3P user agent that indicated the fit between consumer preferences and a website's privacy policy with a color-changing icon with the option to let the consumer gain more background information [68].

User-friendly tools do not exist for all PDLs [26, 54], but should be a key consideration in system design since access to (refined) information is a prerequisite for an individual to (re-)establish data sovereignty [7, 41]. There are multiple

chances [e.g., information provision, specification, analysis, 54] where tools can be combined with PDLs (e.g., with XML bindings) to benefit consumers.

4.3.2 Storage and Deployment of Policy Rules

After policy rules are specified, they need to be deployed. This can be done centrally via a policy repository or retrieval point [24] allowing for optimization through specialized indexes [16]. The decentralized option is to stick policy rules to cryptographically secured data [69]. Sticky policies are useful for data protection rules [48]. What option to choose depends on the application.

While sticky policies appear useful for implementing data sovereignty in a platform ecosystem, they face certain challenges with respect to different versions of policy rules, user roles, and jurisdictions [70]. Additionally, policy rules can contain constraints what data should not be released together [35] or context-dependent information [71]. There could be errors or ambiguities in the reasoning about policy rules when related data is released independently of each other in a platform ecosystem without a central component that checks the logical applicability of policy rules. On the upside, the potential problems of sticky policies can be mitigated by an ecosystem that includes such central component and/or defines clear boundaries to avoid ambiguities for data in the ecosystems [e.g., by LUCON or IDS components, 40, 63, 64]. Then, sticky policies are a chance to realize data sovereignty by connecting consumer preferences with encrypted data.

5 Discussion

We identified the key considerations and corresponding chances and challenges for using PDLs to support data sovereignty in platform ecosystems. On the concept layer, extant PDL-research based on privacy and data protection offers a valuable foundation to specify policy rules that support data sovereignty. However, we could not identify a PDL that addresses all challenges impeding data sovereignty and PDLs for the consumer side (preference languages) are rather underrepresented. The syntax and semantics of PDLs are ideal for communicating consent in standardized human- and machine-readable formats that represent rules and build upon a semantic foundation. Out of the supported operations, prohibitions can be seen as a trade-off between security and practicability. Obligations can be considered as the most valuable operation for protecting data, but realizing and enforcing post-obligations in practice is challenging. Security in the form of access control can be realized by well-established PDLs. While usage

control is already considered valuable for establishing data sovereignty, it is complex in implementation and does not primarily focus on the needs of consumers. The protection of policy rules and credentials by classifying them as sensitive and protecting them accordingly (e.g., by negotiations or metapolices) should be part of security considerations. Good usability, especially for consumers, in form of assessable and easy to specify policy rules can be fostered by tool support or guidance material but is not included in every PDL. For the deployment and association of policy rules with data, a central repository or sticky policies can be used. To prevent the loss of context-sensitive information and relations between data, both approaches should be combined.

It can be assumed that data ecosystems will grow in popularity. Therefore, challenges, for instance, with secondary data use and provenance tracking, will manifest even more often and more regularly in the future, which results in the need for new mechanisms to protect data sovereignty. The identified key considerations are informed by a plethora of distinct PDLs. One could interpret these considerations as requirements for building a new PDL for data sovereignty. However, we see the development of a new holistic data sovereignty PDL with skepticism since we encountered a vast amount of abandoned PDLs in our literature review that were never used in practice. Due to the heterogeneous requirements for establishing data sovereignty, we consider improving the interplay of extant or new PDLs a more promising chance to implement data sovereignty with PDLs than the development of yet another PDL; for instance, by translating between languages [e.g., OSL to DRM PDLs, 61], refining high-level PDLs into more technical languages, or by creating bindings to established PDLs (e.g., XACML).

To address the problem of low adoption and high abandonment of PDLs, the perspective of practitioners needs to be considered; otherwise, there is a risk that PDLs suitable for (re-)establishing data sovereignty will never be used in practice. For practitioners, the adoption of PDLs for data sovereignty can be beneficial to demonstrate legal compliance, increase the usability of their platform ecosystem, or improve the interplay of participating parties (e.g., provisional authorizations). The adoption of PDLs in a data economy that is driven by a decentralized community appears also promising. PDLs could help with the communication between peers and data transfers (e.g., sticky policies), while the decentralized design prevents clustering of data under the control of one authority. Thus benefitting autonomy and data sovereignty.

Despite the various applications PDLs are used for, they are not a cure-all technology. The purpose of PDLs is to specify rules that describe policies or preferences. Without a working enforcement mechanism, those rules can be ignored or

bypassed. Since we did not focus on enforcement-mechanisms and also reviewed PDLs in early design stages without implemented enforcement-mechanisms, some of our identified chances might, for now, only exist in theory. Due to space limitations, we had to focus on the most important considerations and exclude other considerations that should not be neglected, for instance, interoperability, extensibility, scalability, context-sensitivity, and some operations [e.g., filters or delegations, 37]. Moreover, we focused on PDLs and did not address various other technologies that might be suitable for implementing data sovereignty in platform ecosystems. We could only scratch the surface for the well-studied research fields of usability and security and their mechanisms, theories, and principles, which are directly or indirectly connected with the reported key considerations.

In future research, we will further investigate ontologies as a representation of knowledge to help consumers understand their data, relationships between their data, and data provenance. Another promising research direction is the assessment of what characteristics of distributed ledger technologies [DLT, 72, 73] could be useful to enhance certain parts of data sovereignty, for instance, data provenance or use of DLT as an enforcement mechanism for policy rules. To address provenance tracking and secondary data use, the extension of extant systems, for instance, with a usage control system, a suitable preference language, and tools to support consumers seems helpful and promising to ensure free but safe movement of consumer data.

6 Conclusions

We started our manuscript with the question whether PDLs can enable consumers to let their computer say 'NO!' since this will become a necessity in the future. To extend the literature on data sovereignty on a technical level, we conducted a review of the old and vast field of PDLs. The identified key considerations for using PDLs as central building blocks to implement data sovereignty indicate certain challenges that need to be overcome. Still, solution strategies are already offered in extant literature and by implementing them, the chances PDLs offer for (re-)establishing data sovereignty outweigh the challenges they are facing. By building our layer model, we contribute towards seizing these chances and offer a building block for future PDL development and for the development of technologies that (re-)establish data sovereignty for consumers. To conclude, the answer to the question whether PDLs are a beneficial technology to (re-)establish consumers' data sovereignty is 'YES!'.

Acknowledgements This research was partially funded by the German Federal Ministry of Education and Research (BMBF) within the scope of the research project DaWID (Data-driven value creation platform for interactive, assisting service systems; funding reference number: 16SV8383). This work was supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF) and by KASTEL Security Research Labs.

References

1. Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., Shadbolt, N.: Third party tracking in the mobile ecosystem. In: Proceedings of the 10th ACM Conference on Web Science. pp. 23–31. ACM, Amsterdam, Netherlands (2018)
2. Libert, T.: An automated approach to auditing disclosure of third-party data collection in website privacy policies. In: Proceedings of the 2018 World Wide Web Conference. pp. 207–216. International World Wide Web Conferences Steering Committee, Lyon, France (2018)
3. Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C., Gill, P.: Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In: Network and Distributed Systems Security Symposium 2018. NDSS, Sab Duegi, California, USA (2018)
4. De Filippi, P., McCarthy, S.: Cloud computing: Centralization and data sovereignty. Eur. J. Law Technol. **3** (2012)
5. Sunyaev, A., Dehling, T., Taylor, P.L., Mandl, K.D.: Availability and quality of mobile health app privacy policies. J. Am. Med. Inform. Assoc. **22**, e28–e33 (2015)
6. Zuboff, S.: Big other: Surveillance capitalism and the prospects of an information civilization. Journal of Information Technology. **30**, 75–89 (2015)
7. Hummel, P., Braun, M., Augsberg, S., Dabrock, P.: Sovereignty and data sharing. ITU Journal: ICT Discoveries. **25**, (2018)
8. Westin, A.: Privacy and Freedom. Atheneum, New York (1967)
9. Ochs, C., Büttner, B., Lamla, J.: Trading social visibility for economic amenability: Data-based value Translation on a "Health and fitness platform." Sci. Technol. Human Values **46**, 480–506 (2021)
10. Couture, S., Toupin, S.: What does the notion of "sovereignty" mean when referring to the digital? New Media & Soc, **21**, 2305–2322 (2019)
11. Polatin-Reuben, D., Wright, J.: An Internet with BRICS Characteristics: Data sovereignty and the balkanisation of the Internet. In: 4th USENIX Workshop on Free and Open Communications on the Internet. USENIX Association, San Diego, California, USA (2014)
12. Amoore, L.: Cloud geographies: Computing, data, sovereignty. Prog. Hum. Geogr. **42**, 4–24 (2018)
13. Zrenner, J., Moeller, F.O., Jung, C., Eitel, A., Otto, B.: Usage control architecture options for data sovereignty in business ecosystems. J. Enterp. Inf. Manage. **32**, 477–495 (2019)
14. Culnan, M.J.: Policy to avoid a privacy disaster. Journal of the Association for Information Systems. **20**, 848–856 (2019)

15. Reinsel, D., Gantz, J., Rydning, J.: The digitization of the world from edge to core. White Paper #US44413318. Framingham: International Data Corporation (2018).
16. Han, W., Lei, C.: A survey on policy languages in network and security management. Comput. Netw. **56**, 477–489 (2012)
17. Becher, S., Gerl, A., Meier, B., Bölz, F.: Big picture on privacy enhancing technologies in e-Health: A holistic personal privacy workflow. Information. **11**, 356 (2020)
18. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering. pp. 1–10. ACM, London, England (2014).
19. Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: A typology of literature reviews. Information & Management. **52**, 183–199 (2015)
20. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**, 77–101 (2006)
21. Sloman, M., Lupu, E.: Security and management policy specification. IEEE Network **16**, 10–19 (2002)
22. Sloman, M.: Policy driven management for distributed systems. J. Netw. Syst. Manage. **2**, 333–360 (1994)
23. Phan, T., Han, J., Schneider, J.G., Erbinger, T., Rogers, T.: A survey of policy-based management approaches for service oriented systems. In: 19th Australian Conference on Software Engineering. IEEE, Perth, Australia (2008)
24. Oasis, eXtensible Access Control Markup Language (XACML) Version 3.0, https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html#_Toc325047205. Accessed: 8. 10. 2020
25. Kumaraguru, P., Cranor, L., Lobo, J., Calo, S.: A survey of privacy policy languages. In: Proceedings of the 3rd symposium on Usable privacy and security. ACM (2007)
26. Zhao, J., Binns, R., Van Kleek, M., Shadbolt, N.: Privacy languages: Are we there yet to enable user controls? In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 799–806. ACM, Montréal, Québec, Canada (2016)
27. Kasem-Madani, S., Meier, M.: Security and privacy policy languages: A survey, categorization and gap identification. arXiv:1512.00201. (2015)
28. Anderson, A.: A comparison of two privacy policy languages: EPAL and XACML. In: Proceedings of the 3rd ACM workshop on secure web services. pp. 53–60. ACM, Alexandria, Virginia, USA (2006)
29. Lazouski, A., Martinelli, F., Mori, P.: Usage control in computer security: A survey. Comput. Sci. Rev. **4**, 81–99 (2010)
30. Seamons, K.E., Winslett, M., Yu, T., Smith, B., Child, E., Jacobson, J., Mills, H., Yu, L.: Requirements for policy languages for trust negotiation. In: Proceedings Third International Workshop on Policies for Distributed Systems and Networks. pp. 68–79. IEEE, Monterey, California, USA, (2002).
31. Bertino, E., Ferrari, E., Squicciarini, A.: Trust negotiations: concepts, systems, and languages. Comput. Sci. Eng. **6**, 27–34 (2004)
32. Coi, J.D., Olmedilla, D.: A review of trust management, security and privacy policy languages. In: Proceedings of the International Conference on Security and Cryptography. pp. 483–490. INSTICC PRess, Porto, Portugal (2008)
33. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. **284**, 34–43 (2001)

34. Tonti, G., Bradshaw, J.M., Jeffers, R., Montanari, R., Suri, N., Uszok, A.: Semantic web languages for policy representation and reasoning: A comparison of KAoS, Rei, and Ponder. In: International Semantic Web Conference. pp. 419–437. Springer, Sanibel Island, Florida, USA (2003)
35. Duma, C., Herzog, A., Shahmehri, N.: Privacy in the semantic web: What policy languages have to offer. In: Eighth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'07). pp. 109–118. IEEE, Bologna, Italy (2007)
36. Leicht, J., Heisel, M.: A survey on privacy policy languages: Expressiveness concerning data protection regulations. In: 2019 12th CMI Conference on Cybersecurity and Privacy (CMI). pp. 1–6. IEEE, Copenhagen, Denmark (2019)
37. Damianou, N., Dulay, N., Lupu, E., Sloman, M.: The ponder policy specification language. In: International Workshop on Policies for Distributed Systems and Networks. pp. 18–38. Springer, Bristol, United Kingdom (2001)
38. Azraoui, M., Elkhiyaoui, K., Önen, M., Bernsmed, K., De Oliveira, A.S., Sendor, J.: A-PPL: an accountability policy language. In: Data privacy management. autonomous spontaneous security, and security assurance, pp. 319–326. Springer, Wroclaw, Poland (2014)
39. Adonis, A.A.: Critical engagement on digital sovereignty in international relations: Actor transformation and global hierarchy. Glob. J. Polit. Int. **21**, 262–282 (2019)
40. Otto, B., Auer, S., Cirullies, J., Jürjens, J., Menz, N., Schon, J., Wenzel, S.: Industrial Data Space Digitale Souveränität über Daten. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V und Industrial Data Space e. V. (2016)
41. Posch, R.: Digital sovereignty and IT-security for a prosperous society. In: Informatics in the Future. pp. 77–86. Springer, Cham (2017)
42. Harzing, A.-W.: Publish or perish. https://harzing.com/resources/publish-or-perish. Zugegriffen: 27. Mai. 2020
43. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Q. **26**, 13–23 (2002)
44. Henze, M., Hiller, J., Schmerling, S., Ziegeldorf, J.H., Wehrle, K.: Cppl: Compact privacy policy language. In: Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society. pp. 99–110. ACM, New York, NY, USA (2016)
45. Reagle, J., Cranor, L.F.: The platform for privacy preferences. Commun. ACM **42**, 48–55 (1999)
46. van de Ven, J., Dylla, F.: Qualitative privacy description language. In: Annual Privacy Forum. pp. 171–189. Springer, Frankfurt a. M. (2016)
47. Gerl, A., Bennani, N., Kosch, H., Brunie, L.: LPL, towards a GDPR-compliant privacy language: Formal definition and usage. In: Hameurlain, A., Wagner, R. (Hrsg.) Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXVII, pp. 41–80. Springer, Berlin (2018)
48. Benghabrit, W., Grall, H., Royer, J.-C., Sellami, M., Azraoui, M., Elkhiyaoui, K., Önen, M., De Oliveira, A.S., Bernsmed, K.: A Cloud Accountability Policy Representation Framework. In: Proceedings of the 4th International Conference on Cloud Computing and Services Science. pp. 489–498. SCITEPRESS, Barcelona, Spain (2014).
49. Guarda, P., Zannone, N.: Towards the development of privacy-aware systems. Inf. Softw. Technol. **51**, 337–350 (2009)

50. Ulbricht, M.-R., Pallas, F.: YaPPL-a lightweight privacy preference language for legally sufficient and automated consent provision in IoT scenarios. In: Data Privacy Management, Cryptocurrencies and Blockchain Technology. pp. 329–344. Springer (2018)
51. Becker, M.Y., Malkis, A., Bussard, L.: S4P: A generic language for specifying privacy preferences and policies. Technical Report, Microsoft Research (2010)
52. Jiang, H., Bouabdallah, A.: JACPoL: a simple but expressive JSON-based access control policy language. In: IFIP International Conference on Information Security Theory and Practice. pp. 56–72. Springer, Crete, Greece (2017)
53. Uszok, A., Bradshaw, J.M., Jeffers, R., Tate, A., Dalton, J.: Applying KAoS services to ensure policy compliance for semantic web services workflow composition and enactment. In: The Semantic Web–ISWC 2004. pp. 425–440. Springer, Hiroshima, Japan (2004)
54. Morel, V., Pardo, R.: Three dimensions of privacy policies. arXiv preprint arXiv:1908.06814. (2019)
55. Li, N., Chen, H., Bertino, E.: On practical specification and enforcement of obligations. In: Proceedings of the second ACM conference on Data and Application Security and Privacy. pp. 71–82. ACM, San Antonio, Texas, USA (2012)
56. Bonatti, P.A., Duma, C., Fuchs, N., Nejdl, W., Olmedilla, D., Peer, J., Shahmehri, N.: Semantic web policies–a discussion of requirements and research issues. In: ESWC 2006: The Semantic Web: Research and Applications. pp. 712–724. Springer, Budva, Montenegro (2006)
57. Ferguson, D., Albright, Y., Lomsak, D., Hanks, T., Orr, K., Ligatti, J.: PoCo: A Language for specifying obligation-based policy compositions. In: Proceedings of the 2020 9th International Conference on Software and Computer Applications. pp. 331–338. ACM, Langkawi, Malaysia (2020)
58. Bonatti, P.A., Olmedilla, D.: Rule-based policy representation and reasoning for the semantic web. In: Reasoning Web 2007: Reasoning Web. pp. 240–268. Springer, Dresden, Germany (2007)
59. Sandhu, R., Park, J.: Usage control: A vision for next generation access control. In: Gorodetsky, V., Popyack, L., Skormin, V. (Eds.) Second International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security, pp. 17–31. Springer, St. Petersburg, Russia (2003)
60. Gil, G., Arnaiz, A., Diez, F.J., Higuero, M.V.: Evaluation methodology for distributed data usage control solutions. In: 2020 Global Internet of Things Summit. pp. 1–6. IEEE, Dublin, Ireland (2020)
61. Hilty, M., Pretschner, A., Basin, D., Schaefer, C., Walter, T.: A policy language for distributed usage control. In: Biskup, J., López, J. (Hrsg.) Computer Security – ESORICS 2007, pp. 531–546. Springer, Dresden, Germany (2007)
62. Bier, C.: How usage control and provenance tracking get together - a data protection perspective. In: 2013 IEEE Security and Privacy Workshops. pp. 13–17. IEEE, San Francisco, California, USA (2013)
63. Schuette, J., Brost, G.S.: LUCON: Data flow control for message-based IoT systems. In: 2018 17th IEEE International Conference On Trust. Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), pp. 289–299. IEEE, New York, NY, USA (2018)

64. Otto, B., Steinbuß, S., Teuscher, A., Lohmann, S., et. al.: IDS reference architecture model version 3.0. International Data Spaces Association (2019)
65. Karafili, E., Lupu, E.C.: Enabling data sharing in contextual environments: Policy representation and analysis. In: Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies. pp. 231–238. ACM, Indianapolis, Indiana, USA (2017)
66. Pretschner, A., Hilty, M., Basin, D.: Distributed usage control. Commun. ACM **49**, 39–44 (2006)
67. Kolar, M., Fernandez-Gago, C., Lopez, J.: Policy languages and their suitability for trust negotiation. In: DBSec 2018: Data and Applications Security and Privacy XXXII. pp. 69–84. Springer, Bergamo, Italy (2018)
68. Cranor, L.F., Guduru, P., Arjula, M.: User interfaces for privacy agents. ACM Transactions on Computer-Human Interaction. **13**, 135–178 (2006)
69. Mont, M.C., Pearson, S., Bramhall, P.: Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. In: 14th International Workshop on Database and Expert Systems Applications. pp. 377–382. IEEE, Prague, Czech Republic (2003)
70. Karjoth, G., Schunter, M., Waidner, M.: Platform for enterprise privacy practices: Privacy-enabled management of customer data. In: PET 2002: Privacy Enhancing Technologies. pp. 69–84. Springer, San Francisco, California, USA (2002)
71. Kapitsaki, G.M.: Reflecting user privacy preferences in context-aware web services. In: 2013 IEEE 20th International Conference on Web Services. pp. 123–130. IEEE, Santa Clara, California, USA (2013)
72. Sunyaev, A.: Distributed ledger technology. In: Sunyaev, A. (Hrsg.) Internet computing: Principles of distributed systems and emerging internet-based technologies, pp. 265–299. Springer International Publishing, Cham (2020)
73. Kannengießer, N., Lins, S., Dehling, T., Sunyaev, A.: Trade-offs between distributed ledger technology characteristics. ACM Comput. Surv. **53**, 42:1–37 (2020)