Towards Enforceable Usage Policies for Industry 4.0

Sebastian R. Bader[1] and Maria Maleshkova[2]

[1] Fraunhofer IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
sebastian.bader@iais.fraunhofer.de
[2] University of Bonn, Endenicher Allee 19a, 53115 Bonn, Germany
maleshkova@cs.uni-bonn.de

Abstract. Controlling the usage of business-critical data is essential for every company. While the upcoming age of Industry 4.0 propagates a seamless data exchange between all participating devices, facilities and companies along the production chain, the required data control mechanisms are lacking behind. We claim that for an effective protection, both access and usage control enforcement is a must-have for organizing Industry 4.0 collaboration networks. Formalized and machine-readable policies are one fundamental building block to achieve the needed trust level for real data-driven collaborations.
We explain the current challenges of specifying access and usage control policies and outline respective approaches relying on Semantic Web of Things practices. We analyze the requirements and implications of existing technologies and discuss their shortcomings. Based on our experiences from the specification of the International Data Spaces Usage Control Language, the necessary next steps towards automatically monitored and enforced policies are outlined and research needs formulated.

Keywords: usage control · industry 4.0 · semantic web of things

1 Introduction

Industry 4.0, the disruptive development arising from new, internet-based automation and data exchange technologies, has drawn much attention in the recent times. New solutions become evident in the context of the production industry, enabling the interconnection of heterogeneous devices, machines, and facilities but also to link suppliers, manufacturers, and customers along the whole supply chain. However, current solutions are mostly limited to closed domains with strictly defined use cases, which results in isolated silos.

Consequently, only the establishment of secure and effective data protection mechanisms will provide the necessary trust level. In this paper, we explicitly do not focus on methods to establish secure communication channels or to harden a system against malicious attacks. In contrast, we present our insights in the aspect of data access and usage control and how Semantic Web technologies can make a difference if used on top of secure communication channels. To this end, we present our experiences in the development of an end-to-end usage control framework for the International Data Space (IDS) [5].

In this context we make the following contributions: (1) we present the current state of the art on formulating usage policies and their implications on usage enforcement systems, (2) describe our assumptions and the challenges in the current technology stack, (3) outline feasible solutions by relaying on Semantic Web of Things practices, and (4) conclude with a description of the existing research gaps but also depict necessary progress in the attitudes towards the topic from a business decider and legal view on usage control in Industry 4.0 settings.

2 Related Work

One of the most influential usage control models is UCON$_{ABC}$ [7]. The introduced terminology of authorizations, obligations, and conditions are taken over by many of the later publications. Together with RFC 2904 [11] and the introduction of the different *policy points*, these two works form the theoretical foundation for the later ones in the domain of usage control.

Mazzoleni et al. [3] discuss the shortcomings of OASIS eXtensible Access Control Markup Language (XACML) [4]. Even though XACML was mainly created for expressing access restrictions, Mazzoleni et al. also discuss usage control scenarios and their challenges in distributed systems. They propose an algorithm for integrating and resolving of different policies but do not focus on the various identification, integration and organizational aspects.

Pretschner and Walter [8] critically discuss the necessity of usage control negotiations. They argue that hardly any scenario actually requires real negotiation mechanisms but can be mapped to a simpler provisioning and selection of specific options. As this is true for basic offer-accept/reject patterns, we state that for real-world scenarios the therefore necessary transparency is a crucial obstacle. Therefore, because of insufficient knowledge of the opposite party, participants have to repetitively exchange proposals until a suitable solution is found.

Speiser [10] proposes a policy language based on first-order logic and RDF serializations for expressing data-related specifications. The combination of semantic vocabularies simplifies the exchange process but the referencing of data objects is not directly possible. The approach of Wenning and Kirrane [12] and the SPECIAL project goes into a similar direction. They model policies as an OWL 2 language, enabling existing reasoner for the policy evaluation process.

The Open Digital Rights Language (ODRL2.2) [2] is further able to express usage concepts through its RDF vocabulary. Still, only resources identifiable by URIs are regarded which leads to problems when data objects are not exposed by Web endpoints but also inside local systems. The specification allows expressive statements but the implications of many supported constructs are not yet sufficiently understood.

https://www.specialprivacy.eu/

3 Challenges

A formalized usage control language has several advantages. While the most obvious one is the enabling of effective enforcement of usage rules, proactively preventing misuse, an according implementation requires an extensive control regime with strict demands on the deployed IT landscape but also the related interaction processes. However, the sole provision of machine-processable descriptions of contract offers and requests already allows new search and discovery services. A commonly understood, semantically defined language enables the seamless exchange of contractual constraints, capabilities and applicable actions. Even though a comprehensive end-to-end usage control framework still has to be developed, the various enhancements by machine-interpretable policies in combination with cryptographic verification methods and transparent certification processes already increase the trustworthiness through technical approaches.

3.1 Identification

In the context of the Web, identification of resources is done by URIs and, in case of locating, URLs. In SWoT, identifying 'things' with URIs has the benefit of globally unique identifiers. Still, the distinction between the *actual* physical thing and digital representations providing information *about* the physical thing (referred to as *data documents* in the following) is essential. In the context of usage control, a policy may (a) refer to either the thing itself or (b) the documents. In the first case, the challenge is to reliably and transparently identify all data documents, in any possible format or appearance, related to the physical thing. In particular, it is a non-trivial task to connect a potentially anonymous JSON or XML with a specific resource in a situation where usually only incomplete information about its provenance are available. For case (b), any transformation or multiplication must be controlled. For instance, if an edge gateway processes and aggregates a sensor stream, the output may not have any explicit attribution to the facility it origins. Still, an informed expert of a related company can derive critical insights and may gain a competitive advantage.

3.2 Semantic and Syntactic Connectivity

Connectivity in Industry 4.0 has many aspects. The most prominent view on connectivity is the ability to connect the heterogeneous devices and systems of a production plant with each other and further services. In the case of usage control, we take the general connectivity and interoperability challenges as solved by SWoT approaches and primarily focus on the exchange of unambiguous policy formalizations. In this context, a common data format has to be specified, a shared vocabulary with a suitable expressiveness has to be defined and standardized APIs and interaction patterns are needed. Similar to proposed data interoperability by SWoT best practices, policies and contracts need to be shared and have to be unambiguously understood. Still, respective standards in particular for vocabularies and their interpretation for the usage control domain have to be formulated and implemented.

3.3 Decidability of Policy Constructs

The unambiguously interpretability of the effecting policy patterns must be guaranteed. Each statement of a policy must allow the unambiguous derivation to either ‘true’ or ‘false’ for any situation. The resolvement of sequences, requirements but also negotiations or concatenations must be defined.

Several further challenges hinder automatic decision making for usage control systems. We define several stages of increasing complexity in order to structure the topic and illustrate dependencies. At the first stage, only feature-value based comparisons (1) are regarded. Decision factors, for instance number of usages, have to be unambiguously identified and the location and request method defined. Next, the deciding parameters are provided by an external third party (2). The trustworthiness of this third party and its provisioning endpoint depends on its resilience against attacks and manipulations but also on its ability to reliably provide ‘the truth’. Examples are scenarios where a payment services confirms whether or not a payment has been received.

For the next category, crucial information is only locally available at the consumer-side (3). In particular, details of the connected back-end systems or data exchange channels affect the decision-making. In addition, role-based restrictions fall into this category as the membership of user accounts is in general too critical for publication. This is especially demanding as no company is willing, and in many cases also prohibited by anti-trust regulations, to enable full transparency of its IT landscape to an external organization.

Even more demanding are use cases where required events or states are temporally or spatially separated from the usage control system (4). For instance, the obligation to pay before using data is a very reasonable constraint for Industry 4.0 scenarios. Still, the triggering of a payment and the point in time when the payment finally arrives are usually not the same. Usage control systems must be able to react accordingly.

3.4 Trusted Context Provision

Beginning with stage (2), the usage control system needs to integrate external information. Information on made payments or other relevant events cannot be trusted if coming from an operator with related, own business interests. In order to stay independent of manipulations by either the provider- or the consumer-side, the usage control system must rely on an independent, trustworthy component. The identity of this trusted context provider must be technically verifiable and the interaction channel must be secured against manipulation attempts. Most importantly, the provided attributes have to be deployed in a transparent and reliable manner.

3.5 Policy Negotiation

Integration of systems and facilities across company boundaries is at the heart of Industry 4.0. Interactions need to be established and revoked on the fly in order to gain the necessary flexibility in the creation of workflows. We claim that manually negotiated, designed, and maintained connections are not efficient enough for such use cases. Consequently, systems need to become autonomous to some degree.

Still, the sovereignty and protection of its IT landscape is vital for every participant in an Industry 4.0 network. However, the non-resolvable information asymmetry between different companies requires bidirectional negotiations on the level of abilities and requirements. For instance, a data provider cannot reasonably restrict data usage down to the level of used software. It would require complete transparency and control other the consumers IT network. In a typical, decentralized Industry 4.0 setting, with several equally empowered companies, such requirements are unacceptable for the other companies. On the other hand, requesting certain system *abilities*, like e.g. protected data access or non-proliferation of information, is certainly reasonable and gives the opposite party the necessary options to adjust its systems accordingly.

3.6 Legal Impact

Formalized policies currently do not have the same legal quality as typical, full-text contracts. Currently, policies only describe the willing to cooperate in certain manners without any legally enforceable consequences. Even though data-driven developments and business models have a high priority and visibility for politics, the required legal blurriness of contract formulations. As stated before, no usage control system can work with fuzzy or vague constructs, the requirements and obligations of a contract must allow as little interpretation as possible.

4 Semantic Web-enabled Usage Control

The Web of Things comprises several conventions and best practices to target the challenges of the upcoming Industry 4.0. While its core advantages might be seen in its ability to reliably operate decentralized networks and to scale nearly without limits. Still, especially when combined with Semantic Web technologies, the Web of Things provides noteworthy propositions for the previously outlined challenges.

4.1 Identification

URIs have proven their value as global identifiers throughout the internet. They are especially useful when pointing at the corresponding data endpoint. In such cases, there the identifier actually also serves as an URL, location and identification can be solved at the same time. In the context of usage control, URIs and URLs are well suited for distant resources. That are, for instance, stable endpoints for requesting Linked Data but also remote attributes to make decisions upon. Remote attributes are necessary to formulate constraints on events and states which information upon are provided by a third party. Information about a (not) executed payment for instance might be accessible via a URL.

Another category of referents consists of pointers to information, which are only accessible inside a company network. For instance, user roles and affiliation data is usually not opened to external validation. In such scenarios, the restriction to URIs is typically not feasible. Shorter identifiers might be sufficient as long as the deployed usage control system can interpret them. Still, a respective trust level can be achieved by either additional contractual agreements or a third party certifying the correctness of the delivered information.

Sets and collections are another challenge which can be solved by a Web-based approach. Again, using URIs can allow a usage control system to dynamically request information on members. However, a recognized information provider (either external for uncritical data or internal for e.g. user permission management) needs to be operated in a long-term fashion.

In addition, an identification throughout the attributes of data objects can be feasible. Patterns defining required attribute manifestations can enable a membership assignment, similar to the identification through path patterns approach described above. For instance, a policy construct can express its applicability to a set of data assets by requiring an ID attribute in a certain value range. However, indirect assignments like this are more error-prone than explicit meta data as its adaptability to unpredicted changes of data formats or modeling schemes is lower and hard to track down.

4.2 Semantic and Syntactic Connectivity

Connectivity in the context of Industry 4.0 can be examined through several views. For this paper, connectivity is related to the unambiguous exchange of contracts, policies and the additional information resources necessary to interpret and enforce them. The Semantic Web of Things has several contributions to this challenge. First, the ability to dereference resources enables the direct linkage to context information and further descriptions. The application of RDF integrates different data formats and API requirements. Widely used conventions, like RESTful interactions, are standardized in combination with Linked Data by the Linked Data Platform recommendations.

The main advantage for a semantically defined usage control language is its self-descriptive nature. As such, it can contribute best when used as an intermediator between local languages or configurations. As such, the full expressiveness of the whole language is not necessarily supported by every local systems. Still, every participating component must be able to at least specify non-conformance when requested.

4.3 Decidability of Constructs

Depicting attributes by URLs and providing respective information on their current and past manifestation is an obvious approach for category (1). Linked Data promotes a proliferation in such a way.

Comparison operators have to be standardized. While the semantics of commonly used mathematical operators, like *equals*, *lower than*, or *greater than*, is trivially clear, the application of operators on more complex datatypes requires significantly more effort.

Furthermore for the Semantic Web, the issue of datatype comparison has not been solved yet. In order to implement a usage control system working with RDF data, the inability of comparisons like *“2”^^xsd:integer == “2.00”^^xsd:double == “2”^^xsd:string* is a major obstacle. The further propagation of RDF and Linked Data in this context requires standardized methods to solve this issue.

In cases of integrating additional resources or endpoints (see category (2)), the SWoT approaches give direct guidelines. Interacting with remote data by Linked Data or Linked Data Platform specifications standardizes the interactions and gives mature and well-known patterns. These approaches are also applicable when decision factors are only accessible at certain locations (3). Still, there is the risk that the semantics and structure of URIs or other identifiers already exposes protected information. A URI for instance might give hints to back-end components, which existence would not have been recognizable otherwise. More generally, the indirect but necessary outlining of abilities and requirements in the form of policies (or other exchanged meta data) may already constitute a risk for one of the participants.

If such actions are unavoidable through the nature of the necessary decisions, a framing contract between the respective organizations must specify the allowed interactions. In that case, one can argue that the policy formalizations are overruled by the framing contract in any case as this contract must act as the source of trust. Consequently, the necessity of human interactions limits the usability of usage control systems to the current state of explicitly regulated company-to-company interactions backed by extensive, textual contracts.

4.4 Trusted Context Provision

In a typical usage control setting, decision factors that are not observable to one party always reduce the level of trust. In order to avoid such situations, as much as possible should be hosted by a trusted party, which might also be the data provider but does not need to. Such a trusted Policy Information Point (PIP) serves as a source of truth, accepted by both parties.

As a result, we consider a trusted PIP as a necessary technical component with high requirements related to the acquiring of context information but also as an important business role in the usage control environment. One can think of several models to reimburse the provisioning of trustworthy information, similar to other infrastructure service providers like an identity provider, a certification authority and so on.

4.5 Policy Negotiation

The differing requirements and conditions of the involved parties necessarily lead to a process of aligning the respective offers, prohibitions and constraints. For long-term scenarios, the time-consuming dialog between human actors is still sufficient. However, in scenarios with data requests and consumption on the fly, (semi)autonomic agents have the potential to quickly reach agreements in defined ranges. In order to do so, the problem of agreeing on usage contracts needs a mathematical formalization with suitable utility functions and formalized constraints. To the best of our knowledge, no current system is able to translate contracts and contractual clauses into suitable equation systems.

Even though several approaches have been made to tackle the resolution of conflicting policies, we claim that there is still significant research potential. One exemplary fact contributing to this issue is that, regarding the ODRL 2.2 recommendation, prioritization of policies cannot be sufficiently modeled.

Furthermore, an automated negotiation of usage contracts requires a high degree of situation-awareness at all involved components. Even though many papers discuss steps to assign agents with this task ([6] or [9]), neither the necessary technical capability nor the willingness to grant the necessary autonomy to non-human actors can be stated in current implementations.

4.6 Legal Impact

As stated, currently policies are not sufficient to replace textual contracts. While the unambiguous description of their content has advantages for assisted searching or matchmaking, no existing system can reasonably state whether a formalized policy passes an examination at court or not. However, we identify a number of core obstacles in the related (EU-wide) legislation: The prohibition of automized decision making, the primacy of textual contracts, the necessity of interpretation of contractual clauses and the privacy regulations framed by GDPR. It is not the scope of this paper to discuss the necessary evolution of these principles but to make aware of their implications.

The prohibition of automized decision making reserves the final decision making to a human. This restriction is hard to accomplish when decisions are necessary on the fly and with high frequency. Framework agreements might help to describe general patterns but cannot be specific enough for in situ decision-making. The primacy of textual contracts refers to the fact that the final escalation stage in any business context is always a court trial. For now, legal experts will only decide on the textual version of a contract. In contrast to that, the usage control system can only decide based on the formalized policy. The crucial assumption herein is the consistency of the textual and formalized versions. If that cannot be guaranteed – and no system known to the authors is able to do so – unpredictable consequences arise.

5 The IDS Usage Control Language

The International Data Space [5] enables data sovereignty in a decentralized network of companies. As such, trustworthy data exchange and control of its usages are essential building blocks. The IDS usage control specification [1] therefore defines requirements, configurations and interactions for data exchange and collaborations between its members. In this context, the IDS Usage Control Language is a RDF vocabulary comprising an ODRL 2.2 profile for the representation of usage offers, requests and agreements. In particular, three security profiles have been defined as basic Service Level Agreements, determining both minimal requirements for the protected data exchange but also the security level of the technical IDS infrastructure in place. For instance, Level 0 requires only the compliance to the IDS connectivity specifications whereas Security Level 2 demands enhanced capabilities like active usage control engines, integrity protection mechanisms and a verifiable certification process.

While the proposals of the previous section were mainly also introduced into the IDS Usage Control Language, at several points stricter restrictions had to be made in order to avoid currently unsolved problems. The trade-off in the form of a decreased expressiveness was made on purpose in order to avoid some of the outlined pitfalls. For instance, currently no sequences of activities are supported. Even though that may change in the near future, we claim that currently the implicit and explicit consequences are not sufficiently understood.

This results in a very limited amount of supported constructs. We deliberately decided to select this strategy in order to establish a stable foundation for future extensions. For now, only rules with one possible action type and primitive, not nested constraints. Further refinement constructs as introduced by the ODLR 2.2 specification are not supported in order to restrict the complexity of the policies. In order to control these restrictions and to support users with testing and validation capabilities of their policies, SHACL constraint shapes are supplied for all major classes.

The IDS approach sets white-listing as its default behaviour. We state that this is the only reasonable way as otherwise the consequences of allowing usages that should have been restricted result in unforeseeable consequences. Restricting too many events on the other hand may break workflows but is better manageable as the affected environment is still limited. A contrary approach would drastically increase the risk of non-allowed usages and thereby unintentionally made obligations.

6 Conclusion and Outlook

The presented list of challenges depicts the obstacles we faced during the development of the IDS Usage Control Language. This process was driven by the specific requirements of the IDS and therefore can only constitute an incomplete list. However, we think that most of the drawn conclusions are relevant for usage control languages in general.

Choosing RDF vocabularies for a usage control language has many advantages (e.g. self-descriptive, information linking, clear semantics, etc.), especially for sharing and integrating policies in a Web-powered environment. However,

The IDS Usage Control Language is part of the IDS Information Model and can be accessed at `https://github.com/IndustrialDataSpace/InformationModel/`

the inability to directly specify property links, for instance with priorities or provenance information, hampers the formal representation of usage contracts.

Several of the outlined challenges are still open and require future research. These inabilities, from which we only examined the ones in the context of the usage control language, currently prevent a complete, end-to-end usage control system. Still, we want to strengthen the fact that an iterative approach already has significant benefits at every step. The ability to describe offers and requests, capabilities to automatically interpret and derive conclusions and, finally, a gradual increase of control and autonomy of the usage control systems enables new business models at each stage.

While at first proper search engines for actually usable data can be created, a more and more enforced usage control will reduce the concerns regarding misuse and fraud. We see this development as a necessary process to realizable, data-driven business cases for and beyond Industry 4.0.

References

1. Eitel, A., Jung, C., Haas, C., Mader, C., Brost, G., Schütte, J., Pullmann, J., Zrenne, J., Birnstill, P.: Usage Control in the Industrial Data Space. Tech. rep., Fraunhofer IESE (December 2017)
2. Ianella, R., Villata, S.: ODRL Information Model 2.2. Tech. rep., W3C ODRL Community Group (2018), https://www.w3.org/TR/odrl-model/
3. Mazzoleni, P., Crispo, B., Sivasubramanian, S., Bertino, E.: Xacml policy integration algorithms. Transactions on Information and System Security **11**(1) (2008)
4. Moses, T., et al.: eXtensible Access Control Markup Language (XACML). OASIS Standard (February 2005)
5. Otto, B., Lohmann, S., Steinbuss, S., Teuscher, A.: IDS REFERENCE ARCHITECTURE MODEL. Tech. rep., International Data Spaces Association (2018)
6. Park, H.A., Zhan, J., Lee, D.H.: Privacy-aware access control through negotiation in daily life service. In: International Conference on Intelligence and Security Informatics. pp. 514–519. Springer (2008)
7. Park, J., Sandhu, R.: The UCON ABC usage control model. ACM Transactions on Information and System Security (TISSEC) **7**(1), 128–174 (2004)
8. Pretschner, A., Walter, T.: Negotiation of Usage Control Policies - Simply the Best? In: 3rd International Conference on Availability, Reliability and Security. IEEE (2008)
9. Rahwan, I., Kowalczyk, R., Pham, H.H.: Intelligent agents for automated one-to-many e-commerce negotiation. In: Australian Computer Science Communications. vol. 24, pp. 197–204. Australian Computer Society, Inc. (2002)
10. Speiser, S., Studer, R.: A self-policing policy language. In: International Semantic Web Conference. pp. 730–746. Springer (2010)
11. Vollbrecht, J., Calhoun, P., Farrell, S., Gommans, L., Gross, G., de Bruijn, B., de Laat, C., Holdrege, M., Spence, D.: RFC 2904: AAA Authorization Framework. Request For Comment, Network Working Group (2000)
12. Wenning, R., Kirrane, S.: Compliance Using Metadata. In: Hoppe, T., Humm, B., Reibold, A. (eds.) Semantic Applications, pp. 31–45. Springer, Heidelberg (2018)