Towards a Purpose-Based Access Control Model Derived from the Purpose Limitation Principle

Kebede, M.G.; van Binsbergen, T.; van Engers, T.; van Vuurden, D.G.

DOI
10.3233/FAIA230958

Publication date
2023

Document Version
Final published version

Published in
Legal Knowledge and Information Systems

Legal Knowledge and Information Systems
G. Sileno et al. (Eds.)

doi:10.3233/FAIA230958

Towards a Purpose-Based Access Control Model Derived from the Purpose Limitation Principle

MILEN G. KEBEDE [a,1], THOMAS VAN BINSBERGEN [b], TOM VAN ENGERS [b], DANNIS G. VAN VUURDEN [b]

[a] *University of Amsterdam*
[b] *University of Amsterdam,TNO, Princess Maxima Center for Pediatric Oncology*
ORCiD ID: Milen G. Kebede https://orcid.org/0000-0003-4790-7024 , Thomas van Binsbergen https://orcid.org/0000-0001-8113-2221, Tom van Engers https://orcid.org/0000-0003-3699-8303, Dannis G. van Vuurden https://orcid.org/0000-0002-1364-9007

Abstract. The purpose limitation principle is a GDPR cornerstone that aims to minimize data processing risks by limiting instances of personal data access and usage. We model purpose as an action or sequences of actions and formalize action relationships to derive purpose-based permissions. Based on these permissions, we introduce a novel purpose-based access control model with a purpose matching algorithm illustrated with a healthcare research use case.

Keywords. GDPR, Purpose limitation principle, Purpose based access and usage

1. Introduction

Privacy regulations, such as the General Data Protection Regulation(GDPR) and the California Consumer Privacy Act(CCPA), impose stringent requirements for compliant data processing. The GDPR aims to empower data subjects by granting them more control over the usage of their personal data and facilitating data exchange while preserving privacy. The GDPR, as per Article 4(1), deems "any information linked to a distinctly identifiable natural person and applies to a wide array of data types, including healthcare-related data" as personal data.

Data processing that falls under the jurisdiction of the GDPR has to comply with its seven principles: lawfulness, fairness and transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability. This work focuses on the purpose limitation principle. The purpose limitation principle mandates data controllers collect data for specific, explicit, legitimate purposes (Article 5(b)). Consequently, applications and services that process personal data must adhere to specific purposes for which data subjects have consented. The evolving landscape of

[1] Corresponding Author: Milen G. Kebede, m.g.kebede@uva.nl

big data applications presents a technological challenge in achieving this goal. Hence, striking a balance between data access and evolving privacy requirements is a persistent challenge as illustrated in the following use case.

Use case. The SIOPE DIPG/DMG consortium, established to advance Diffuse Intrinsic Pontine Gliomas (DIPG) disease research, collects DIPG disease data from consortium members [1]. Members submit pseudonymized datasets from consented donors(DIPG patients) to the DIPG Registry. Researchers can access datasets given project proposals are approved by the DIPG network's executive committee based on, among other factors, projects that fall under DIPG research purposes. Consequently, data analysis programs applied on DIPG datasets must align with these research objectives. However, current purpose-based access control methods (PBAC) focus on abstract purposes that can cover an array of actions that can be performed by a user, requiring human decision-making and lacking scalability.

In this paper, we how we can use a generic norm represenation language ((e)Flint) for defining purposes and relationships between purposes, derivation of allowed and prohibited purposes based on these relationships and a purpose-based access control mechanism incorporating these formalisations.

2. Purpose Limitation Principle

Legal data processing, under the jurisdiction of the GDPR, has to comply with the six legal bases (Article 6), among which consent is the most common legal bases. A valid consent must be given freely is specific, informed, and unambiguous (Article 7, Recital32). A purpose associated with a given consent covers all processing activities under that consent. When a controller intends to use data for a different purpose than what the data subject consented to, the GDPR places a duty on the controller to inform the data subject (Article13(3)). We identify two sub-components under this principle.

Firstly, **Purpose Specification**. Article 5(b) stipulates that 'Personal data shall be collected for specific, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes.' 'Specific purpose' refers to sufficiently detailed purposes to enable the implementation of data protection safeguards. The 'explicit purpose' refers to clearly, unambiguously, and adequately expressed purposes. While 'Legitimate purposes' refers to purposes that align with applicable laws, including national laws, contractual agreements, and EU labor laws.

Secondly, **Compatible use**. In cases where the need to use data for subsequent processing arises, the controller must perform a compatibility test to assess whether the new purpose aligns with the initial one (Article 5, Recital 50). Article 6(4) articulates further details on compatibility tests, considering factors such as the relationship between the original and new purposes, the context of data collection, the nature of data, potential consequences of further processing, and the presence of appropriate safeguards.

3. Related Work

Early work on PBAC, such as Byun et al. [2], define purpose as the underlying rationale for data collection and access. In their work, purposes are organized hierarchically,

and users are authorized to access data for specific purposes. Pallas et al. [3] propose a purpose-based access control framework at the application layer, establishing a purpose hierarchy based on compatibility relationships. The Enterprise Privacy Authorization Language (EPAL) represents purposes as atomic values that model intended data processing services [4]. The Open Digital Rights Language (ODRL), a policy specification language about usage content and services, expresses access purposes through constraints [5], which are boolean expressions that refine action semantics among other functionalities.

The eXtensible Access Control Markup Language (XACML), an attribute-based access control policy language, allows purpose expressions as a condition rule by refining rule applicability [6]. De Mesellis et al. propose a declarative framework grounded in first-order temporal logic that offers precise semantics of purpose [7]. Purpose is specified with temporal logic, as a set of actions within a plan to achieve that purpose. Jafari M. et al. propose an approach where purposes are organized as abstract and concrete purposes, and access is granted when access purpose aligns with the intended resource purpose [8]. The work by [9] delves into purpose semantics, using a formalism based on planning and modified Markov Decision Processes (MDPs) to determine if action sequences align with specific purposes.

The methods and approaches discussed above are significant for specifying and enforcing purposes, although some challenges in determining purpose remain. The main challenge arises from the need for granular purposes, resulting in difficulty to map purposes to technical implementations. This is especially evident when dealing with policy languages that operate on elementary actions such as 'read' or 'write', while purposes comprise higher-level actions such as 'marketing' or 'scientific research'. PBAC treats purposes as atomic values organized within trees or graphs. Permissions are granted for individual actions, while some purposes require a sequence of actions to fulfill a purpose. The challenge highlights a need for more alignment between the conventional PBAC model and emerging data processing purposes.

Another challenge emerges in validating the authenticity purposes, whereas in PBAC, trust is placed in user-stated purposes. However, this approach can lead to insider abuse, where a user can declare false purposes, leading to data misuse. While ex-post compliance measures can be argued for addressing such violations, the issue remains: unauthorized access has occurred, posing potential GDPR fines. A third challenge centers around GDPR's requirement for compatible purposes, which to our knowledge, existing literature has yet to address adequately.

4. Purpose Based Access Control

The purpose based access control mechanism proposed in this section specifies and enforces purposes using the eFLINT language and reasoner [10]. For more about eFLINT, we refer the reader to previous work on the constructs of the language [10] and [11]. Furthermore, in this work, we use the terms 'purpose' and 'action' interchangeably, with actions considered as 'low-level' and purposes as 'high-level'.

definition 1. *A* consented purpose *is a purpose for which the data subject has given consent (GDPR Article 6.1.a). An action is permitted if it is equal to a consented purpose.*

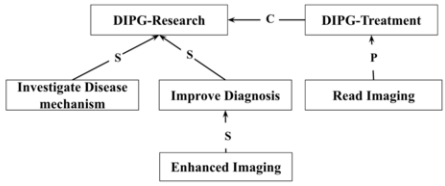

Figure 1. Relationship between DIPG research purposes

definition 2. *An* obliged purpose *is a purpose for which the processor has a legal obligation (GDPR Article 6.1.c). An action is permitted if it is equal to an obliged purpose.*

Consented and obliged purposes form base cases in a recursive search process in which arguments for lawful processing are sought, starting from a (processing) action on an asset. The following relations are used to derive permissions in this process.

definition 3. *A purpose is a* specific-of *of another purpose if it belongs to a more abstract, semantically higher level activity. Permission is generated by the specific-of relation as follows: If P_2 is a specific-of of P_1, then P_2 is permitted if P_1 is permitted.'*

definition 4. *A purpose P_1 is a* generic-of *purposes P if and only if P_1 is the most specific purpose that is more general than all the purposes in P. P_1 is permitted if all of P are consented purposes. Generic-of is derived from specific-of and the consented purposes.*

definition 5. *A purpose P_1 is a* prerequisite of *purpose P_2 if and only if P_1 needs to be performed in order to obtain P_2. P_1 is permitted if P_2 is permitted.*

definition 6. *The* compatible with *relation says that if P_1 and P_2 are compatible, then purpose P_2 is permitted if P_1 is permitted.*

Similar to the work [12], the specific-of, compatible-with and prerequisite-of relations form a 'purpose graph', defined below. An example is given in Figure 1.

definition 7. *A* purpose graph (V,S,P,C) *is a directed acyclic graph (DAG) with purposes in V labelling nodes and with three sets of edges S, P, and C corresponding to the specific-of, prerequisite-of and compatible-with relations respectively.*

The following eFLINT code specifies a purpose graph, using fact-types to represent the relations. Our example concerns a single asset `DIPGAsset` and data subject `Subject`.

```
+compatible-with(DIPGTreatment, DIPGResearch).
+specific-of(Investigate, DIPGResearch).
+specific-of(ImproveDiagnosis, DIPGResearch).
+specific-of(EnhancedImaging, ImproveDiagnosis).
+prerequisite-of(ReadImaging, DIPGTreatment).
Fact asset Identified by DIPGData
Fact subject Identified by Subject.
+subject-of(Subject, DIPGData).
```

The process constructing an argument for legal processing starts with an access request.

definition 8. Access request *is a tuple (S,A,O) where A is an action, S is the actor performing the action and O is the asset on which the action will be performed.*

Two approaches are considered for evaluating an access request, corresponding to two instantiations of the action A in the above definition. In the first approach, an access request is considered a conventional access control request where the action A is expected to be a node in the purpose graph. An action is thus a (low-level) purpose, but potentially abstract, such as 'improve diagnosis' and 'enhance imaging.' In the second approach, the action A corresponds to a submitted program to perform some processing on the asset. In this case, the purpose – i.e., the node in the purpose graph – is computed by analyzing the (source code of the) program. This procedure is left out of the scope of this work.In both cases, the same search process is applied to determine whether the request is permitted.

The following eFLINT code defines actions corresponding to two nodes in the example purpose graph. They are 'Physical', distinguishing them from institutional actions. The actions synchronize with an institutional action `process` that, when performed, corresponds to an access request as defined above. The physical actions are 'qualified' as being an instance of the institutional action and inherit its pre- and post-conditions [11].

```
Physical enhance-imaging
    Syncs with process(actor, EnhancedImaging, DIPGData)
Physical read-imaging
    Syncs with process(actor, ReadImaging, DIPGData).
```

The components of `process` are the actor performing the physical action, the node of the purpose graph associated with the physical action, and the asset processed. To experiment with our approach, a modified version of the eFLINT reasoner has been developed that gives a special treatment to the `process` action. The action is considered to be enabled (permitted), if an argument for legal processing can be found, given a specific purpose graph and a set of consented and obliged purposes (and is otherwise prohibited).

Action matching procedure Given a triple (S,A,O), forming an instance of `process` and an access request, a path of edges in the purpose graph is sought that links the action A to one of the obliged or consented purposes (for all subjects, in the case of consent). In the following scenario, the `enhance-imaging` action is considered lawful as it is more specific than the consented purpose DIPGResearch, invoking the specific-of relation twice. The `read-imaging` action is considered lawful by invoking the prerequisite-of and compatible-with relations (in that order). Below, the comments show the output of the reasoner.

```
+consent(Subject, Member, DIPGData, DIPGResearch).
enhance-imaging(Member). // Lawful:
// EnhancedImaging -s-> ImproveDiagnosis -s-> Consented(DIPGResearch)
read-imaging(Member). // Lawful:
// ReadImaging -p-> DIPGTreatment -c-> DIPGResearch
```

The following scenario shows that the generic-of relation generates a permission for a purpose (DIPGResearch) that is more general than those for which consent have been given. (Unlike in the previous scenario, DIPGResearch is not a consented purpose here.)

```
+consent(Subject, Member, DIPGData, Investigate).
+consent(Subject, Member, DIPGData, ImproveDiagnosis).
process(actor=Member, action=DIPGResearch). // Lawful:
// DIPGResearch -g-> {Investigate, ImproveDiagnosis}
```

5. Discussion and Conclusion

This paper presents an approach to formalize purposes in a fine-grained manner within access control systems, ensuring compliance with the purpose limitation principle. Policy authors can use purpose graphs to express reasons for lawful processing. In other contributions we explained the expressiveness of the eFLINT language for expressing norms. Some challenges remain however. For instance, different interpretations of purposes described in natural language texts (regulations, contracts, etc.) may exist, and deciding on the right one may require alignment processes and sometimes some form of dispute settlement and ultimately court procedures. Another limitation lies in the determination if two purposes can/should be considered 'compatible', typically relying on contextual information and human judgment. One approach is to view purposes as functions with preconditions and post-conditions and formalizing compatability as the absence of conflicts between the post-conditions of purposes.

References

[1] Joshua Baugh, Ute Bartels, James Leach, Blaise Jones, Brooklyn Chaney, Katherine E Warren, Jenavieve Kirkendall, Renee Doughman, Cynthia Hawkins, Lili Miles, et al. The international diffuse intrinsic pontine glioma registry: an infrastructure to accelerate collaborative research for an orphan disease. *Journal of neuro-oncology*, 132(2):323–331, 2017.

[2] Ji-Won Byun, Elisa Bertino, and Ninghui Li. Purpose based access control of complex data for privacy protection. In *Proceedings of the tenth ACM symposium on Access control models and technologies*, pages 102–110, 2005.

[3] Frank Pallas, Max-R Ulbricht, Stefan Tai, Thomas Peikert, Marcel Reppenhagen, Daniel Wenzel, Paul Wille, and Karl Wolf. Towards application-layer purpose-based access control. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 1288–1296, 2020.

[4] Paul Ashley, Satoshi Hada, Günter Karjoth, Calvin Powers, and Matthias Schunter. Enterprise privacy authorization language (epal). *IBM Research*, 30:31, 2003.

[5] Renato Iannella and Serena Villata. Odrl information model 2.2. *W3C Recommendation*, 15, 2018.

[6] OASIS Standard. extensible access control markup language (xacml) version 3.0. *A:(22 January 2013). URl: http://docs. oasis-open. org/xacml/3.0/xacml-3.0-core-spec-os-en. html*, 2013.

[7] Riccardo De Masellis, Chiara Ghidini, and Silvio Ranise. A declarative framework for specifying and enforcing purpose-aware policies. In *Security and Trust Management: 11th International Workshop, STM 2015, Vienna, Austria, September 21-22, 2015, Proceedings 11*, pages 55–71. Springer, 2015.

[8] Mohammad Jafari, Reihaneh Safavi-Naini, and Nicholas Paul Sheppard. Enforcing purpose of use via workflows. In *Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society*, pages 113–116, 2009.

[9] Michael Carl Tschantz, Anupam Datta, and Jeannette M Wing. Formalizing and enforcing purpose restrictions in privacy policies. In *2012 IEEE Symposium on Security and Privacy*, pages 176–190. IEEE, 2012.

[10] L. Thomas van Binsbergen, Lu-Chi Liu, Robert Van Doesburg, and Tom Van Engers. eflint: a domain-specific language for executable norm specifications. In *Proceedings of the 19th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, pages 124–136, 2020.

[11] L. Thomas van Binsbergen, Milen G Kebede, Joshua Baugh, Tom Van Engers, and Dannis G van Vuurden. Dynamic generation of access control policies from social policies. *Procedia Computer Science*, 198:140–147, 2022.

[12] Mohammad Jafari, Philip W.L. Fong, Reihaneh Safavi-Naini, Ken Barker, and Nicholas Paul Sheppard. Towards defining semantic foundations for purpose-based privacy policies. In *Proceedings of the first ACM conference on Data and application security and privacy*, pages 213–224, 2011.