Privacy-aware Linked Widgets

Javier D. Fernández
Vienna University of Economics and Business
Complexity Science Hub Vienna
javier.fernandez@wu.ac.at

Fajar J. Ekaputra, Peb Ruswono Aryan, Elmar Kiesling
Vienna University of Technology
name.surname@tuwien.ac.at

Amr Azzam
Vienna University of Economics and Business
amr.azzam@wu.ac.at

ABSTRACT

The European General Data Protection Regulation (GDPR) brings new challenges for companies, who must demonstrate that their systems and business processes comply with usage constraints specified by data subjects. However, due to the lack of standards, tools, and best practices, many organizations struggle to adapt their infrastructure and processes to ensure and demonstrate that all data processing is in compliance with users' given consent. The SPECIAL EU H2020 project has developed vocabularies that can formally describe data subjects' given consent as well as methods that use this description to automatically determine whether processing of the data according to a given policy is compliant with the given consent. Whereas this makes it possible to determine whether processing was compliant or not, integration of the approach into existing line of business applications and ex-ante compliance checking remains an open challenge. In this short paper, we demonstrate how the SPECIAL consent and compliance framework can be integrated into Linked Widgets, a mashup platform, in order to support privacy-aware ad-hoc integration of personal data. The resulting environment makes it possible to create data integration and processing workflows out of components that inherently respect usage policies of the data that is being processed and are able to demonstrate compliance. We provide an overview of the necessary meta data and orchestration towards a privacy-aware linked data mashup platform that automatically respects subjects' given consents. The evaluation results show the potential of our approach for ex-ante usage policy compliance checking within the Linked Widgets Platforms and beyond.

CCS CONCEPTS

• **Security and privacy → Information accountability and usage control**; **Privacy protections**.

KEYWORDS

Privacy; GDPR; Compliance; Linked Data

ACM Reference Format:
Javier D. Fernández, Fajar J. Ekaputra, Peb Ruswono Aryan, Elmar Kiesling, and Amr Azzam. 2019. Privacy-aware Linked Widgets. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308560.3317591

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3317591

1 INTRODUCTION

The recent European General Data Protection Regulation (GDPR) [1] defines a set of obligations for controllers and processors of personal data. Among other requirements, companies must be transparent about their processing of personal data and about its sharing within and between organizations. Furthermore, companies also need to demonstrate that their systems and business processes comply with usage constraints specified by data subjects.

Thus, the GDPR implicitly fosters requirements for more accountable systems that are aware of restrictions on personal data that is being collected, used, and shared. Designing and implementing a GDPR-compliant infrastructure that fosters privacy-by-design, is, however, a challenge. Several tools [9, 13, 14] aim to assist companies in assessing their GDPR compliance. These tools are limited in that they focus on privacy impact assessment based on standard questionnaires, but they cannot be used to automatically check compliance with subjects' policies.

Other initiatives [2, 16] use semantic web technologies to represent policies in a manner so that they can be checked automatically. In this context, the EU H2020 SPECIAL[1] project provides (i) a GDPR-oriented policy language for subject's consent, (ii) vocabularies to represent data processing events, and (iii) a complete architecture [11] for GDPR transparency and compliance.

In this paper, we focus on building *ex-ante* compliance mechanisms into data processing infrastructures. This is motivated by the need to ensure that any potential violations of subjects' restrictions on the use of their data are detected before any infringing processing occurs. To this end, we propose an approach to integrate the SPECIAL policy language and compliance checking algorithm [2] into a mashup environment – the Linked Widgets Platform (LWP) – [18] and its constituent components.

LWP is a framework for defining Linked Data-based components (e.g. data ingestion, semantification, search, integration, analysis, etc.) and to interconnect them in a processing workflow that can consume and manage data from heterogeneous sources, making use of the flexibility and expressivity of semantic technologies. The privacy-aware LWP extension, consisting of SPECIAL-based policy metadata and compliance orchestration, automatically checks that a LWP workflow is compliant with the subject's policies of the data being processed, preventing non-compliant workflows from being executed. Our initial setup and experiments with realistic data and policies in the smart energy domain, outlined in Section 3, have yielded encouraging results and demonstrate the feasibility of the ex-ante (i.e. at runtime) GDPR checking approach and the applicability of SPECIAL in existing Line of Business systems.

[1] https://www.specialprivacy.eu/

The remainder of the paper is organized as follows. Section 2 provides background on the SPECIAL framework and the LWP. Section 3 then presents our proposal towards a privacy-aware LWP, integrating the SPECIAL components to provide ex-ante GDPR compliance of the LWP workflows. Section 4 evaluates the privacy-aware LWP prototype in our case study, focused in the smart energy domain. Finally, Section 5 summarizes the state of the art, and Section 6 concludes and devises future work.

2 BACKGROUND

In this section we briefly review the SPECIAL usage policy language and compliance mechanism and provide an overview of the Linked Widget Platform, which will be extended in the next section.

2.1 SPECIAL Transparency and Compliance

The SPECIAL consent, transparency and compliance framework consists of two primary components, (i) the SPECIAL Consent component that uses the SPECIAL policy language to represent the consent from the data subject in the form of a usage policy; and the (ii) the SPECIAL Transparency and Compliance Component that uses the SPECIAL log vocabulary to represent data processing events, and the SPECIAL compliance checking mechanisms to verify compliance of such events with the usage policies. In the following we briefly present the SPECIAL usage policy and the compliance checking mechanisms. The log vocabulary[2] follows on from the usage policy and is out of the scope of this paper.

The SPECIAL Policy Language. Conceptually, a *usage policy* is meant to specify a *set of authorized operations*. According to the GDPR, these policies shall specify clearly (i) which data are collected, (ii) what is the purpose of the collection, (iii) what processing will be performed, (iv) where and for how long is the storage of the data, and (v) whether or not the data will be shared with others. The SPECIAL policy language follows these five principles and represent them using semantic technologies. Thus, a SPECIAL usage policy, P_s, is composed of one or more *basic usage policies*, each of which is an OWL 2 [7] expression of the form:

```
ObjectIntersectionOf(
 ObjectSomeValuesFrom(spl:hasData SomeDataCategory)
 ObjectSomeValuesFrom(spl:hasProcessing SomeProcessing)
 ObjectSomeValuesFrom(spl:hasPurpose SomePurpose)
 ObjectSomeValuesFrom(spl:hasRecipient SomeRecipient)
 ObjectSomeValuesFrom(spl:hasStorage SomeStorage) )
```
(1)

SPECIAL provides taxonomies that represent general categories for each case, which can be extended for particular cases. For instance, Listing 1 shows an example of a SPECIAL usage policy to represent that a subject consents to collect the energy consumption, to integrate other data sources and to perform profiling on the anonymous data in order to allow the company to optimize the energy infrastructure, who can store the data in EU indefinitely. This example makes use of the aforementioned SPECIAL auxiliary vocabularies (*spl*, *svpr*, *svl*, *svdu*, *svr*), an existing extension for cyber-physical social systems[3] (*svd-cpss*), and an exemplary vocabulary potentially defined by a company (*eg*).

[2] http://purl.org/specialprivacy/splog
[3] https://w3id.org/cityspin/ontology/special-cpss/0.1.0/index-en.html

Listing 1: Example of a SPECIAL usage policy on energy data

```
ObjectIntersectionOf(
    ObjectSomeValueFrom( spl:hasData eg:EnergyConsumption )
    ObjectSomeValueFrom( spl:hasProcessing
        ObjectIntersectionOf(
            eg:Profiling svpr:Anonymize eg:Integration svpr:Collect ))
    ObjectSomeValueFrom( spl:hasPurpose svd-cpss:Optimizing )
    ObjectSomeValueFrom( spl:hasStorage
        ObjectIntersectionOf(
            ObjectSomeValuesFrom( spl:hasLocation svl:EU )
            DataSomeValuesFrom( spl:hasDuration svdu:Indefinitely ))
    ObjectSomeValueFrom( spl:hasRecipient svr:Ours ) )
```

Using SPECIAL for Compliance Checking. In SPECIAL, policies and log events are described in semantically unambiguous terms aligned to the same taxonomies defining usage policies, hence it facilitates transparency and automatic compliance checking. Regarding this latter, the usage policy enforced by a data controller contains the operations that are permitted within the data controller's organization. Therefore, the usage U_c attached to a SPECIAL log entry *complies* with the usage policy P_s in the data subject's consent if and only if all the authorizations in U_c are also authorized by P_s, that is, U_c complies with P_s if and only if

$$U_c \subseteq P_s. \tag{2}$$

Thus, in OWL 2 terminology, this implies checking whether the following axiom is *entailed* (implied) by the combined ontology O containing the SPECIAL policy language ontology plus the aforementioned auxiliary vocabularies:

$$\texttt{SubClassOf(}U_c\ P_s\texttt{)}. \tag{3}$$

This is inherently supported by general inference engines for OWL 2 (e.g. HermiT and FaCT++). Further details of the compliance checking mechanism can be found in [2, 10].

For instance, a log entry can specify that there is a process of type `eg:SensorGathering` on location data. This entry is compliant with a potential usage policy stating that the controller can collect (`svpr:Collect`) location data *iff* `eg:SensorGathering` is a subclass of `svpr:Collect`.

This mechanism can be used for ex-post compliance checking, i.e. based on event logs, as well as ex-ante compliance, i.e. operations to be performed on the subject's data. In the next section, we focus on the latter, integrating it into the Linked Widgets Platform.

2.2 Data Mashups and Linked Widgets Platform

Mashup environments are designed to support non-expert users in combining and processing data from multiple sources to create a single new service displayed in a graphical interface [6]. In corporate settings, mashups can facilitate lightweight composition of heterogeneous enterprise applications in a shorter time to cover the long tail of user needs [8]. The term implies easy, fast integration, frequently made possible by access to open APIs and data sources to produce results beyond the predictions of the data owners [3]. This focus on ad-hoc data integration is a major strength of the mashup paradigm, but when personal data is involved, it raises significant privacy concerns. In particular, flexible and unconstrained ad-hoc integration and processing of personal data can easily clash with requirements for informed consent and transparency. Hence,

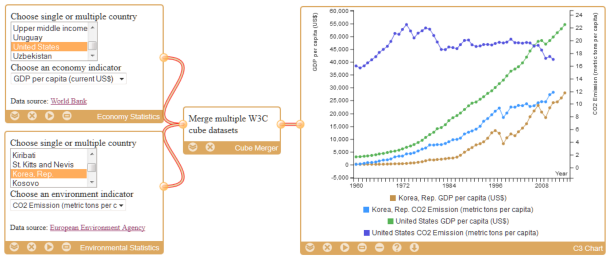

Figure 1: A mashup example on statistical data (adapted from [17])

it is important to (i) provide data subjects with the means to explicitly state the permissible use of their data, and (ii) make sure that a mashup platform that processes personal data is aware of and respects these usage policies. Given that mashups are typically constructed in an ad-hoc manner, it is necessary to build ex-ante conformance checking directly into the platform in order to ensure compliance at runtime.

We illustrate this by extending the *Linked Widgets Platform* [19] with privacy-aware mashups and processing components. This platform combines semantic web and mashup concepts to allow users to collaboratively and interactively integrate data in an ad-hoc and distributed manner. Each stakeholder can contribute their data and computing resources to a shared data processing flow. In addition, the platform facilitates both ad-hoc and persistent data integration on multiple devices.

Figure 1 illustrates a simple example mashup created in the Linked Widget Platform that integrates data on economic and environmental indicators from independent sources. The mashup consists of four widgets: the *Economic Statistics* and *Environmental Statistics* widgets provide data from the World Bank[4] and European Environment Agency[5] SPARQL endpoints, respectively. Based on the chosen parameters of countries and indicators, these widgets execute SPARQL queries. The results are processed and transformed into W3C Data Cube vocabulary[6] data sets using the StatSpace engine [4]. In each data set, the two dimensions are country and date; a single measure is the selected indicator. Next, the two data cubes are integrated by the *Cube Merger* widget and finally, the processed data is visualized by the *C3 Chart* widget.

3 THE PRIVACY-AWARE LWP

In this section, we present how the SPECIAL framework can be integrated into LWP to allow *ex-ante* compliance checking of usage policies. First, we provide a brief overview of the envisioned *ex-ante* compliance scenario (Section 3.1). Then, we detail the LWP SPECIAL-based metadata and orchestration to perform such scenario (Section 3.2), hence fostering the development of a privacy-aware LWP.

[4]http://worldbank.270a.info/sparql
[5]http://digital-agenda-data.eu/data/sparql
[6]https://www.w3.org/TR/vocab-data-cube/

3.1 *Ex-ante* Compliance: Scenario and Setup

Our case study is the smart (energy) building domain, where the building management (data controller) would like to analyze the energy consumption data of tenants (data subject) in relation with user behaviour as well as ambience data (e.g., outside temperature). The building management has access to the temperature data coming from building sensors, while the tenant provides data about their energy consumption as well as their personal profile.

Figure 2 provides an overview of the *ex-ante* compliance checking process. First, we assume a scenario where data subjects (e.g., a building tenant in our case study, a user of a company APP, etc.) give explicit consent to data controllers (e.g. the energy provider, or the company providing a service in an APP), according to the GDPR. We also assume that each data subjects' consent is represented as a usage policy, P_s, following the SPECIAL Policy Language (c.f. see Section 2.1). Automatic means of representing or obtaining such consents are out of the scope of this paper. Then, our scenario considers that the company (for simplicity, the same controller) wants to perform a data-intensive process (e.g. to analyze energy consumption patterns), making use of both personal data from the aforementioned data subjects and non-personal data (e.g., environmental sensor or other external sources). In this context, before any processing is performed, the company can build a "workflow" constraint, W_c, considering the SPECIAL dimensions, i.e., (i) the category of data to be processed, (ii) the purpose of the processing, (iii) which kind of processing will be performed, (iv) where and for how long is the storage of the data, and (v) whether or not the processing requires to share data with others. The workflow constraint will be then represented with the SPECIAL Policy Language, hence it can be then checked against the individual data subject's usage policy before the data processing is conducted (i.e., *ex-ante*) to make sure that only personal data compliant to the data controller policies being processed. Similarly to the *ex-post* scenario considering data logs, the *ex-ante* compliance checking process consists of a simple inference task to assure that:

$$\texttt{SubClassOf(}W_c\ P_s\texttt{).} \tag{4}$$

Note that *ex-ante* compliance checking can be performed a) at run-time, i.e., the first step of the data processing workflow considers an input stream of personal data and the algorithm automatically verifies that each "record" is compliant with the full workflow constraint, discarding the record otherwise, or b) as a filtering batch process, i.e., the full personal data stored is checked against the workflow constraint, and the adequate candidates are filtered and stored separately. In the following, we are agnostic of these possibilities and we assume that usage policies are performed once per usage policy. Efficient mechanisms of performing batch compliance checking (e.g. grouping usage policies based on common hierarchical elements) is devoted to future work.

3.2 Making LWP "SPECIAL"

The flexible, semantic-based LWP framework enables an efficient integration of the presented scenario for *ex-ante* compliance checking. Figure 3 shows an schematic example considering a processing workflow composed of six tasks, each of them wrapped as a single widget: Personal Data collection, Anonymization, Integration

(e.g. with building sensor data), Aggregation, Analysis (e.g. using Machine Learning) and Reporting.

The main SPECIAL extension to LWP consists of two metadata levels. First, at widget level, we semantically represent the specific processing conducted by a specific widget, and we align it with the corresponding SPECIAL vocabularies. In concrete, we specify i) the categories of the data that are processed (both as input and output of the widget), ii) the concrete process tasks and iii) the required storage (location and duration). Note that all these categories are optional, as some widgets perform general tasks or they abstract the data being processed (e.g. they do not restrict the category of data). These "semantic annotations" using the SPECIAL vocabularies are intrinsic of the widget regardless of their usage in different LWP mashups.

Then, at mashup level, i.e. for a particular processing workflow in LWP, we allow the data controller to specify the designated purpose(s) and recipient(s) of the mashup, using the corresponding SPECIAL vocabularies. Similarly, both types can be optionally present, but a more fine-grained specification enables to obtain a result for the *ex-ante* compliance checking that closely reflects the adequation to the constraints of data subjects.

Finally, the combination of widget and mashup-level metadata using the aforementioned SPECIAL vocabularies serves as the basis for constructing the data controller "workflow" constraints. It is worth mentioning that two "aggregation" methods are possible to obtain the final constraints:

- **Mashup-level constraint aggregation**. In this method, a single workflow constraint is created, representing the constraints of the entire mashup. To this end, we combine the individual categories of each widget i) applying `owl:unionOf` for the data category (i.e the mashup contains a union of the different data categories processed in the workflow), ii) using `owl:intersectionOf` for the process category (i.e. the mashup consists of a conjunction of several processes), and iii) joining the different storage periods [10]. In each category, we additionally remove classes that are sub-classes of other class.
- **Widget-level constraint aggregation**. Within this method, instead of having a single usage policy for the mashup, we provide one usage policy per widget within a mashup. To this end, we add the mashup types (purpose and recipient) to each individual widget, creating one constraint for each connected widget.

In both cases, at the end of the process, if a type is empty, we use the top level class of each type. Note that mashup-level constraint aggregation enables a direct application of the compliance checking algorithm shown above. In contrast, the algorithm should be modified to consider widget-level constraint aggregation, as special restrictions in the data subjects' policies, such as `owl:intersectionOf` relations, should be checked against different workflow constraints. Thus, in what follows, we focus on mashup-level constraint aggregations, while widget-level constraint aggregations are considered for future work.

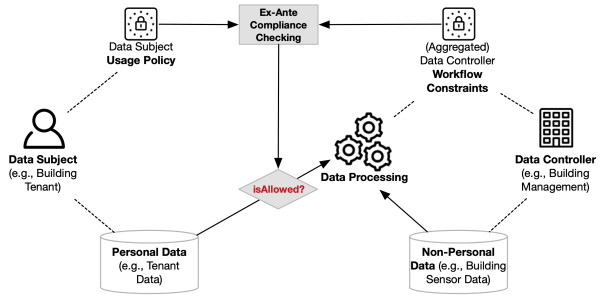

Figure 2: The overview figure of ex-ante usage-policy compliance checking

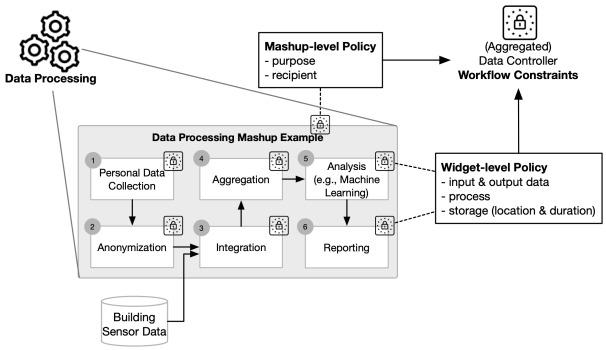

Figure 3: The components for generating data controller usage-policies

4 EVALUATION

The LWP SPECIAL extension and the corresponding ex-ante compliance checking mechanisms have been implemented in a fully functional prototype[7]. Note that, to verify the implementation, we defined a set of criteria that a privacy-aware mashup environment should fulfill:

(1) Support ex-ante compliance checking,
(2) Conformance checks must provide correct results,
(3) Ensure that no personal data is processed without given consent,
(4) Ability to support arbitrary number of heterogeneous usage policies from various data subjects,
(5) Optionally, ability to identify the specific components (or lack of) leading to consent violations.

The prototypical implementation of our privacy-aware mashup platform fulfills all the mandatory requirements, while the optional identification of problematic components is considered as future work. Note also that while we programmatically ensure that all processed data fulfill subject consents, we do not focus here on other irrevocable techniques (e.g. cryptographic methods).

In the following we first evaluate our proposal on our smart energy building scenario and the workflow depicted in Figure 3.

[7]Our privacy-aware LWP prototype and the test data are available at: https://github.com/linkeddatalab/LWP-SPECIAL

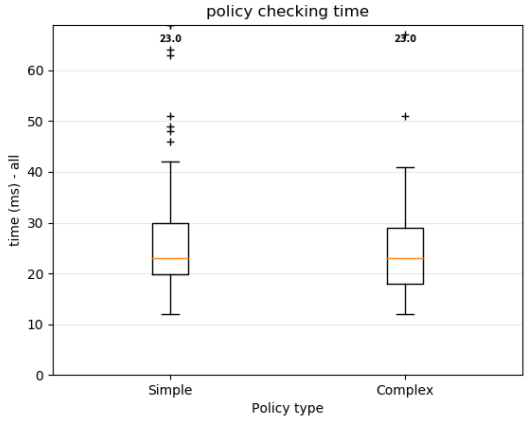

Figure 4: Results of the ex-ante compliance checking for the smart energy building workflow.

Then, we perform an scalability test on multiple workflows of different size.

All experiments run on an Ubuntu 14.05.05 Linux Server in a Intel Xeon CPU E5-2620 2GHz machine with 2GB memory.

4.1 Realistic Scenario

Following from our case study on smart energy building, we implement a LWP workflow including the components and relationships depicted in Figure 3. Table 1 shows the concrete metadata to annotate the mashup-level properties and each of the workflow components, using the SPECIAL vocabularies. Note that we split the personal data collection process into three components, which corresponds to the location, consumption data and temperature gathering processes. The last row of the table shows the final categories of the mashup-level constraint aggregation (as explained in Section 3.2).

Then, in order to test the performance of the ex-ante compliance checking on the provided workflow, we define different data subject policies. In particular, we take as input the same vocabulary used in the LWP workflow and we extend it with additional terms (for each category) in the SPECIAL auxiliary vocabularies. We then randomly generate 100 simple policies (i.e. only considering one term per category) and 100 complex policies (also considering disjunction and conjunction of terms).

The performance results of the ex-ante compliance checking, for both simple and complex policies, are shown in Figure 4. Several comments are in order. First, note that the median time to check the compliance of each policy is 23ms per policy, in both simple and complex cases. This result shows the feasibility of the system in a realistic scenario. A closer study to the difference between simple and complex policies shows that the 0.75 quantile of simple policies performs in less than 20ms, while in complex policies, the same quantile carries out the compliance checking in 18ms. This shows that the presence of very restricted policies can produce early fails in the compliance checking process and that, in general, the influence of complex policies in the performance is negligible.

4.2 Scalability Test

Our scalability test regards the ability of the approach to scale, not only to the complexity of the data subject policies, but the number of components in the workflow (i.e. the complexity of the workflow constraints). To this aim, we consider as input the same categories as our simple and complex policies, and we randomly generate and annotate random LWP workflows with 5, 10, 20, 50 and 100 components (i.e., widgets).

Figure 5 shows the performance results, for simple and complex policies, of random workflows at increasing number of processing components. The small variations in performance (a median of 23-27ms per policy, even in the case of 100 components) show that the proposed ex-ante compliance mechanisms in LWP is able to scale w.r.t the number of processing components in the mashup.

5 STATE OF THE ART

As for GDPR compliance, the Information Commissioner's Office (ICO) in the UK [9], Microsoft [13], and Nymity [14] have developed compliance tools that enable companies to assess the compliance of their applications and business processes by completing a predefined questionnaire. Recent works also look at the challenges of representing GDPR concepts and obligations [15, 16] as well as informed consent [5]. In contrast to existing approaches, SPECIAL proposes vocabularies [10] that can be used to record both usage policies and data processing and sharing events in a manner that supports automatic compliance checking.

Event management for business process compliance monitoring and process mining [12, 20] can be seen as orthogonal work.

6 CONCLUSIONS AND FUTURE WORK

The General Data Protection Regulation (GDPR) has brought the need of more transparent and accountable systems, assuring that any data process respects the recorded data subjects' consent. This work advances in this direction, providing an extension of an existing semantic data processing tool, the Linked Widget Platform (LWP), that automatically checks for ex-ante compliance w.r.t. the defined subject's consents. To this aim, we integrate the semantic policies and inference mechanisms provided in SPECIAL, an EU H2020 project on GDPR transparency and compliance. Our initial results on a realistic smart energy scenario, performing ex-ante compliance checking in 23ms per policy, shows the feasibility and efficiency of the SPECIAL and LWP combination, promoting the development of privacy-aware tools on the basis of semantic technologies.

Our proposal focuses on identifying potential hazardous processes before their execution. Irrevocable means of preventing the execution is out of the scope of the paper and is subject of future work. We also plan to provide a more fine-grain report of consent violations, identifying both problematic components in the workflow or missing pieces to fulfill the constrains. Finally, we consider to apply the privacy-aware LWP in real-world environments in the more general context of smart cities.

Table 1: SPECIAL vocabularies for the smart energy building workflow.

Component	SPECIAL category				
	Data	Processing	Purpose	Storage	Recipient
Mashup-level	-	-	svpu:Develop	-	svr:Ours
Personal Data Collection 1	svd:Location,svd:UniqueId	svpr:Transfer	-	-	-
Personal Data Collection 2	svd:Location,cpss:ConsumptionData	svpr:Transfer	-	-	-
Personal Data Collection 3	svd:Location,cpss:Temperature	svpr:Transfer	-	-	-
Anonymization	svd:Anonymized	svpr:Anonymize	-	-	-
(Geo)Integration	svd:Location	cpss:Integration	-	-	-
Aggregation	svd:Derived	svpr:Aggregate	-	-	-
Analysis	svd:Derived	svpr:Analyse	-	-	-
Report	-	-	-	svl:EU, spl:AnyDuration	-
Mashup-level Constraint Aggregation	owl:unionOf(svd:Location, svd:UniqueId, svd:ConsumptionData, cpss:Temperature, svd:Anonymized, svd:Derived)	owl:intersectionOf(svpr:Transfer, svpr:Anonymize, cpss:Integration, svpr:Aggregate, svpr:Analyze)	svpu:Develop	svl:EU, spl:AnyDuration	svr:Ours

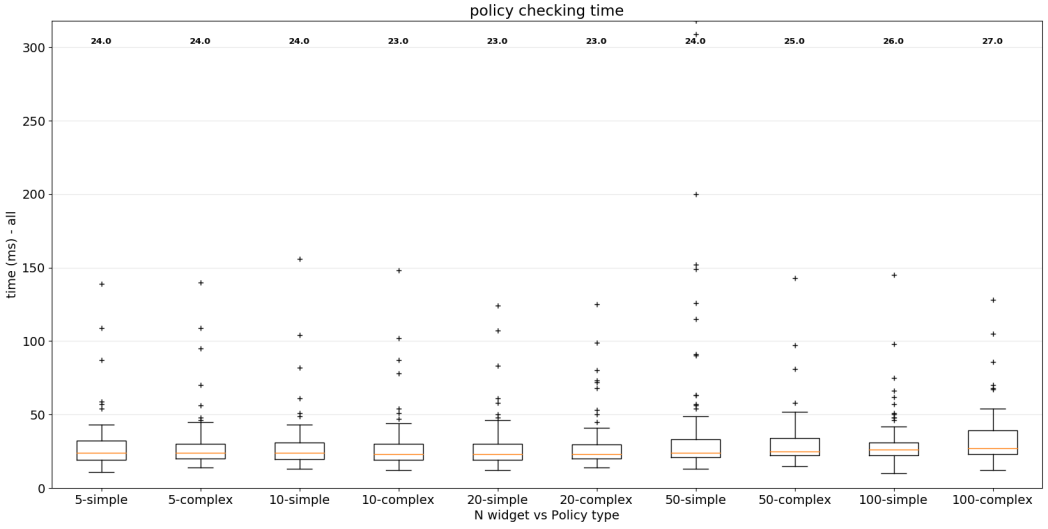

Figure 5: Results of the ex-ante compliance checking at increasing number of workflow components.

ACKNOWLEDGEMENTS

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant 731601 (SPECIAL), by the Austrian Research Promotion Agency (FFG): grant no. 861213 (CitySPIN) and 867559 (EXPEDiTE), and by the Austrian Federal Ministry for Digital, Business and Enterprise and the National Foundation for Research, Technology and Development.

REFERENCES

[1] [n. d.]. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). L119 ([n. d.]), 1–88. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[2] Piero A Bonatti. 2018. Fast Compliance Checking in an OWL2 Fragment.. In *IJCAI*. 1746–1752.

[3] Paul de Vrieze, Lai Xu, Athman Bouguettaya, Jian Yang, and Jinjun Chen. 2011. Building enterprise mashups. *Future Generation Computer Systems* 27, 5 (2011), 637 – 642. http://www.sciencedirect.com/science/article/pii/S0167739X10001974

[4] Ba-Lam Do, Peter Wetz, Elmar Kiesling, Peb Ruswono Aryan, Tuan-Dat Trinh, and A Min Tjoa. 2016. Statspace: A unified platform for statistical data exploration. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer, 792–809.

[5] Kaniz Fatema, Ensar Hadziselimovic, Harshvardhan J. Pandit, Christophe Debruyne, Dave Lewis, and Declan O'Sullivan. 2017. Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model. In *Proc of PrivOn*.

[6] Darlene Fichter. 2010. What is a mashup. In *Library mashups: Exploring new ways to deliver library data*. Information Today, Incorporated.

[7] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F Patel-Schneider, and Sebastian Rudolph. 2009. OWL 2 web ontology language primer. *W3C recommendation* 27, 1 (2009), 123.

[8] Volker Hoyer, Katarina Stanoevska-Slabeva, Till Janner, and Christoph Schroth. 2008. Enterprise Mashups: Design Principles towards the Long Tail of User Needs. In *2008 IEEE International Conference on Services Computing*, Vol. 2. IEEE Computer Society, Washington, 601–602. https://www.alexandria.unisg.ch/44891/

[9] Information Commissioner's Office (ICO) UK. 2017. Getting ready for the GDPR. (2017). https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/getting-ready-for-the-gdpr/

[10] Sabrina Kirrane, Piero Bonatti, Javier D. Fernández, Clemente Galdi, Luigi Sauro, Daniele Dell'Erba, Iliana Petrova, and Ida Siahaan. 2018. SPECIAL Deliverable 2.8: Transparency and Compliance Algorithms V2. (2018). https://www.specialprivacy.eu/images/documents/SPECIAL_D2.8_M23_V1.0.pdf

[11] Sabrina Kirrane, Javier D. Fernández, Wouter Dullaert, Uros Milosevic, Axel Polleres, Piero A. Bonatti, Rigo Wenning, Olha Drozd, and Philip Raschke. 2018. A Scalable Consent, Transparency and Compliance Architecture. In *Proc. of ESWC*. 131–136. https://doi.org/10.1007/978-3-319-98192-5_25

[12] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil MP van der Aalst. 2015. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Information systems* 54 (2015),

209–234.
[13] Microsoft Trust Center. 2017. Detailed GDPR Assessment. (2017). http://aka.ms/gdprdetailedassessment
[14] Nymity. 2017. GDPR Compliance Toolkit. (2017). https://www.nymity.com/gdpr-toolkit.aspx
[15] H.J. Pandit and D Lewis. 2017. Modelling provenance for gdpr compliance using linked open data vocabularies. In *Proc of PrivOn*.
[16] Harshvardhan J. Pandit, Declan O'Sullivan, and Dave Lewis. 2018. Queryable Provenance Metadata For GDPR Compliance. *Proc. of SEMANTICS* 137 (2018), 262 – 268. http://www.sciencedirect.com/science/article/pii/S1877050918316314
[17] Tuan-Dat Trinh, Peb R Aryan, Ba-Lam Do, Fajar J Ekaputra, Elmar Kiesling, Andreas Rauber, Peter Wetz, and A Min Tjoa. 2017. Linked data processing provenance: towards transparent and reusable linked data integration. In *Proc. of the International Conference on Web Intelligence*. ACM, 88–96.
[18] Tuan-Dat Trinh, Ba-Lam Do, Peter Wetz, Amin Anjomshoaa, and A Min Tjoa. 2013. Linked widgets: An approach to exploit open government data. In *Proc. of Information Integration and Web-based Applications & Services*. ACM, 438.
[19] Tuan-Dat Trinh, Peter Wetz, Ba-Lam Do, Elmar Kiesling, and A Min Tjoa. 2015. Distributed mashups: A collaborative approach to data integration. *IJWIS* 11, 3 (2015), 370–396.
[20] Wil MP Van der Aalst. 2011. Process Mining. In *Process Mining*. Springer, 95–123.