A Scalable Consent, Transparency and Compliance Architecture

Sabrina Kirrane[1], Javier D. Fernández[1], Wouter Dullaert[2], Uros Milosevic[2], Axel Polleres[1], Piero A. Bonatti[3], Rigo Wenning[4], Olha Drozd[1], and Philip Raschke[5]

[1] Vienna University of Economics and Business, Austria
[2] Tenforce, Belgium
[3] Universita' di Napoli Federico II, Italy
[4] W3C, Sophia-Antipolis, France
[5] Technical University Berlin, Germany

Abstract. In this demo we present the SPECIAL consent, transparency and compliance system. The objective of the system is to afford data subjects more control over personal data processing and sharing, while at the same time enabling data controllers and processors to comply with consent and transparency obligations mandated by the European General Data Protection Regulation. A short promotional video can be found at `https://purl.com/specialprivacy/demos/ESWC2018`.

1 Introduction

Data, which is commonly touted as the oil of the 21st century, is not only fueling the success of the tech giants (i.e. Google, Apple, Facebook, Amazon) but also driving innovation in enterprises in general, as evidenced by the rise in data science across a variety of domains. Although personal data is particularly valuable, in Europe the General Data Protection Regulation (GDPR) stipulates obligations with respect to personal data processing and sharing that must be fulfilled by data controllers and processors. Such obligations relate to obtaining consent from data subjects, the provision of transparency with respect to personal data processing and sharing, and ensuring compliance with usage restrictions.

Although a number of tools that focus on GDPR compliance [4, 5, 6] have recently been released, such tools are targeted at self assessment, whereby companies are given information on their obligations after completing a standard questionnaire. In contrast the system described herein can be used by companies to automatically check if existing data processing and sharing practices comply with data protection related obligations. In this demo paper, we describe the SPECIAL[6] consent, transparency and compliance system, which can be used not only to record consent but also to provide transparency to data subjects concerning the use of their personal data.

The contributions of the paper can be summarized as follows: we (i) demonstrate how usage constraints, data processing and sharing events can be expressed using the Resource Description Framework (RDF); and (ii) propose a

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

transparency and compliance system that can automatically verify that data processing and sharing complies with the relevant usage control policies.

2 Related Work

The traditional way to obtain consent is to have a human readable description of the processing where the data collected is described in some very general terms. *Dynamic consent* is a relatively new framework that refers to the use of modern communication mediums to provide transparency, enable consent management and to elicit greater involvement of data subjects from a consent perspective [3].

When it comes to transparency with respect to data processing, relevant work primarily relates to the re-purposing of existing logging mechanisms as the basis for personal data processing transparency and compliance [2]. Many of the works analyzed by Bonatti et al. [2] use a secret key signing scheme based on Message Authentication Codes (MACs) together with a hashing algorithm to generate chains of log records that are in turn used to ensure log confidentiality and integrity [1]. MACs are themselves symmetric keys that are generated and verified using collision-resistant secure cryptographic hash functions. However, only a few works [7, 8] focused on personal data processing.

As for GDPR compliance, recently the Information Commissioner's Office (ICO) in the UK [4], Microsoft [5], and Nymity [6] have developed compliance tools that enable companies to assess the compliance of their applications and business processes by completing a predefined questionnaire. In contrast to existing approaches, we propose a system that can be used to record both usage policies and data processing and sharing events in a manner that supports automatic compliance checking.

3 RDF Vocabularies for Usage Policies and Events

The vocabularies described in this section are based on the SPECIAL usage policy language[7] and log vocabulary[8], which were derived from in-depth legal analysis of use cases that require the processing and sharing of personal data for improved information and communication technology and financial services. SPECIAL usage policies can be used to denote the following information at different levels of granularity:

- *Data* describes the personal data collected from the data subject.
- *Processing* describes the operations that are performed on the personal data.
- *Purpose* represents the objective of such processing.
- *Storage* specifies where data are stored and for how long.
- *Recipients* specifies with whom the data is shared.

[7] http://purl.org/specialprivacy/policylanguage
[8] http://purl.org/specialprivacy/splog

In this paper we use the standard namespace prefixes for both `rdf` and `rdfs`, and adopt the SPECIAL vocabulary prefixes represented in *Listing* 1.1.

Listing 1.1. SPECIAL Namespace Prefixes

```
PREFIX spl: <http://www.specialprivacy.eu/langs/usage-policy#>
PREFIX splog: <http://www.specialprivacy.eu/langs/splog#>
PREFIX svd: <http://www.specialprivacy.eu/vocabs/duration#>
PREFIX svl: <http://www.specialprivacy.eu/vocabs/locations#>.
```

Usage policies. Using the SPECIAL usage policy language it is possible to specify basic usage policies as OWL classes of objects, as denoted in *Listing* 1.2 (represented using the OWL functional syntax for conciseness). Whereby the permission to perform *SomeProcessing* of *SomeDataCategory* for *SomePurpose* has been given to *SomeRecipient* in compliance with *SomeStorage* restrictions.

Listing 1.2. Structure of a Usage Control Policy

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

Data processing and sharing events. The SPECIAL policy log vocabulary is used to represent data processing and sharing events. The event log extract represented in *Listing* 1.3 (represented using turtle), relates to a new processing event corresponding to a data subject identified as `befit:Sue` on the `03.01.2018` at `13:20` (i.e., validity time). The event was recorded few seconds later (i.e., transaction time). The actual data captured can be traced via the `splog:eventContent` property, which is detailed in *Listing* 1.4, and usually stored in a separate knowledge base. While a hash of the content is stored in the event log.

Listing 1.3. A new event for Sue's BeFit device

```
befit:entry3918 a splog:ProcessingEvent;
splog:dataSubject befit:Sue;
dct:description "Store location in our database in Europe"@en;
splog:transactionTime "2018-01-10T13:20:50Z"^^xsd:dateTimeStamp;
splog:validityTime "2018-01-10T13:20:00Z"^^xsd:dateTimeStamp;
splog:eventContent befit:content3918;
splog:inmutableRecord befit:iRec3918.
```

Listing 1.4. The content of a new event for Sue's BeFit device

```
befit:content3918 a splog:LogEntryContent;
    spl:hasData svd:Location;
    spl:hasProcessing befit:SensorGathering;
    spl:hasPurpose befit:HealthTracking;
    spl:hasStorage [spl:haslocation svl:OurServers];
    spl:hasRecipient [a svr:Ours].
```

Compliance checking. In order to verify that data processing and sharing events comply with the corresponding usage policies specified by data subjects, we use OWL reasoning to decide whether the authorized operations specified by a data subject through their given consent, subsume the specific data processing records in the transparency log.

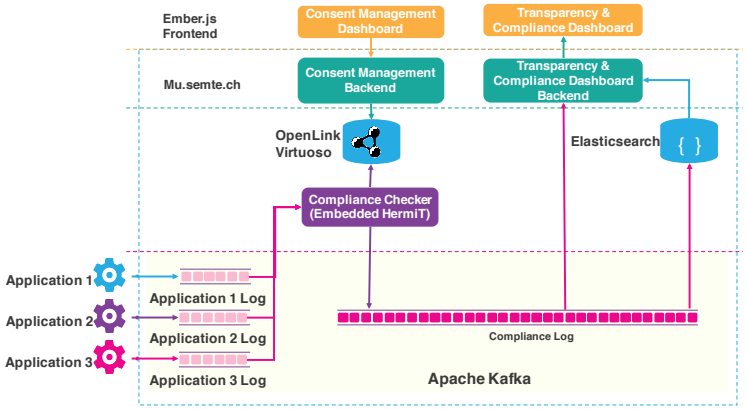

Fig. 1. A Scalable Consent, Transparency and Compliance Architecture

4 A Scalable Consent, Transparency and Compliance Architecture

The SPECIAL demo system architecture, which is depicted in *Figure* 1, enables transparency and compliance checking based on usage policies and events expressed using the aforementioned vocabularies.

Kafka and Zookeeper. Data processing and sharing event logs are stored in the Kafka[9] distributed streaming platform, which in turn relies on Zookeeper[10] for configuration, naming, synchronization, and providing group services. Each application log is represented using a distinct Kafka topic, while a separate compliance topic is used to store the enriched log after compliance checks have been completed.

Virtuoso Triple Store Based on our current use case requirements, we assume that consent updates are infrequent and as such usage policies and the respective vocabularies are represented in a Virtuoso triple store.

Compliance Checker. The compliance checker, which includes an embedded HermiT[11] reasoner uses the consent saved in Virtuoso together with the application logs provided by Kafka to check that data processing and sharing complies with the relevant usage control policies. The results of this check are saved onto a new Kafka topic.

Elasticsearch. As logs can be serialized using JSON-LD, it is possible to benefit from the faceting browsing capabilities of Elasticsearch[12] and the out of the box visualization capabilities provided by Kibana.

[9] https://kafka.apache.org/
[10] https://zookeeper.apache.org/
[11] http://www.hermit-reasoner.com/
[12] https://www.elastic.co/products/elasticsearch

Consent and Transparency & Compliance Backends. Interaction between the various architectural components is managed by mu.semte.ch[13] an open source micro-services framework for building RDF enabled applications.

Consent and Transparency & Compliance Dashboards. Users interact with the system via the consent management and the transparency and compliance dashboards. The former supports granting and revoking consent for processing/sharing. While, latter provides the data subject with transparency with respect to data processing and sharing events in a digestible manner.

5 Conclusion

The objective of this demo paper is to introduce the SPECIAL consent, transparency and compliance system, which is built around the Kafka distributed streaming platform. Future work includes the benchmarking of the various system components of the SPECIAL system from a performance and a scalability perspective, and the hardening of the system against various security attacks.

Acknowledgments. Supported by the European Union's Horizon 2020 research and innovation programme under grant 731601.

References

1. M. Bellare and B. Yee. Forward integrity for secure audit logs. Technical report, Technical report, Computer Science and Engineering Department, University of California at San Diego, 1997.
2. P. Bonatti, S. Kirrane, A. Polleres, and R. Wenning. Transparent personal data processing: The road ahead. In *International Conference on Computer Safety, Reliability, and Security*, pages 337–349. Springer, 2017.
3. I. Budin-Ljøsne, H. J. Teare, J. Kaye, S. Beck, H. B. Bentzen, L. Caenazzo, C. Collett, F. D'Abramo, H. Felzmann, T. Finlay, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. *BMC medical ethics*, 18(1):4, 2017.
4. Information Commissioner's Office (ICO) UK. Getting ready for the GDPR, 2017. URL `https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/getting-ready-for-the-gdpr/`.
5. Microsoft Trust Center. Detailed GDPR Assessment, 2017. URL `http://aka.ms/gdprdetailedassessment`.
6. Nymity. GDPR Compliance Toolkit. URL `https://www.nymity.com/gdpr-toolkit.aspx`.
7. T. Pulls, R. Peeters, and K. Wouters. Distributed privacy-preserving transparency logging. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*, 2013.
8. S. Sackmann, J. Strüker, and R. Accorsi. Personalization in privacy-aware highly dynamic systems. *Communications of the ACM*, 49(9), 2006.

[13] https://mu.semte.ch/