POLICIES, POLICY LANGUAGE AND VOCABULARY TO AUTOMATICALLY HANDLE PRIVACY ISSUES IN FUTURE PERSONAL AND PROFESSIONAL TRANSPORTATION

Dr. Martin Kurze, Deutsche Telekom

12 September, 2019

(acknowledging Sabrina Kirane, Piero Bonnattl & Rudy Jacob for their work/contribution)
ACKNOWLEDGEMENT: EU PROJECT „SPECIAL“: PROJECT PARTNERS AND CREDITS

This work is a result of H2020 project SPECIAL and was supported by

The project SPECIAL (Scalable Policy-awarE linked data arChitecture for privacy, trAnsparency and compLiance) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731601 as part of the ICT-18-2016 topic Big data PPP: privacy-preserving big data technologies.
GDPR
GLOBAL DATA-BUSINESS PREVENTER (??)

General Data Protection Regulation –
Not as bad as it might seem to the superficial observer.

To make use of personal data
(as collected during/for transportation) you need:
• Informed consent
• For a specific purpose
• Transferrable to specific 3rd party
• Transparency & control for the user

plus: data security (on device and in the cloud); plus: trust
SPECIAL Objectives

• Policy management framework
  ❖ Gives users control of their personal data
  ❖ Represents usage policies, legislative requirements and business policies in a machine readable format

• Transparency and compliance framework
  ❖ Provides information on how data is processed and with whom it is shared
  ❖ Allows companies to verify that processing is in line with data subject usage policies and legal requirements
  ❖ Allows data subjects to take corrective action

• Scalable policy-aware Linked Data architecture
  ❖ Build on top of the Big Data Europe (BDE) platform scalability and elasticity mechanisms
  ❖ Extended BDE with robust policy, transparency and compliance protocols
  ❖ Enable personal data value chains

• Pilot implementation and evaluation
  ❖ The architecture will be validated in the context of personal data sharing use cases for the telecoms and financial services sectors

• Collaboration, Dissemination & Standardisation
  ❖ Create real-world impact in the form of a sustainable solution that we disseminate actively

source: Sabrina Kirrane
Transportation is undergoing a fundamental change today:

From Mono-modality based on one vehicle owned by the person/company to “multi-modal systems of shared vehicles, orchestrated by online-transportation service providers” (like UBER 😊).

→ data is potentially PII (personally identifiable information)

→ under special legal protection

→ high value for (re)use outside the transportation field.
PROBLEM IDENTIFIED, SOLUTION IS “POSSIBLY FEASIBLE”

integrate means for

• tracking and
• controlling PII

whenever they appear in the different stages and processes of transportation.

The policy model, its components and applicability in GDPR-context and beyond (i.e. in transportation)
THE SPECIAL APPROACH: POLICY LANGUAGE, VOCABULARY AND POLICY ENGINE

The H2020 project SPECIAL has already developed a rich framework for consent management and automated compliance checking that should be of interest for most of the participants to this workshop.

Funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement N. 731601.
COMPONENTS OF THE FRAMEWORK

A policy model for data usage, specifically designed to address the requirements of the GDPR. This model covers both data subjects' consent and data controllers' data usage policies (business policies).

Vocabularies (ontologies) for the main usage policy elements, such as purposes, personal data categories, recipients, storage duration, etc, as well as GDPR's legal bases and varieties of consent. A first version of the vocabularies is being released by the W3C Data Privacy Vocabularies and Controls Community Group, promoted by SPECIAL.

An encoding of the policy model and the vocabularies in the standard OWL2. The encoding provides the model with a formal semantics - sorely needed for interoperability and for guaranteeing the coherence of compliance checking, explanations, and policy validation.

A scalable engine that can execute a compliance check every 0.5 ms without resorting to parallelism. The engine can process all expressions in a new profile of OWL2 called PL (Policy logic) that is general enough to accommodate a number of possible extensions to the current model, and follow the evolution of the vocabularies. SPECIAL developed also a parallel big data architecture for compliance checking for further speed-ups.

A transparency infrastructure where (i) data controllers can log personal data processing operations and monitor their internal processes, (ii) data subjects can enquire how their personal data are processed, (iii) data protection officers can audit personal data processing activities.
Envisaged roles of DT in tomorrow’s world transportation:

1. **Trusted 3rd party**

DT could ensure that (PII) data is transferred between partners in a secure and privacy aware way. Using the SPECIAL approach, DT could also apply user-defined policies for data sharing. E.g., DT could strip PII datasets down to a level compliant with the user’s desire (formulated as policy). This of course requires advanced and interoperable vocabularies and ontologies.

2. **Service Provider**

telecommunications companies have the need/opportunity to monetize this data as other ("OTT") companies do anyway. These large data sets will become even larger and more valuable with upcoming technologies such as “5G”.
MAIN LEGAL BASE OF UTILIZING PII:

- General Data Protection Regulation (GDPR)
Analysing & Modelling the GPDR
Lessons learned

ODRL was heavily guided by licensing use cases.....

Therefore when modelling regulatory requirements we need to decide how closely we stick to the original model
Usage policy language
The minimal core model

Collected data ➔ Purpose of data collection and processing ➔ How is data processed ➔ Where are collected data and profiles stored ➔ For how long are the data stored ➔ Disclosure to third parties

Transparency Ledger

Who ➔ What

Why ➔ When

How ➔ Where

Telekom Innovation Laboratories

[source: Sabrina Kirrane]
Usage policy language
Syntax and expressivity

- Usage policy language, which can be used to express both the data subjects’ consent, data controllers usage requests, fragments of the GDPR, and business policies
- The foundation of the policy language was the Minimal Core Model (MCM)
- We propose a new policy language that extensively re-uses standards based privacy-related vocabularies
- We are able to leverage existing Web Ontology Language (OWL) based reasoners out of the box

[source: Sabrina Kirrane]
Usage policy language
SPECIAL resources

The SPECIAL Usage Policy Language
version 0.1

Unofficial Draft 06 April 2018

Editor:
Javier D. Fernández (Vienna University of Economics and Business)

Authors:
Piero Bonatti (Università di Napoli Federico II)
Sabrina Kirrane (Vienna University of Economics and Business)
Ivana Minerva Petrosa (Università di Napoli Federico II)
Luigi Gaioni (Università di Napoli Federico II)
Eva Strehlein (Ernst-Moritz-Arndt-Universität zu Greifswald)

This document is licensed under a Creative Commons Attribution 3.0 Unported License.

Abstract
This document specifies usage policy language of SPECIAL both the data subjects’ consent and the data usage policy by a computer so as to automatically verify that the usage of data is in accordance with the requirements specified in the data usage policy.

The ontology defined in this document is publicly available.

Vocabulary .../langs/usage-policy#

Bert Bos © Last Updated: 17 April 2018

• Detailed in D2.1 Policy Language V1
• Available for download via the SPECIAL website https://www.specialprivacy.eu/langs/usage-policy
• An unofficial draft specification has been published online http://purl.org/specialprivacy/policylanguage
• Feeds into the standardisation efforts conducted in the W3C Data Privacy Vocabularies and Controls Community Group

Provenance/event information

The model

- Development of a log vocabulary that reuses well-known vocabularies such as PROV for representing provenance metadata
- Demonstrate how provenance can be used to support transparency in data value chains

A Scalable Consent, Transparency and Compliance Architecture
Sabrina Kirrane, Javier D. Fernández, Wouter Dullaert, Uros Milosevic, Axel Polleres, Piero Bonatti, Rigo Wenning, Olha Drozd and Philip Raschke
Proceedings of the Posters and Demos Track of the Extended Semantic Web Conference (ESWC 2018)
Provenance/event information
Syntax and expressivity

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

```log
befit:entry3918 a splug: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:immutableRecord befit:iRec3918.
```

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

```log
befit:content3918 a splug:LogEntryContent;
  splug:hasData svid:Location;
  splug:hasProcessing befit:SensorGathering;
  splug:hasPurpose befit:HealthTracking;
  splug:hasStorage [splug:haslocation svid:OurServers];
  splug:hasRecipient [a svid:Ours].
```

A Scalable Consent, Transparency and Compliance Architecture
Sabrina Kirrane, Javier D. Fernández, Wouter Dullaert, Uros Milosevic, Axel Polleres, Piero Bonatti, Rigo Wenning, Olha Drozd and Philip Raschke
Proceedings of the Posters and Demos Track of the Extended Semantic Web Conference (ESWC 2018)

Telekom Innovation Laboratories
Transparency and compliance checking platforms

- Data processing and sharing event logs are stored in the Kafka distributed streaming platform, which in turn relies on Zookeeper for configuration, naming, synchronization, and providing group services.

- We assume that consent updates are infrequent and as such usage policies and the respective vocabularies are represented in a Virtuoso triple store.

- The compliance checker, which includes an embedded

  A HermiT 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.

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

A Scalable Consent, Transparency and Compliance Architecture
Sabrina Kirrane, Javier D. Fernández, Wouter Dullaert, Uros Milosevic, Axel Polleres, Piero Bonatti, Rigo Wenning, Olha Drozd and Philip Raschke
Proceedings of the Posters and Demos Track of the Extended Semantic Web Conference (ESWC 2018)
Provenance/event information

- In D2.3 we frame the SPECIAL policy, transparency and compliance components within the wider scope of a general Enterprise setting.

- SPECIAL can be used in conjunction with existing Line of Business (LOB) and in Business Intelligence (BI) / Data Science (DS) settings.

- Key role of the Personal Data processing Inventory.

[source: Sabrina Kirrane]
CONCLUSIONS AND INVITATION TO COLLABORATE

Transportation and telecommunication are just two industries dealing with data, some of it “personal/private” in nature. A few questions and open issues are intended for discussion during the workshop. The following list is just a starting point for a hopefully lively discussion and exchange of experience:

• How to get “user consent” to exchange PII?
• Do we need “one huge universal” ontology (for transportation and privacy) or several smaller ones?
• Would a “policy model” be applicable for transportation too?
• Transparency for the data subject and privacy protection might contradict the “open” use of PII (such as location information for advances traffic management)
EU HORIZON2020 PROJECT „SPECIAL“ CONSIDERS THIS AN OPPORTUNITY FOR EUROPE, NOT A CHALLENGE

Together with academia, industry and even regulatory authorities, the consortium developed a solution for the needs above:

Basic concept: „sticky policies“ attached to data

- Policy language to express „rules“ how data (Personally Identifiable Data, PII) may or may not be used
- Policy engine to check (and enforce) compliance
- Full transparency and control for the user (“data subject”)

More details on the web: https://www.specialprivacy.eu/
ACKNOWLEDGEMENT: EU PROJECT „SPECIAL“: PROJECT PARTNERS AND CREDITS

This work is a result of H2020 project SPECIAL and was supported by

The project SPECIAL (Scalable Policy-awarE linked data arChitecture for privacy, trA transparency and compLiance) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731601 as part of the ICT-18-2016 topic Big data PPP: privacy-preserving big data technologies.