

# SPECIAL: Scalable Policy-aware Linked Data arChitecture for prIvacy, trAnsparency and complIance

Javier D. Fernández<sup>1</sup>, Sabrina Kirrane<sup>2</sup>, Axel Polleres<sup>1</sup>, and Rigo Wenning<sup>3</sup>

<sup>1</sup> Vienna University of Economics and Business & Complexity Science Hub, Vienna, Austria

<sup>2</sup> Vienna University of Economics and Business, Vienna, Austria

<sup>3</sup> W3C, Sophia-Antipolis, France

jfernand@wu.ac.at, sabrina.kirrane@wu.ac.at, axel.polleres@wu.ac.at,  
rigo@w3.org

**Abstract.** SPECIAL is a research and innovation action, which is funded under the H2020 ICT-18-2016 Big data PPP: privacy-preserving big data technologies call. The SPECIAL project aims to address the contradiction between Big Data innovation and privacy-aware data protection by proposing a technical solution that makes both of these goals realistic.

**Keywords:** Big Data, Compliance, Data protection, Legislation, Linked Data, Privacy, Robustness, Scalability, Transparency

## 1 Introduction

The SPECIAL<sup>4</sup> platform, which is routed in Semantic Web technologies and Linked Data principles [1]: (i) supports the acquisition of user consent at collection time and the recording of both data and metadata (consent, policies, event data, context) according to legislative and user-specified policies; (ii) caters for privacy-aware, secure workflows that include usage/access control, transparency and compliance verification; (iii) demonstrates robustness in terms of performance, scalability and security all of which are necessary to support privacy preserving innovation in Big Data environments; and (iv) provides a dashboard with feedback and control features that make privacy in Big Data comprehensible and manageable for data subjects, controllers, and processors. SPECIAL shall allow citizens and organisations to share more data, while guaranteeing data protection compliance, thus enabling both trust and the creation of valuable new insights from shared data. In order to support transparency across company borders Linked Data principles such as allocating unique IRIs to data resources and using these IRIs to associate metadata (i.e. policies and event data) with said resources will be employed.

---

<sup>4</sup> <https://www.specialprivacy.eu/>

## 2 R&D activities and goals

The SPECIAL project, will be realised by combining and significantly extending big data architectures to handle Linked Data, harnessing them with sticky policies [4] as well as scalable queryable encryption [2] [3], and developing advanced user interaction and control features. SPECIAL builds on top of the Big Data Europe<sup>5</sup> and PrimeLife<sup>6</sup> projects, exploits their results, and further advances the state of the art of privacy enhancing technologies. Against this background, the vision of SPECIAL translates into the following, concrete activities and goals:

- A policy management framework is required in order to ensure that data subjects can associate access and usage policies with their personal data and to support the derivation of policies for processed data. At the centre of this objective is the need for a policy language that is able to represent not only access/usage policies in machine readable format, but also legal rules, business rules, provenance data, and contextual information. Additionally, there is a need to develop automated policy synthesis techniques that can be used to derive policies for data produced by Big Data processing algorithms (aggregation, mining, etc.).
- A transparency and compliance framework is needed in order to generate an immutable record of events that are linked to data and associated policies. Encryption, hashing and digital signatures are required in order to ensure both the integrity and non-repudiation of policies and events. In order to support traceability in terms of both the processing and sharing of personal data it is necessary to link data, policies and provenance/events with contextual information relating to the user and/or the environment. While, from a compliance perspective, it is necessary to automatically verify that the processing and sharing of personal data is inline with access and usage policies and also with the data protections legislation, and to inform the relevant parties in a nonintrusive manner.
- Both the policy management and the transparency and compliance frameworks need to be realised in the form of a scalable architecture. The proposed scalable policy-aware Linked Data architecture will be evaluated under real-world conditions not only in terms of functionality, but also placing a strong emphasis on nonfunctional requirements, such as performance, scalability and security.

**Acknowledgements.** Supported by the European Union’s Horizon 2020 research and innovation programme under grant 731601.

## References

1. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5, pp. 1-22.

<sup>5</sup> <https://www.big-data-europe.eu/>

<sup>6</sup> <http://primelife.ercim.eu/>

2. Brakerski, Z., & Vaikuntanathan, V. (2014). Efficient fully homomorphic encryption from (standard) LWE. *SIAM Journal on Computing*, 43(2), pp. 831-871.
3. Cash, D., Jaeger, J., Jarecki, S., Jutla, C., Krawczyk, H., Rosu, M., & Steiner, M. (2014). Dynamic searchable encryption in very-large databases: Data structures and implementation. *IACR Cryptology ePrint Archive*, 2014:853
4. Trablesi, S., Neven, G., Ragget, D. et al. (2011). Report on design and implementation. PrimeLife Deliverable D5.3.4. Available at: [http://primelife.ercim.eu/images/stories/deliverables/d5.3.4-report\\_on\\_design\\_and\\_implementation-public.pdf](http://primelife.ercim.eu/images/stories/deliverables/d5.3.4-report_on_design_and_implementation-public.pdf).