Scalable Policy-aware linked data architecture for privacy, transparency and compliance

### Reporting

**Project Information**

<table>
<thead>
<tr>
<th>Grant agreement ID: 731601</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Closed project</strong></td>
</tr>
<tr>
<td><strong>EC signature date</strong></td>
</tr>
<tr>
<td><strong>Start date</strong></td>
</tr>
<tr>
<td><strong>End date</strong></td>
</tr>
</tbody>
</table>

**Funded under**

INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)

**Total cost**

€ 3,991,388,75

**EU contribution**

€ 3,991,388,75

**Coordinated by**

GEIE ERCIM

France

This project is featured in...
The value of the EU data economy was more than €285 billion in 2015, and is expected to rise to €739 billion by 2020 [1]. The expectation is that the mining of Big Data will bring significant advances for business, science and society at large, however there are concerns with respect to privacy and data protection. In order to address this challenge, SPECIAL enables the trustful usage and sharing of personal data even across company borders. It does so by providing an “automated means using technical specifications” to support consent and transparency requirements that are specified in the European General Data Protection Regulation (GDPR). Against this background, the vision of SPECIAL translates into the following, concrete objectives:

- The development of a policy management framework that can be used to ensure that data subjects can associate usage policies with their personal data, and to support the derivation of policies for inferred and aggregated data. At the center of this objective is the need for a policy language that is able to represent not only usage policies in machine readable format, but also legal rules, business rules, provenance data, and contextual information.
- Additionally, there is a need for a transparency and compliance framework, which can be used to verify that personal data usage policies are being adhered to. 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. From a compliance perspective, it is necessary to automatically verify that the processing and sharing of personal data is inline with usage policies and also with the data protections legislation, and to inform the relevant parties in a nonintrusive manner. While, encryption, hashing and digital signatures are required in order to ensure both the integrity and non-repudiation of policies and events.
- Finally, it is necessary to ensure that SPECIAL outputs are suitable for big data uses cases such as the telecoms and financial services pilots being developed in SPECIAL. Towards this end both the policy management and the transparency and compliance frameworks need to be realized in the form
of a scalable transparency and compliance architecture. The proposed scalable policy-aware Linked Data architecture needs to be evaluated under real-world conditions not only in terms of functionality, but also placing a strong emphasis on robustness requirements, such as performance, scalability and security.


Work performed from the beginning of the project to the end of the period covered by the report and main results achieved so far

From an operational perspective, the consortium has put in place several templates, processes and procedures in order to ensure the smooth operation of the project, including but not limited to: a quality, risk and data management plan, ethical guidelines and procedures, an exploitation plan, and a standardisation plan.

At the start of the project the Telecoms and Financial Services use cases underwent a thorough legal and technical analysis. Both the uses cases and the insights gaining from the legal and technical analysis were used to develop the SPECIAL usage policy language and supporting vocabularies. Which, subsequently lead to the development of a log vocabulary that can be used to record data processing and sharing events, and the compliance checking algorithm that can to used for both ex-post and ex-ante compliance checking.

In order to allow for the iterative development of project outcomes based not only on a more in-depth understanding of personal data processing and sharing use cases, but also on the evolving legal landscape, the project was designed in a manner that enables work published in early deliverables to be iteratively refined based on feedback obtained both internally from downstream activities and externally via dissemination and standardisation activities. Thus, in the first 18 months of the project the agile methodology enabled the SPECIAL consortium to develop several consent interfaces, a control interface and an initial transparency dashboard, and to gain insights from the initial usability testing. In parallel, the consortium has developed and deployed the first major release of the big data transparency and compliance checking platform and devised a benchmark that can be used to compare this early release to alternative platforms (that are also under development), and to continuously refine and enhance the backend from both a functionality and a robustness perspective.

Both Proximus and Deutsche Telekom are currently evaluating the effectiveness of the existing consent user interfaces in the context of their use cases, while Thomson Reuters are currently evaluating the consent and transparency engine. Additionally the policy language and vocabularies formed the basis of initial discussions around standardisation. Towards this end, SPECIAL launched a W3C Data Privacy Vocabularies and Controls Community Group (DPVCG) on the May 25th 2018 (the day the GDPR came into effect). The objective of the DPVCG is to provide a platform for engagement with the wider community, to gather additional use cases and to develop standard vocabularies that can be used for personal data processing consent, transparency and GDPR compliance.

From a dissemination perspective, the work performed to date has already resulted in 27 deliverables,
and 18 publications in peer reviewed journals, conference and workshops. Already in the first reporting period activities in relation to the SPECIAL have been presented in major venues spanning: Privacy, Semantic Web, Legal Informatics, Artificial Intelligence, and Big Data, and the consortium have presented the project at several different Dagstuhl seminars.

Progress beyond the state of the art and expected potential impact (including the socio-economic impact and the wider societal implications of the project so far)

SPECIAL significantly advances the state of the art in the following concrete areas:
- From a policy perspective, SPECIAL builds upon sophisticated policy frameworks introduced in previous projects and existing standardisation efforts and adapts them so as to reach the right balance between the expressiveness and scalability of the usage control policy language.
- From a transparency perspective, SPECIAL enables data transactions (i.e. who shared what data with whom and under what usage conditions) to be stored in a manner that prevents tampering and repudiation from any of the involved peers (i.e. those owning, disclosing, and acquiring data, respectively), and to ensure that all recorded transactions have actually taken place.
- SPECIAL extends the Big Data Europe (BDE) platform, an open source and multi-purpose data management environment, with transparency and compliance checking capabilities.
- The SPECIAL dashboard is a generic visualisation platform that is able to show users the information that data controllers and processor know about them, and the relevant metadata (policies, event data, context) attached to this data. While, the consent and control interfaces enable users to effectively manage permissions in an understandable manner.
- The legal partners are working closely with the technological development partners in order to provide a solutions based on the principles of privacy by design and privacy by default.

Last update: 13 September 2020