Data Privacy Vocabularies and Controls: Semantic Web for Transparency and Privacy

Piero A. Bonatti[5], Bert Bos[2], Stefan Decker[4], Javier D. Fernández[1], Sabrina Kirrane[1], Vassilios Peristeras[3], Axel Polleres[1], and Rigo Wenning[2] *

[1] Vienna University of Economics and Business, Austria
{firstname.lastname}@wu.ac.at
[2] W3C/ERCIM
{firstname}@w3.org
[3] International Hellenic University, Greece
v.peristeras@ihu.edu.gr
[4] RWTH Aachen University, Germany, Germany
decker@informatik.rwth-aachen.de
[5] Universita' degli Studi di Napoli Federico II, Italy
piero.bonatti@unina.it

Abstract. Managing Privacy and understanding the handling of personal data has turned into a fundamental right—at least for Europeans—since May 25th with the coming into force of the General Data Protection Regulation. Yet, whereas many different tools by different vendors promise companies to guarantee their compliance to GDPR in terms of consent management and keeping track of the personal data they handle in their processes, interoperability between such tools as well uniform user facing interfaces will be needed to enable true transparency, user-configurable and -manageable privacy policies and data portability (as also—implicitly—promised by GDPR). We argue that such interoperability can be enabled by agreed upon vocabularies and Linked Data.

1 Why Privacy needs Interoperability

The level of privacy and trust concerns has raised to a point where regulators, citizens and companies have started to take action. Services on the Web are often very complex orchestrations of cooperations between multiple actors, and the processing of personal data in Big Data environments is becoming intransparently complex. The level of complexity will increase further if the upcoming Internet of Things is taken into account. If trust in such services is eroded, the growth of the Web and the digital economy itself, and therefore the prosperity of society in general are endangered. The challenge is how to convey the transparency to the user to allow for informed personal data self determination. This includes especially methods to gather and manage user consent, even in an IoT environment.

While building privacy-by-design[3] into systems is a much wider scope, we already lack the tools and best practices for those wanting to be good citizens of the Web to provide interoperable and understandable privacy controls, with the exception maybe of work on Permissions[6] and on tracking protection[5], but even those only cover partial aspects.

We argue that existing work on Standards in this domain have *not* yet enabled interoperabilty. In particular:

- There are no standard vocabularies to describe and interchange personal data – these are needed to support the data subjects' right to personal data portability.
- There are – with the exception maybe of starting points in P3P [4] no agreed upon vocabularies or taxonomies for describing *purposes* of personal data handling and *kinds of processing*: the GDPR requires that both consent and data processing are tied to a concrete purpose and processing data for other than this purpose is unlawful. Consequently personal data processing should be logged with a standard reference to a purpose which complies with the users consent. While e.g. the PROV ontology [9] vocabulary provides a means to express processing and provenance of processed data, and ODRL [6] allows to describe certain permissions and obligations, concrete taxonomies for such processing and permissions in the context of personal data handling are not yet standardized.
- There are no agreed upon vocabularies or ontologies that align *the terminology of privacy legislations*, such as the GDPR, with the above mentioned vocabularies, that would allow software vendors to claim compliance with such regulations.

2 Finding agreement on areas for Standardisation

As a first step towards closing these gaps, on the 17th and 18th April 2018, some forty people took part in a W3C workshop on data privacy controls and vocabularies in Vienna. The initial idea was that linked data annotations can help tackle the issue of privacy in modern data environments and will allow for the creation of the next generation of privacy enhancing technologies. Of course, the enactment of the GDPR was also prominent in the discussions. The agenda, developed on the basis of position statements submitted by various stakeholders active in the area of standard solutions for interoperable privacy[6] consisted of sessions on four themes,'relevant vocabularies and initiatives', 'industry perspective', 'research topics' and 'the governmental side & initiatives'. The workshop ended with a discussion on next steps and in particular on priorities for standardisation. In order from highest to lowest priority, the goal is to develop a:

1. Taxonomy of regulatory privacy terms (including all GDPR terms).
2. Taxonomy for personal data.
3. Taxonomy of purposes.

[6] All the over 30 position statements and expressions of interests by various stakeholders and organisations are available at https://www.w3.org/2018/vocabws/people.html

4. Taxonomy of disclosure and consent.
5. Metadata related to the details of anonymisation.
6. Log vocabularies to describe and exchange logs on personal data processing.

3 The Data Privacy Vocabularies and Controls Community Group

One of the primarily outputs of the workshop was the setup of a 'W3C Community Group (CG)' entitled *Data Privacy Vocabularies and Controls CG (DPVCG)*[7] which was created on the 25th of May, the official starting date of the General Data Protection Regulation.

The goal of the CG is to harmonize related efforts and bring together stakeholders that already have put forward proposals to develop vocabularies to enable semantic interoperability and interchange of transparency logs about personal data processing, enable data portability for data subjects, etc. The exact scope of use cases related to making personal data processing interoperable by respective standards in order to ease proof of compliance with the GDPR and related privacy protection regulations will be the first deliverable of the CG.

More concretely, the following steps and deliverables are planned so far.

1. **Use cases and requirements**: in a first step we will collect and align common requirements from industry and also from other stakeholders to identify areas where interoperability is most needed in the handling of personal data. The outcome shall be a prioritized list of requirements for what needs to be covered by shared vocabularies to enable interoperability in the identified use cases.
2. **Alignment of vocabularies and identification of overlaps**: in a second document, we will collect existing vocabularies and standardization efforts to identify their overlaps and suitability as starting points to minimally and extensibly cover the requirements prioritized in step one.
3. **Glossary of GDPR terms**: a third deliverable will be an understandable glossary of common terms from the GDPR and how they shall be covered by the agreed vocabularies.
4. **Vocabularies** based on the heterogeneity or homogeneity of the agreed upon use cases and requirements, we plan define a set of vocabularies for exchanging and representing interoperably: personal data, purposes/processing, disclosure/consent, anonymisation, and transparency logs.

Some starting points for these vocabularies e.g. for transparency logs that companies can keep on their collected consent and personal data handling [8], along with lightweight reasoning techniques [1, 7, 2] to check GDPR-compliance over these logs have already been created in draft versions by the EU H2020 project SPECIAL[8], in which several of the co-authors are actively involved.

[7] https://www.w3.org/community/dpvcg/

[8] Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance, cf. http://specialprivacy.eu

Starting points in terms of a description of existing vocabularies and gaps between them can be found in [10].

4 Input and Collaboration

Interoperability and agreed upon vocabularies are in our opinion crucial for truly implementing the goals of stricter, more user-centric privacy controls for data subjects, but also to alleviate the burden for companies in the context of fulfilling their compliance obligations in the context of GDPR. We invite any interested stakeholders to join the Data Privacy Vocabularies and Controls Community Group and provide feedback to its drafts. The first Face-2-face meeting is being planned at the MyData2018 conference in Helsinki, Finland, in August, and the second 2nd Face-2-face meeting being planned to be co-located with the European Big Data Value Forum in Vienna, Austria, in 12-14 November 2018. We hope the SW4CG workshop at ISWC also offers a stage to discuss the emerging needs for interoperability in terms of privacy and personal data handling, and how Semantic Web technologies and standards can help to satisfy these needs.

Acknowledgments. This work has been supported by the European Union's Horizon 2020 research and innovation programme under grant 731601 (SPECIAL).

References

1. Bonatti, P., Kirrane, S., Petrova, I., Sauro, L., Schlehahn, E.: Deliverable D2.1: Policy language v1 (Dec 2017), SPECIAL public project deliverable
2. Bonatti, P.A.: Fast compliance checking in an OWL2 fragment. In: Proc. IJCAI (2018)
3. Cavoukian, A.: Privacy by design. Take the challenge. Information and privacy commissioner of Ontario, Canada (2009)
4. Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., Reagle, J.: The platform for privacy preferences 1.0 (p3p1.0) specification (Apr 2002), https://www.w3.org/TR/P3P/, W3C Rec.
5. Fielding, R.T., Singer, D.: Tracking preference expression (dnt) (Oct 2017), https://www.w3.org/TR/2017/CR-tracking-dnt-20171019/, W3C Candidate Rec.
6. Iannella, R., Steidl, M., Myles, S., Rodríguez-Doncel, V.: ODRL Vocabulary & Expression 2.2 (Feb 2018), https://www.w3.org/TR/odrl-vocab/, W3C Rec.
7. Kirrane, S., Bonatti, P.A., Fernández, J.D., Galdi, C., Sauro, L.: Deliverable D2.4: Transparency and compliance algorithms v1 (2018), SPECIAL public project deliverable
8. Kirrane, S., Fernández, U.M.J.D., Polleres, A.: Deliverable D2.3: Transparency framework v1 (2018), SPECIAL public project deliverable
9. Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J.: Prov-o: The prov ontology (Apr 2013), https://www.w3.org/TR/prov-o/, W3C Rec.
10. Polleres, A., Kirrane, S., Wenning, R.: Deliverable D6.3: Plan for community group and standardisation contribution (2017), SPECIAL public project deliverable