Privacy: Standards and Vocabularies for Transparency & Interoperability

Axel Polleres Joint work with: Piero Bonatti, Bert Bos, Stefan Decker, Javier D. Fernández, Sabrina Kirrane, Vassilios Peristeras, Rigo Wenning, Martin Kurze, Ben Whittam Smith & the members of the W3C Data Privacy Vocabularies and Controls CG
Background...

Project Launch:
SPECIAL (a Scalable Policy-awareE linked data arChitecture for privacY, trAnsparency and compLiance)

Axel Polleres, Vienna University of Economics and Business (WU Wien)
MyData 2017, Tallinn/Helsinki
10/08/2017
Use Cases for Transparency and Interoperability in Privacy:

- **Companies:** Ensuring Regulatory Compliance for Companies
- **Regulators:** Checking and enforcing GDPR
- **Data Subjects:** Personal Data Markets: from “Data Collection” to “Data Donations”

Different roles have different use cases.
Components of Personal Data Processing (not exhaustive...)

- Rules/Policies
  - Consent
  - Regulations
- Purpose
- Processing
- Storage
- Personal Data (categories, formats)
Regulatory Compliance for Big Companies

- Many heterogeneous systems that process personal data
- Potentially many different places that store and hold consent
- How to deal with GDPR data requests at scale?
- How to prove to the regulator and to the customer that personal data has been handled in compliance to consent only?
Regulatory Compliance for Small Companies

- No resources to build their own compliance infrastructure
- How to deal with GDPR data requests at scale?
- How to prove to the regulator and to the customer that personal data has been handled in compliance to consent only?

\[ P_4 \subseteq P_2 \]
Semantic Interoperability boils down to:
- What is a common core to address these use cases?
- How do we benefit them all at the same time?

Data Privacy Controls and Vocabularies

A W3C Workshop on Privacy and Linked Data

17–18 April 2018, WU Vienna, Vienna, Austria, Europe  
https://www.w3.org/2018/vocabws/report.html
Semantic Interoperability boils down to:
- What is a common core to address these use cases?
- How do we benefit them all at the same time?

Rough workshop outcome / scoping:

1. Taxonomy of regulatory privacy terms (including all GDPR terms).
2. Taxonomy for personal data.
3. Taxonomy of purposes.
4. Taxonomy of disclosure/processing.
5. Metadata (e.g. related to processing details of anonymization).
7. Taxonomy of linkage operations.
8. Taxonomies of human behavior.
Semantic Interoperability boils down to:
- What is a common core to address these use cases?
- How do we benefit them all at the same time?

→ **Foundation of a W3C Community Group (25\textsuperscript{th} May 2018)**

- Collect concrete **Use cases**
- Collect **Existing Vocabularies**
- Align **Core Vocabularies**

1. Taxonomy of regulatory privacy terms (including all GDPR terms).
2. Taxonomy for personal data.
3. Taxonomy of purposes.
4. Taxonomy of disclosure-processing.
5. Metadata (e.g. related to processing details of anonymization)
7. Taxonomy of linkage operations.
8. Taxonomies of human behavior.
Semantic Interoperability boils down to:
- What is a common core to address these use cases?
- How do we benefit them all at the same time?

→ Foundation of a W3C Community Group (25th May 2018)

→ Collect concrete Use cases
→ Collect Existing Vocabularies
→ Align Core Vocabularies

Use-Cases [edit]
- SPECIAL/Proximus use case - personalized touristic recommendations
- SPECIAL/DT use case - mobile network quality measurements
- SPECIAL/TR use case - 'Know Your Customer' (finance, anti-money-laundering)
- DECODE/DECO1 use case - Online voting system with privacy
- DECODE/DECO2 use case - Rental Register
- DECODE/DECO3 use case - Sharing sensor data

Vocabularies [edit]
- CDMM Consent Ontology
- COEL
- Data Protection Ontology by Bartolini et. al
- GDPRov
- GDPREXT
- IEEE 7012
- ODRL
- P3P
- P-Plan
- Privacy Preference Ontology
- PROV-O
- SPECIAL Usage Policy
- SPECIAL Policy Log

We need your input!
→ Join DPVCG!

https://www.w3.org/community/dpvcg/wiki/Use-Cases,_Requirements,_Vocabularies
Starting Point: Use Cases/Vocabularies from SPECIAL
Three Distinct Use Cases:

1. **THOMSON REUTERS**
   - Know-Your-Customer services for the banking industry

2. **proximus**
   - Recommendation engine for subscribers

3. **T**
   - Service quality monitoring
One Compliance Solution:

Processing requires **PERMISSIONING**

Permissions must be compliant with the **GDPR**

Permissions must be compliant with **Consent**

i.e., **COMPLIANCE** is a logical operation
What to Standardise:

Core Logic
Core Vocabularies
Compliance Services

Against What Criteria:

Completeness and Correctness: ≤
Market adoption
SPECIAL’s view on Core Interoperability Components:

Rules/Policies: SPECIAL Usage Policy Language (SPL)

- Purposes
- Processing
- Storage
- Data
- Recipients

SPECIAL namespaces:

@prefix spl: <http://www.specialprivacy.eu/langs/usage-policy#>.
@prefix svpu: <http://www.specialprivacy.eu/vocabs/purposes#>.
@prefix svpr: <http://www.specialprivacy.eu/vocabs/processing#>.
@prefix svd: <http://www.specialprivacy.eu/vocabs/data#>.
@prefix srv: <http://www.specialprivacy.eu/vocabs/recipients#>.
@prefix splog: <http://www.specialprivacy.eu/langs/splog#>.

...
The data controller will collect financial and judicial information from public sources and analyse it for “know your customer” purposes. This information will be stored on the controller’s servers and released to specific third parties.
ObjectIntersectionOf(
  ObjectSomeValueFrom(spl:hasData
    ObjectUnionOf(svd:Financial svd:Judicial ))
  ObjectSomeValueFrom(spl:hasProcessing
    ObjectUnionOf(tr:Collect-public svpr:Analyze ))
  ObjectSomeValueFrom(spl:hasPurpose tr:KYC )
  ObjectSomeValueFrom(spl:hasStorage
    ObjectIntersectionOf(
      ObjectSomeValueFrom(spl:hasLocation spl:ControllerServers)
      DataSomeValuesFrom(spl:durationInDays
        DatatypeRestriction(xsd:integer xsd:mininclusive "0"^^xsd:integer ))
    ))
  ObjectSomeValueFrom(spl:hasRecipient svr:AnyRecipient )
)
Discussion...
How to structure those taxonomies?
What are the important use cases to cover?
Components of Personal Data Processing (not exhaustive...)

Rules/Policies
- Consent
- Regulations

Purpose

Processing

Storage

Personal Data (categories, formats)
The MyData model could be integrated on the “high level” it presents data as

- human-centrically grouped into “areas of interest”
- as well as how it’s
- processed
- and used
# Personal Data (categories, formats)

the World Economic Forum model

## Figure 10: A taxonomy of personal data by origin

<table>
<thead>
<tr>
<th>TYPE</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individually provided</td>
<td>• Photos&lt;br&gt;• Blogs&lt;br&gt;• Emails&lt;br&gt;• Tweets&lt;br&gt;• Online transaction data&lt;br&gt;• Registration forms &amp; job applications</td>
</tr>
<tr>
<td>Observed</td>
<td>• Internet browsing preferences&lt;br&gt;• Surveillance video&lt;br&gt;• Location data&lt;br&gt;• Call detail records</td>
</tr>
<tr>
<td>Inferred</td>
<td>• Credit scores&lt;br&gt;• Consumer profiles&lt;br&gt;• Predictive traffic flows&lt;br&gt;• Patterns in the spread of infectious diseases&lt;br&gt;• Targeted advertisement</td>
</tr>
</tbody>
</table>

Source: Information Accountability Foundation, World Economic Forum, Marc E. Davis

Quelle: MyData
Quelle: http://reports.weforum.org/rethinking-personal-data/?doing_wp_cron=1538384793.1468729972839355468750
Personal Data (categories, formats)

How should we structure our taxonomy?

- Which one is more fit for purpose?
- Which one covers 80:20 use cases?

How to structure our other taxonomies?

What expressivity do we need (conditionals, etc.)?
Call for Action: Join DPVCG!

• More use cases matter!
• Existing efforts for interoperability/vocabularies matter!

Joining is easy!
→ The group is Open to everyone!
→ Just create a W3C account

https://www.w3.org/community/dpvcg/

• Looking forward to discussions...