Privacy Model and Annotation for DaaS

Michaël Mrissa[1], Salah-Eddine Tbahriti[1], Hong-Linh Truong[2]

[1]LIRIS Lab., Université de Lyon, France
firstname.surname@liris.cnrs.fr
[2]Distributed Systems Group, Vienna University of Technology, Austria
truong@infosys.tuwien.ac.at

December 2, 2010

Outline

Research context of our work

A quick reminder on DaaS

- Recent evolution of WS models (SaaS, cloud computing)
- Data as a Service (DaaS) $\Rightarrow$ data has a central place
- *Service* provides *data* to *consumers*
- Same concerns as Web services (privacy, QoS, etc.)

BUT...DaaS have specific characteristics

1. Three players in the game
   - data consumer (gets data from services)
   - service provider (gathers data sources)
   - data provider (releases data)
2. Various flavours of DaaS
   - Service API
   - SOAP and REST-based
   - Data: structured, semi-structured data (XML), unstructured data (zip dataset, images..)

Research context of our work

A quick reminder on DaaS

- Recent evolution of WS models (SaaS, cloud computing)
- Data as a Service (DaaS) $\Rightarrow$ data has a central place
- *Service* provides *data* to *consumers*
- Same concerns as Web services (privacy, QoS, etc.)

BUT...DaaS have specific characteristics

1. Three players in the game
   - data consumer (gets data from services)
   - service provider (gathers data sources)
   - data provider (releases data)
2. Various flavours of DaaS
   - Service API
   - SOAP and REST-based
   - Data: structured, semi-structured data (XML), unstructured data (zip dataset, images..)

Problems raised

General problems

- Several data concerns influencing data usage
- Complex relationships between the three players wrt data concerns and service contracts bound to these concerns

Let us focus on the privacy problem for DaaS only

- Lack of consideration for privacy in the DaaS lifecycle
- Lack of explicit description for privacy aspects
  $\Rightarrow$ at the **data**, **service** and **user** level
- Lack of correlation between the two problems

Missing information leads to bad interpretation of results

Privacy-awareness for DaaS

Several questions:

- How can data consumers recognize privacy information associated to the services they use and the related data resources ?
- How to describe privacy policies so that:
  $\Rightarrow$ they are available when accessing the service ?
  $\Rightarrow$ they still can be associated to data resources ?
- How can data providers work with service providers to ensure privacy ?

At the architectural level - current problems

- The current interaction model
  - ignores **data providers' concerns**
  - does not support explicit desc. of privacy concerns
- which results in
  - handmade contract between service and data providers
  - no way to ensure privacy constraints are respected
  - privacy aspects not visible to data consumers

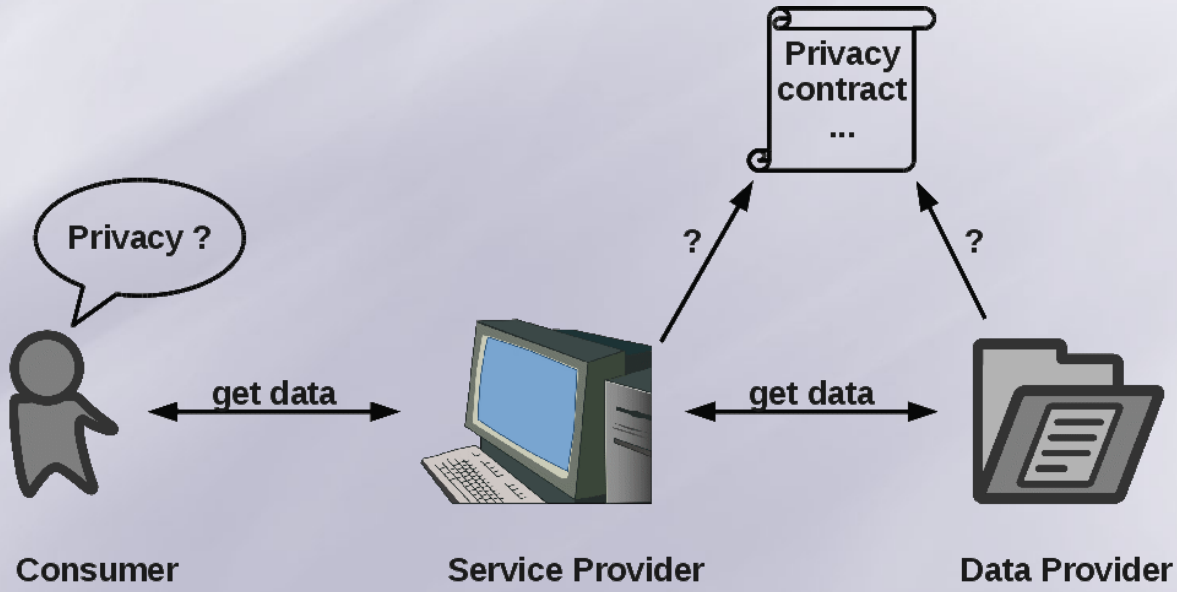

Figure: Typical architectural view

At the architectural level - what should be done

- Need to make privacy requirements explicit
- Need for a formal background ⇒ underlying privacy model
- Need for a link to DaaS descriptions ⇒ annotation
- and to data resources ⇒ PDT (developed later)

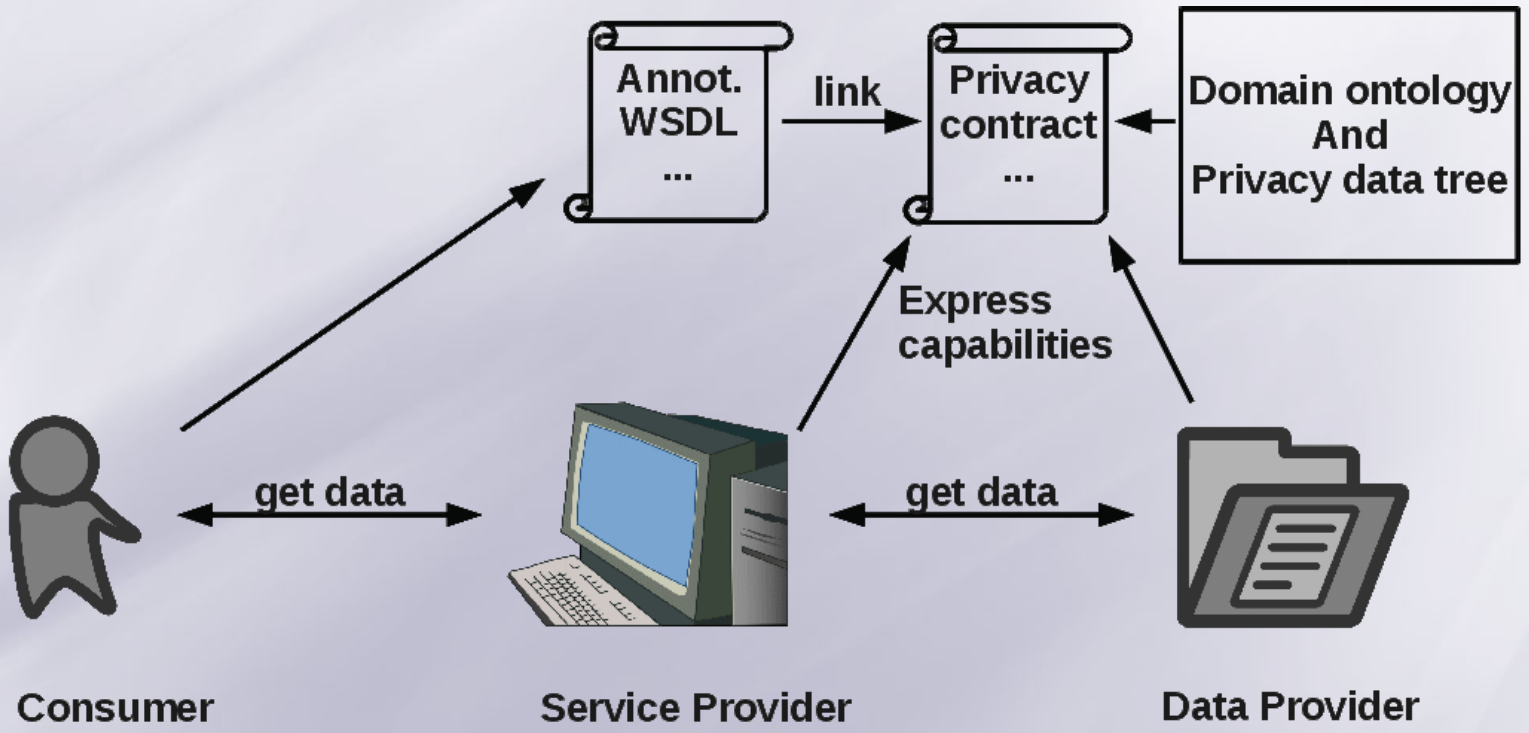

Figure: Architectural view: our proposal

The privacy concern

Handling Privacy is about

respecting restrictions about data (disclosure, usage) according to the requirements of data owners and service providers

Related Work

1. **Modeling privacy**
   - Some works exist for Web services
   - Data and data providers' concerns ignored
   - Sometimes modeled as a QoS option (confidentiality)
2. **Privacy and WS composition**
   - No formal privacy model
   - Mostly for structured data
3. **Privacy in data integration**
   - Access control mostly
   - Focus on algorithms

The privacy concern

Handling Privacy is about

respecting restrictions about data (disclosure, usage) according to the requirements of data owners and service providers

Related Work

1. **Modeling privacy**
   - Some works exist for Web services
   - Data and data providers' concerns ignored
   - Sometimes modeled as a QoS option (confidentiality)
2. **Privacy and WS composition**
   - No formal privacy model
   - Mostly for structured data
3. **Privacy in data integration**
   - Access control mostly
   - Focus on algorithms

Expliciting privacy requirements for DaaS

Published Privacy Requirements		Data Provider's Purpose			Data Form	
Category	Requirements	Organizational work	Pay-per-use	Free/Public	Structured	Unstructured
concern	privacy-preserving methods		+	+	+	+
	types of privacy data	+	+	+	+	+
	data rights		+	+	+	+
scope	individual data resources	+	+	+	+	+
	service operation	+	+	+	+	
	service as a whole	+	+	+	+	

Table I
REQUIREMENTS FOR DAAS PROVIDERS TO PUBLISH PRIVACY CONCERNS

Figure: Classification of DaaS requirements

- Different requirements
  - concern
    - methods: apply to data
    - types: type of data concerned
    - rights: usage restrictions/permissions
  - scope: data resource, service operation, service as a whole
- Depend on the **data provider's purpose**
- and on the **form of data**

Our model for privacy concern

Data privacy capability

Machine-interpretable representation of data privacy possibilities w.r.t. data and service providers

Formal definition

- $DPC \Rightarrow$ set of data privacy capabilities for a DaaS
- $DPC = \{dpc_1, dpc_2, \cdots, dpc_n\}$
- $dpc = (CPI, scope)$, where
  - $CPI = \{po(pdt) \cup up(pdt)\}$
  - $scope = \{data\ resource, operation, service\}$
- CPI is defined as a set of privacy operations and/or usage permissions to be associated to data (via PDT)

Underlying data model

- "Data" scope ⇒ data model needed to describe capabilities
- Privacy Data Tree (PDT) to represent data structure
  - Domain-independent, -specific and custom nodes
  - Possibility to express **data rights**
  - Incrementally built ontology

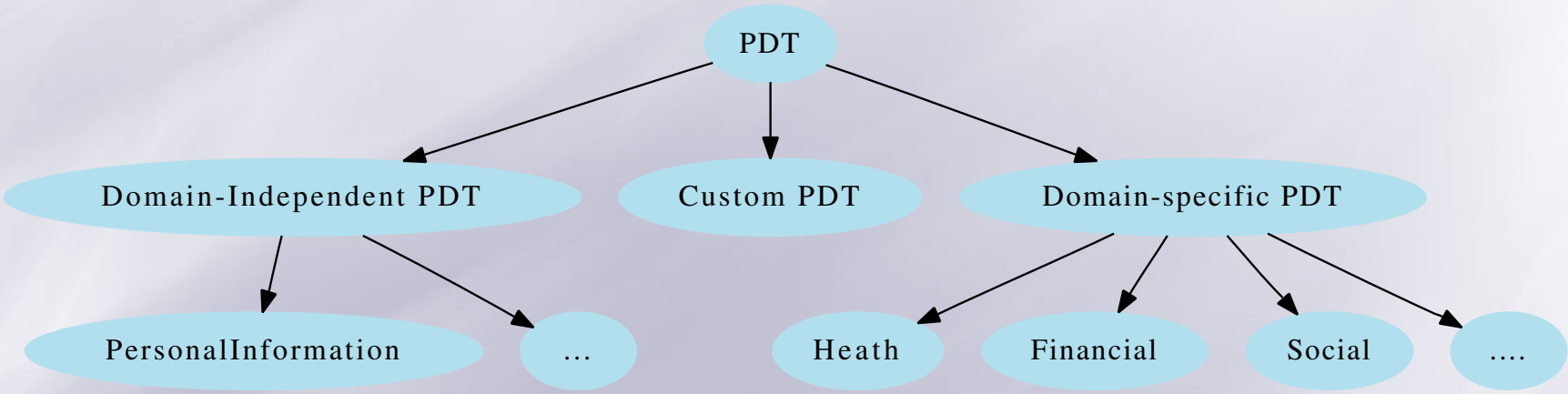

Figure: Overview of a PDT structure

Example

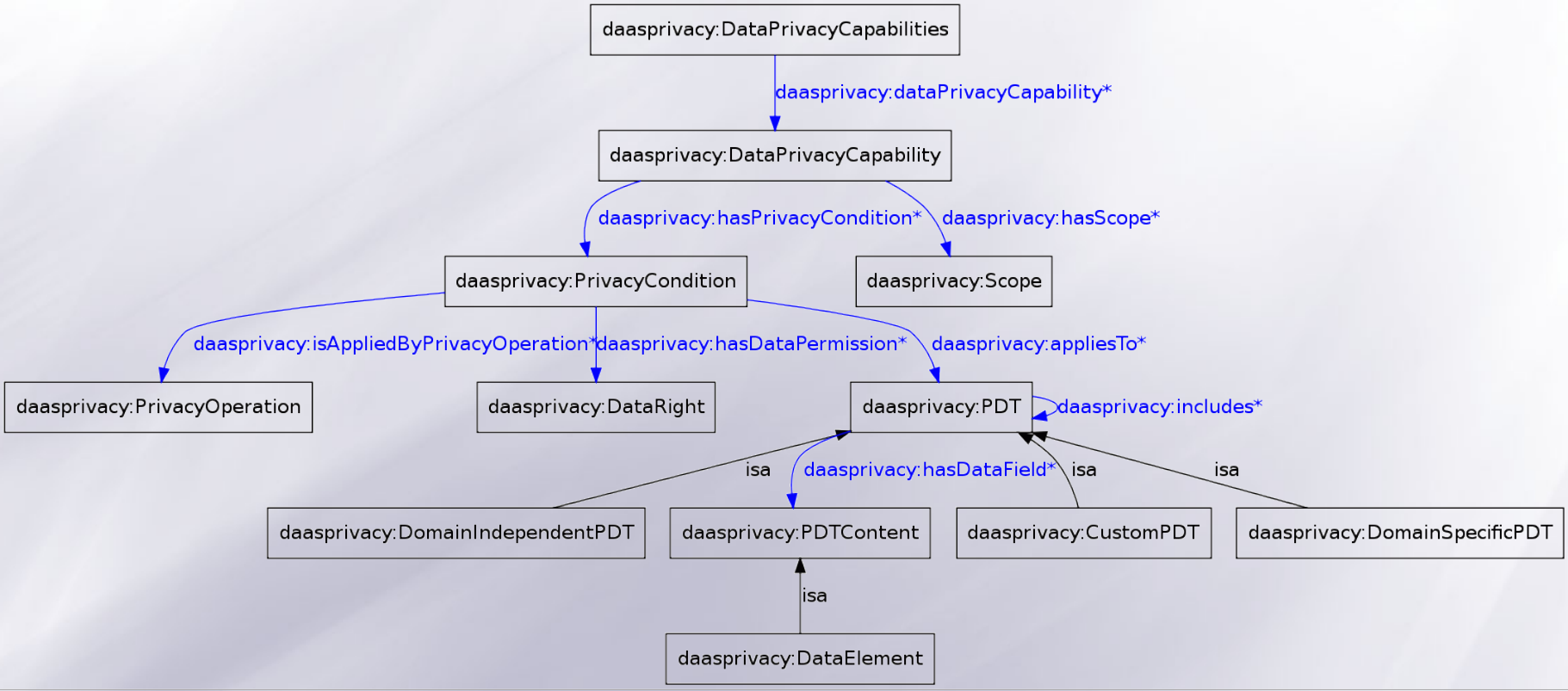

Figure: Our DaaS privacy model in RDF (edited with Protégé)

DaaS Annotation

How to annotate service descriptions?

- WSDL 2.0 $\Rightarrow$ easy, extension elements allowed everywhere
- WSDL 1.1 $\Rightarrow$ less straightforward, use “attrExtensions” element
- RESTful $\Rightarrow$ extension to the MicroWSMO model (syntax-independent, we recommend RDFa)

```
<?xml version="1.0" encoding="UTF-8" ?>
<wsdl:definitions targetNamespace="http://jsonservice.
      liris.cnrs.fr/" name="JSONWSService"
   xmlns:pr="http://www.infosys.tuwien.ac.at/SOD1/
        dataconcerns/daasprivacy.owl#">
...
<portType name="JSONWS">
    <sawsdl:attrExtensions pr:dataprivacycapabilities=
          http://liris.cnrs.fr/~mmrissa/ECOWS/daasprivacy
          .rdf>
...
</portType>
...
</wsdl:definitions>
```

Listing 1. Excerpt of WSDL 1.1 annotation

Figure: Overview of our WSDL 1.1 annotation

The Haiti earthquake posts on Twitter

- Use case
  - Posts about the Haiti earthquake, data provider = Twitter
  - Service provider gives different access rights to data
    - Not authentified, authentified, premium user (levels of trust)
    - Determines the different privacy policies to apply
- Prototype
  - Java Servlet on a Glassfish Server
  - JSON file from Twitter around 100Mo (json-rpc lib.)
  - RDF/XML file for describing privacy capabilities (Jena lib.)
- Development of a request
  - 1° Reception of a request from the data consumer to the service endpoint
  - 2° Parsing of the privacy file attached to the service
  - 3° Fetching data from the data provider (could be a service)
  - 4° Applying privacy policies to data
  - 5° Sending reply to the data consumer

Example

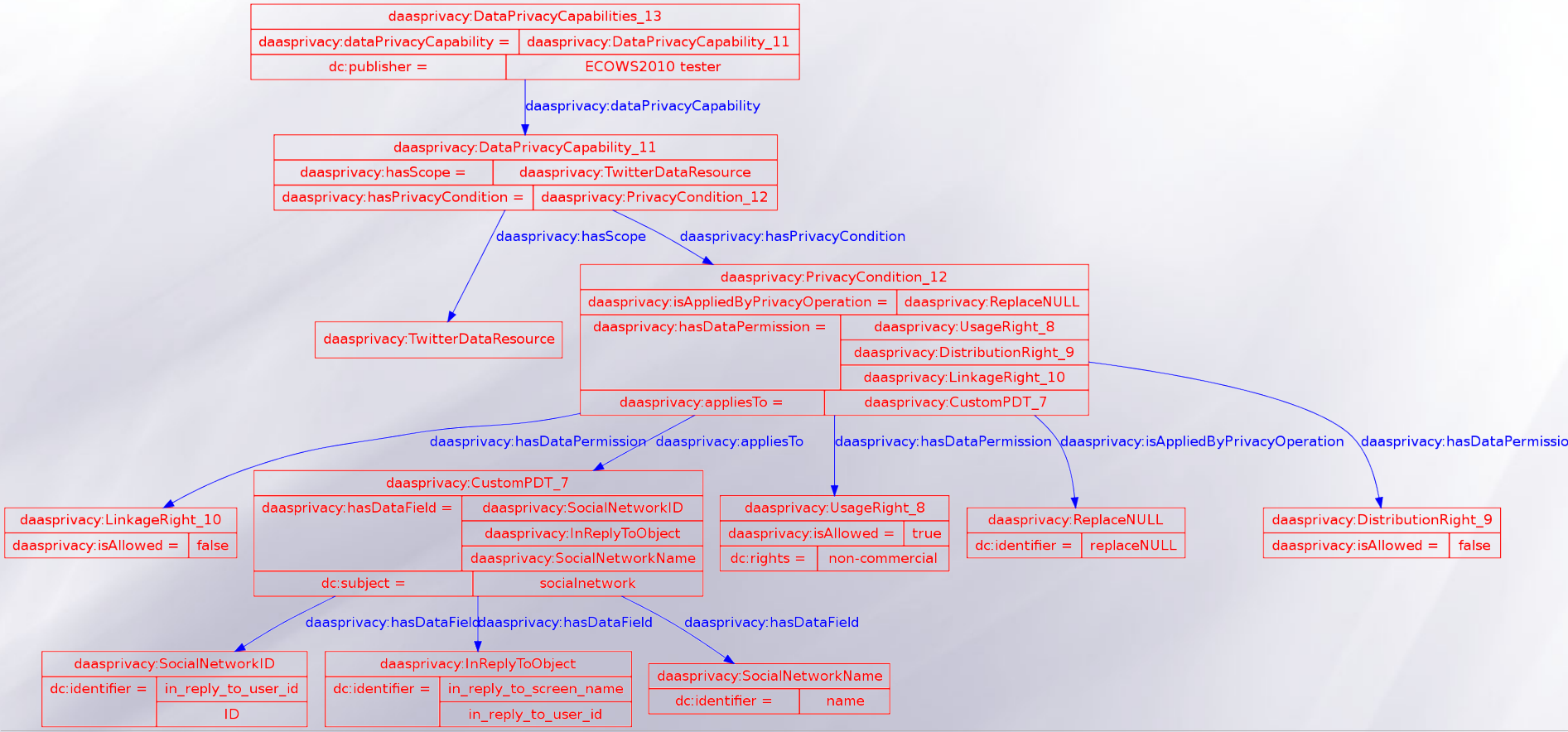

Figure: A RDF instance of DaaS privacy capability (edited with Protégé)

Conclusion

- Discussion
  - Privacy increasingly important for DaaS
  - Needed underlying framework and technical background
    $\Rightarrow$ to foster automatic, privacy-aware interactions
  - This paper is a first step towards this direction
- Open issues and future work
  - What about the interpretation of capability ?
    - Constraints, policies ?
  - How to make sure a privacy policy is fully respected?
  - What about other concerns ?
  - Other questions come from you. . .