Applying open data provenance and licensing to research data

Serena Villata
INRIA Sophia Antipolis
Outline

- Introduction to ANR DataLift project
- Scenarios from requirements of data providers
- Shi3ld - access control for SPARQL endpoints
- Licenses compatibility and composition
- Perspectives: limitations and future challenges
goal of datalift
accelerate the lifting from raw data
to linked public data
ANR DataLift - Partners
DataLift - Main workpackages

- Identify appropriate schemas to assist data selection (LOV catalog)

- Convert data to structured format using appropriate schemas to assist data selection (CSV2RDF, XML2RDF)

- Publish data

- Interlink data with other data sources present on the Web of Data

http://datalift.org/
DataLift - Main workpackages

- Identify appropriate schemas to assist data selection (LOV catalog)
- Convert data to structured formats using appropriate schemas to assist data selection (CSV2RDF, XML2RDF)
- Publish data
- Interlink the data with other data sources present on the Web of Data

http://datalift.org/
Publishing Linked Open Data
Publishing Linked Open Data

Example from DataLift data providers:

IGN requirements for its own data

- Public data (released under specific license)
- Data whose access is restrained to specific persons (e.g., they paid for the service, they are from academia, etc)

Two requirements:

- Access control over data accessible through SPARQL endpoints
- Data licensing
Shi3ld

Context-Aware Access Control for RDF Graph Stores

SELECT ...
WHERE {...}
Shi3ld - Main features

Semantic Web languages **only**
> No new Policy languages

Pluggable to any RDF store
> SPARQL 1.1

Granularity from triples to whole graphs
> Named Graphs
  [Carroll et al, WWW2005]
  RDF 1.1

Mobile context in the loop
> Context Awareness
  [Schilit and Theimer, 94]
  [Dey, 01]
Shi3ld - Vocabularies

Diagram showing relationships between vocabularies:
- AccessConditionSet
- AccessCondition
- AccessPolicy
- AccessPrivilege
- AccessEvaluationContext
- Context
- Device
- User
- Environment
Shi3ld - Examples

ASK {?resource dcterms:creator ?provider . ?provider rel:hasFriend ?consumer . } ARE YOU A FRIEND OF THE DATA PROVIDER ? ✅ ✗


ASK {?resource dcterms:creator ?provider . ?provider rel:hasParent ?consumer . } ARE YOU A PARENT OF THE DATA PROVIDER ? ✅ ✗

ASK{?resource dcterms:creator ?provider . ?provider rel:hasColleague ?consumer . } ARE YOU A COLLEAGUE OF THE DATA PROVIDER ? ✅ ✗
Shi3ld - Examples

?consumer sioc:member_of ?group . }  
ARE YOU A MEMBER OF THE SAME GROUP OF THE DATA PROVIDER ?  

ASK { ?consumer a foaf:Person .
FILTER(?consumer = <http://example#John>) }  
ARE YOU JOHN ? IF SO ✓

ASK { ?consumer a foaf:Person .
FILTER(!(?consumer = <http://example#John>)) }  
ARE YOU JOHN ? IF SO ✗

ASK { FILTER(rand()>0.5) }  
DO YOU GET A NUMBER BIGGER THAN 0.5 ? ✓ ✗
ASK {?
  context a prissma:Context;
  prissma:environment ?env.
  ?env tl:start "2012-10-26T12:00:00Z"^^xsd:dateTime;
  tl:duration "PT5H"^^xsd:duration.
}

ASK {?
  context a prissma:Context;
  prissma:device ?dev;
  prissma:user ?consumer;
  prissma:environment ?env.
  ?consumer a foaf:Person;
    rel:employedBy <http://example#Bob>.
  ?dev a prissma:Device;
    soft:deviceSoftware ?devsw.
  ?devsw a soft:DeviceSoftware;
    soft:operatingSystem ?opsys.
  ?opsys a soft:OperatingSystem;
    common:name "Android".
}

ARE YOU LOCATED IN THE LOUVRE MUSEUM AND IS IT OCTOBER 26TH, 2012 AFTER 12 a.m.?

ARE YOU LOCATED IN THE LOUVRE MUSEUM, ARE YOU EMPLOYED BY BOB, AND ARE YOU USING ANDROID?
Shi3ld - How it works

1. Query **Contextualization**

```sql
INSERT DATA {
  GRAPH :ctx1{
    [user, time, phone, location, ... ]
  }
}

SELECT ... 
WHERE { ... }
```

```
:ctx1
```
2. Access **Policy Evaluation**

```
ASK { ?context
  a prissma:Context;
  prissma:environment ?env.
  ?poi prissma:radius "500";
  ?p geo:lat "43.615811";
  geo:long "7.068532". }

BINDINGS ?context { (:ctx1) }
```

SHIELD - How it works
3. Query Execution on accessible Named Graphs

SELECT ... WHERE {...}

SELECT ...
FROM :ng2,:ng3
WHERE {...}
Licenses in the Web of Data

“the absence of clarity for data consumers about the terms under which they can reuse a particular dataset, and the absence of common guidelines for data licensing, are likely to hinder use and reuse of data”

Heath and Bizer, 2011
Licenses in the Web of Data
Licenses in the Web of Data
Licenses in the Web of Data

![Bar Graph]

- Licensed
- Not Licensed

The graph shows a significant difference between licensed and not licensed data sources on the Web of Data.
Licenses Compatibility and Composition

CLIENT QUERY
SELECT...
WHERE{...}

SPARQL QUERY RESULT
XML + <link URI-Lc>

CLIENT QUERY
LICENCES COMPATIBILITY AND COMPOSITION MODULE
LICENCES SELECTION
COMPATIBILITY and COMPLIANCE EVALUATION
LICENCES COMPOSITION
QUERY RESULT
Licenses Compatibility and Composition

1. Combination of Semantic Web languages (machine-readable licenses) - defeasible deontic logic;

2. Extension of existing proposals in service license analysis and CC licenses;

3. Heuristics for licenses composition (SPINdle defeasible reasoner).
Real world example

- **Facts:** \( F = \{ \text{KeepOpen} \} \)
- **Licenses:** \( L = \{ l_{\text{CC-zero}}, l_{\text{ODbl}}, l_{\text{BY-NC-ND}} \} \)
- **Rules:** \( R = \{ \)
  \[ r_1 : l_{\text{ODbl}} \Rightarrow l_{\text{ODbl}} \text{ Attribution} \]
  \[ r_2 : l_{\text{ODbl}} \Rightarrow l_{\text{ODbl}} \text{ ShareAlike} \]
  \[ r_3 : l_{\text{BY-NC-ND}} \Rightarrow l_{\text{BY-NC-ND}} \text{ Attribution} \]
  \[ r_4 : \text{KeepOpen} \Rightarrow l_{\text{CC-zero}} \text{ Reproduction} \]
  \[ r_5 : \text{KeepOpen} \Rightarrow l_{\text{CC-zero}} \text{ Distribution} \]
  \[ r_6 : \text{KeepOpen} \Rightarrow l_{\text{CC-zero}} \text{ Derivative} \]
  \[ r_7 : l_{\text{ODbl}} \Rightarrow l_{\text{ODbl}} \text{ Sharing} \]
  \[ r_8 : l_{\text{ODbl}} \Rightarrow l_{\text{ODbl}} \text{ Adaptation} \]
  \[ r_9 : l_{\text{BY-NC-ND}} \Rightarrow l_{\text{BY-NC-ND}} \text{ Sharing} \]
  \[ r_{10} : l_{\text{BY-NC-ND}} \Rightarrow l_{\text{BY-NC-ND}} \sim \text{Commercial} \]
  \[ r_{11} : l_{\text{BY-NC-ND}} \Rightarrow l_{\text{BY-NC-ND}} \sim \text{Derivative} \]
  \[ r_{12} : \text{Derivative} \rightarrow_{c} \text{Adaptation} \}
- **Priorities on licenses:** \( l_{\text{ODbl}} \succ l_{\text{BY-NC-ND}} \)
Real world example

- Facts: \( F = \{ \text{KeepOpen} \} \)
- Licenses: \( L = \{ \text{CC-zero}, \text{ODbl}, \text{BY-NC-ND} \} \)
- Rules: \( R = \{ r_1 : \Rightarrow \text{ODbl Attribution}, r_2 : \Rightarrow \text{ODbl ShareAlike}, r_3 : \Rightarrow \text{BY-NC-ND Attribution}, r_4 : \text{KeepOpen} \Rightarrow \text{CC-zero Reproduction}, r_5 : \text{KeepOpen} \Rightarrow \text{CC-zero Distribution}, r_6 : \text{KeepOpen} \Rightarrow \text{CC-zero Derivative}, r_7 : \Rightarrow \text{ODbl Sharing}, r_8 : \Rightarrow \text{ODbl Adaptation}, r_9 : \Rightarrow \text{BY-NC-ND Sharing}, r_{10} : \Rightarrow \text{BY-NC-ND \sim Commercial}, r_{11} : \Rightarrow \text{BY-NC-ND \sim Derivative}, r_{12} : \text{Derivative} \rightarrow \text{Adaptation} \} \)
- Priorities on licenses: \( \text{ODbl} \succ \text{BY-NC-ND} \)

\textbf{AND-composition} is not admissible.
\textbf{OR-composition} is admissible: conflict between rule \( r_6 \) and rule \( r_{11} \).

Deontic conclusions:
+ \( \text{ODbl Attribution} \), + \( \text{ODbl ShareAlike} \), + \( \text{ODbl Reproduction} \), + \( \text{ODbl Distribution} \), + \( \text{ODbl Sharing} \), + \( \text{ODbl Adaptation} \), + \( \text{ODbl \sim Commercial} \),
- \( \text{ODbl Derivative} \), - \( \text{ODbl Commercial} \), - \( \text{ODbl Adaptation} \)

@prefix l4lod: http://ns.inria.fr/l4lod/.
@prefix : http://example/licenses.

:\licC a l4lod:License;
\@l4lod:obliges \@l4lod:Attribution;
\@l4lod:obliges \@l4lod:ShareAlike;
\@l4lod:permits \@l4lod:Reproduction;
\@l4lod:permits \@l4lod:Distribution;
\@l4lod:permits \@l4lod:Sharing;
\@l4lod:permits \@l4lod:Adaptation;
\@l4lod:prohibits \@l4lod:Commercial.
Perspectives

Limitations

- Few licenses expressed in machine-readable format
- Need to ensure trustworthiness of contextual information
- I don’t speak RDF, SPARQL, SPINDle syntax...

Future challenges

- Tell me your requirements, I’ll tell you the license which better suits your needs
- License-based access control
References


- Shi3ld http://wimmics.inria.fr/projects/shi3ld/

- Licenses for Linked Open Data http://ns.inria.fr/l4lod/