Privacy for Linked Data

Silvia Ghilezan Svetlana Jakšić Jovanka Pantović

University of Novi Sad

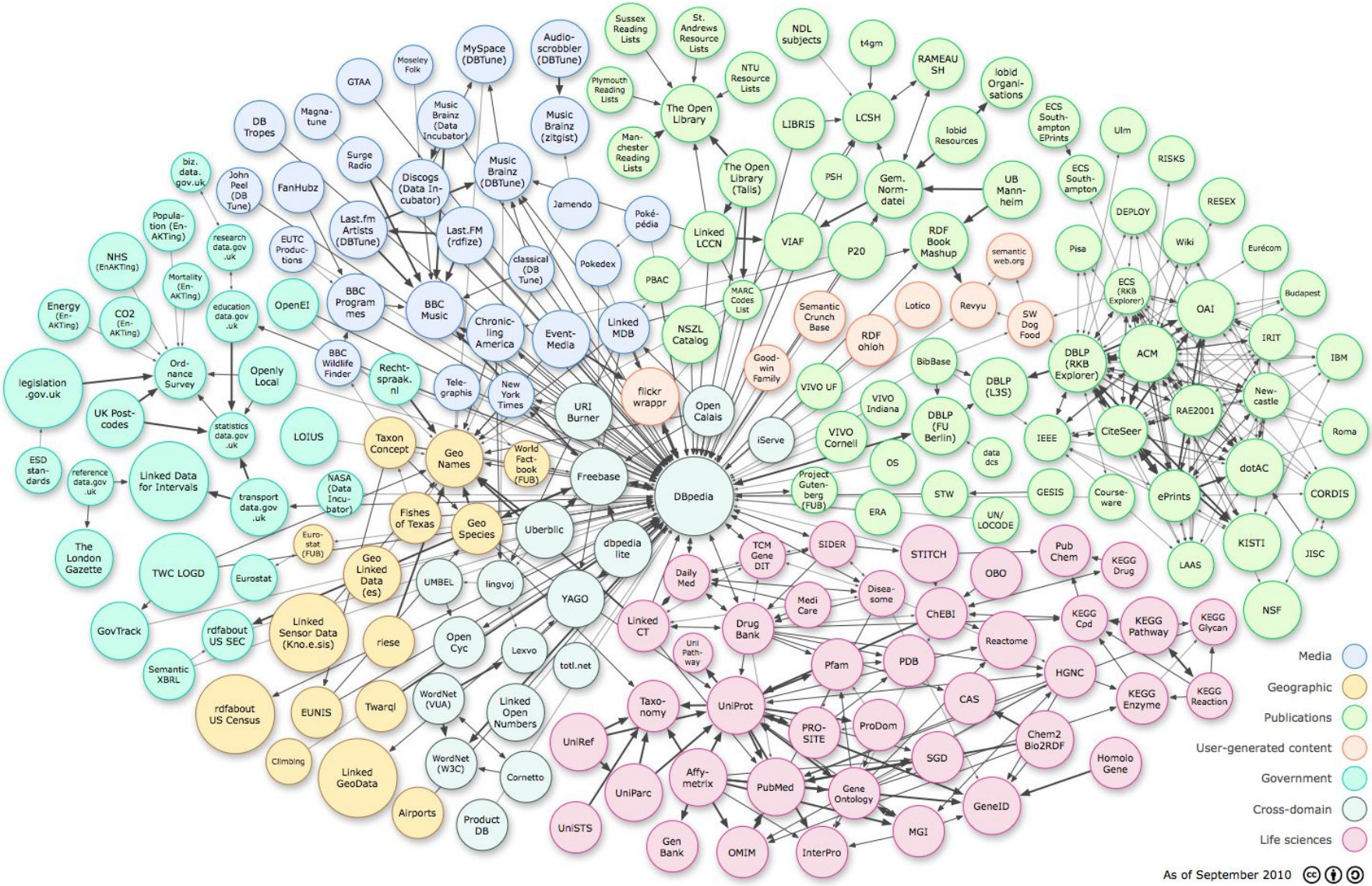
COST IC0901, Haifa

Outline

Linked Data

- Web of Linked Data
- Technologies: URIs (Uniform Resource Identifiers), RDF (Resource Description Framework), SPARQL,...
- W3C project: Semantic Web
  `http://www.w3.org/standards/semanticweb/`
- Published Data: media, publications, life sciences, geographic data, DBpedia, e-government, user-generated content including profiles from social networks and blogs,...

Linked Data Cloud



Privacy

- Alan Westin defined the privacy as “the ability to control who has access to information and to whom that information is communicated"

- Privacy may not include just private status of some data but also significance or no significance of data for some group and ability of readers to understand the data properly.

Related work


Horne, R. and Sassone, V. (2011) A Typed Model for Linked Data. *Technical Report*, available online at http://eprints.ecs.soton.ac.uk/21996/5/paper.pdf.


Horne, R. and Sassone, V. (2011) A Verified Algebra for Linked Data. In Mohammad Reza Mousavi and António Ravara, editors, *FOCLASA*, *EPTCS* 58, 20–33.


Sacco, O. and Passant, A. (2011) A Privacy Preference Ontology (ppo) for Linked Data. In *Proceedings of the Linked Data on the Web Workshop (LDOW2011)*.


Dezani-Ciancaglini, M. and Ghilezan, S. and Jakšić, S. and Pantovič J. (2010) Types for Role-Based Access Control of Dynamic Web Data, In *WFLP'10*, *Lecture Notes in Computer Science* 6559, 1–29.

Idea

- Stored data $\underline{D}$
- Privacy policy D_1
- Users' profiles C
- Queries, processes P

$$\underline{D}^{D_1} \qquad \{P\}_C$$

RDF data

Stored RDF content

- The data in RDF is modelled as a parallel composition of triples (s, p, o).
- Subject s and predicate p represent URIs (names) while object o reperesnts URI or basic data value (literal).
- Stored RDF content is parallel composition of stored triples grouped in default and named graphs.
- Stored triple:

$$\underline{post_1} = \underline{(\texttt{paper}, \texttt{is printed in}, \text{`}\texttt{journal}\text{'})}$$

- Named graph:

$$Blog = \texttt{AB}[\underline{post_1 \otimes post_2 \otimes post_3 \otimes post_4}]$$

RDF data

Stored RDF content

$$\underline{C} ::= \quad \text{stored RDF content (data)}$$
$$\underline{(a, a, o)}^{C} \quad \text{stored triple}$$
$$\underline{C} \parr \underline{C} \quad \text{par}$$

$$G ::= \quad \text{stored graphs}$$
$$a[\underline{C}] \quad \text{graph named } a$$
$$\varepsilon[\underline{C}] \quad \text{the default graph}$$

SPARQL queries

$U ::=$	query
$\tau[(\alpha, \alpha, \gamma)]$	ask
ϕ	constraint
$U \oplus U$	choice
$U \otimes U$	join
$\bigvee u.U$	select name
$\bigvee x.U$	select literal
$*U$	iteration

$\phi ::=$	constraint
I	true
0	false
$\phi \wedge \phi$	and
$\phi \vee \phi$	or
$\neg\phi$	not
$\ldots$	etc.

$P ::=$	pure process
$\bot$	nothing
U	query
$P; P$	then
$\texttt{update}(\tau, (\alpha, \alpha, \gamma), D)$	policy update

Processes

$S ::=$ process

- $\{P\}_C$ pure process with a RDF profile
- G stored RDF graphs
- $S \otimes S$ par

Reduction relation

1. describes and controls the interaction between stored RDF triples and queries;

2. describes the administration of privacy protection policies by processes with profiles.

Reduction

[Ask] $$\frac{C \otimes |D| \rhd C}{\tau\left[\underline{(a,b,c)}^{D}\right] \otimes \{\tau\,[(a,b,c)]\}_{C} \rhd \tau\left[\underline{(a,b,c)}^{D}\right]}$$

[Update] $$\tau\left[\underline{(a,b,c)}^{D}\right] \otimes \{\texttt{update}(\tau,(a,b,c),D_1)\}_{C} \rhd \tau\left[\underline{(a,b,c)}^{D_1}\right]$$

Reduction

[Filter] $$\frac{\vDash \phi}{\{\phi\}_C \rhd \{\bot\}_C}$$

[Weakening] $$\{*U\}_C \rhd \{\bot\}_C$$

[ChooseLeft] $$\frac{S \parr \{U\}_C \rhd S'}{S \parr \{U \oplus V\}_C \rhd S'}$$

[ChooseRight] $$\frac{S \parr \{V\}_C \rhd S'}{S \parr \{U \oplus V\}_C \rhd S'}$$

[Tensor] $$\frac{S \parr \{U\}_C \rhd S' \quad T \parr \{V\}_C \rhd T'}{S \parr T \parr \{U \otimes V\}_C \rhd S' \parr T'}$$

[Dereliction] $$\frac{S \parr \{U\}_C \rhd S'}{S \parr \{*U\}_C \rhd S'}$$

[Contraction] $$\frac{S \parr \{*U \otimes *U\}_C \rhd S'}{S \parr \{*U\}_C \rhd S'}$$

[Guard] $$\frac{S \parr \{P\}_C \rhd S'}{S \parr \{P; Q\}_C \rhd S' \parr \{Q\}_C}$$

[SelectName] $$\frac{S \parr \{U\{\alpha/u\}\}_C \rhd S'}{S \parr \{\vee u.U\}_C \rhd S'}$$

[SelectLiteral] $$\frac{S \parr \{U\{\mu/x\}\}_C \rhd S'}{S \parr \{\vee x.U\}_C \rhd S'}$$

[BlankNode] $$\frac{\tau[\underline{C}] \parr S \rhd \tau[\underline{D}] \parr S'}{\tau[\wedge a.\underline{C}] \parr S \rhd \tau[\wedge a.\underline{D}] \parr S'} \; a \notin \mathtt{fn}(S, S')$$

[Context] $$\frac{S \rhd S'}{S \parr T \rhd S' \parr T}$$

Typing rules

$$\frac{\mathcal{T}(a) = C \quad \vdash C : \mathit{Profile}}{\vdash a : \mathit{Name}(C)}\ (\text{Name}) \qquad \frac{}{\vdash l : \mathit{Literal}}\ (\text{Literal})$$

$$\frac{\Gamma \vdash \alpha : \mathit{Name}(C_1) \quad \Gamma \vdash \beta : \mathit{Name}(C_2) \quad \Gamma \vdash \gamma : \mathit{Name}(C_3) \quad \vdash \tau : \mathit{Name}(E) \quad C_1 \ni D \quad C_2 \ni D \quad C_3 \ni D \quad D \ni E \quad C \otimes |E| \rhd C \quad \vdash C : \mathit{Profile} \quad \vdash D : \mathit{Profile}}{\vdash \texttt{update}(\tau, (\alpha, \beta, \gamma), D) : \mathit{Proc}(C)}\ (\text{UpdateName})$$

$$\frac{\vdash P : \mathit{Proc}(C)}{\vdash \{P\}_C : \mathit{Process}}\ (\text{Proc})$$

$$\frac{\vdash a : \mathit{Name}(C_1) \quad \vdash b : \mathit{Name}(C_2) \quad \vdash c : \mathit{Name}(C_3) \quad \vdash \tau : \mathit{Name}(E) \quad \vdash D : \mathit{Profile} \quad C_1 \ni D \quad C_2 \ni D \quad C_3 \ni D \quad D \ni E}{\vdash \tau[\underline{(a, b, c)}^D] : \mathit{Process}}\ (\text{TripleName})$$

Novelties

After proving that the subject reduction holds, we come to the discussion of crucial properties obtained by the following novelties that we introduced:

1. assigning privacy policies to names;
2. assigning privacy policies to triples;
3. assigning profiles to processes.

Privacy

Privacy properties

- Alan Westin defined the privacy as “the ability to control who has access to information and to whom that information is communicated"

- Privacy may not include just private status of some data but also significance or no significance of data for some group and ability of readers to understand the data properly.

Privacy properties

Theorem

If $\vdash S : \mathit{Process}$ *and* $S \to \tau[\underline{(a,b,c)^D}] \otimes \{P\}_C \otimes S_1$, *then*

(i) *there are RDF contents* C_1, C_2, C_3, E *such that* $\vdash a : \mathit{Name}(C_1)$ *and* $\vdash b : \mathit{Name}(C_2)$ *and* $\vdash c : \mathit{Name}(C_3)$ *and* $\vdash \tau : \mathit{Name}(E)$ *and* $\vdash C : \mathit{Profile}$ *and* $\vdash D : \mathit{Profile}$ *and*

$$D \Supseteq E \text{ and } C_1 \Supseteq D \text{ and } C_2 \Supseteq D \text{ and } C_3 \Supseteq D.$$

(ii) $S \equiv \tau[\underline{(a,b,c)^D}] \otimes \{\tau[(a,b,c)]\}_{D_1} \otimes \{P\}_C \otimes S_1$ *implies* $D_1 \otimes |\overline{D}| \rhd C$.

(iii) $P \equiv \mathtt{update}(\tau, (a,b,c), D_1)$ *implies* $\vdash D_1 : \mathit{Profile}$ *and*

$$D_1 \Supseteq E \text{ and } C_1 \Supseteq D_1 \text{ and } C_2 \Supseteq D_1 \text{ and } C_3 \Supseteq D_1.$$