Exercise 06:

Sticky Policies, DPV, RDF & Turtle

Privacy-Preservation Technologies in Information Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22

Task 1:
Sticky Policies

Privacy-Preservation Technologies in Information Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

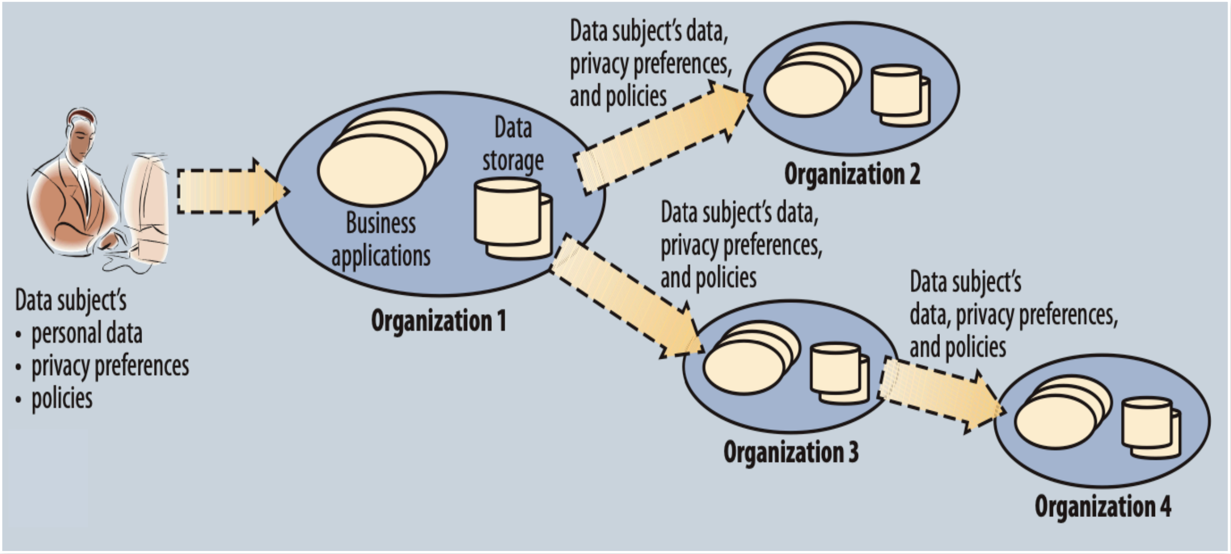
WS 21/22

Definition

- **Sticky policies** represent one approach to improve **user** (Data Subject)' control over their personal **data**.
- In such an approach, **machine-readable** policies are attached to personal data.
- They are called '*sticky*' in that they travel together with data, as data travels across multiple parties

Concept I

- The data subject owns all the access permissions, the encryption keys and the credentials needed for identifying himself/herself and for ciphering the data to be sent to data controller.
  - ➡ Each data controller that wants to decrypt such data has to own the related policies and credentials.
- The data subjects have to trust the data controller (or the data controllers) with which they share policies and credentials.
  - ➡ They have also to trust that their data are collected and shared with other domains in compliance with the agreed rules.
- Such policies/credentials could be updated or revoked and, therefore, a system for their synchronization among all the involved data consumers must be put in place.


Data subject's
• personal data
• privacy preferences
• policies
Business applications
Data storage
Organization 1
Data subject's data, privacy preferences, and policies
Organization 2
Data subject's data, privacy preferences, and policies
Organization 3
Data subject's data, privacy preferences, and policies
Organization 4

Traditional approach

Concept II

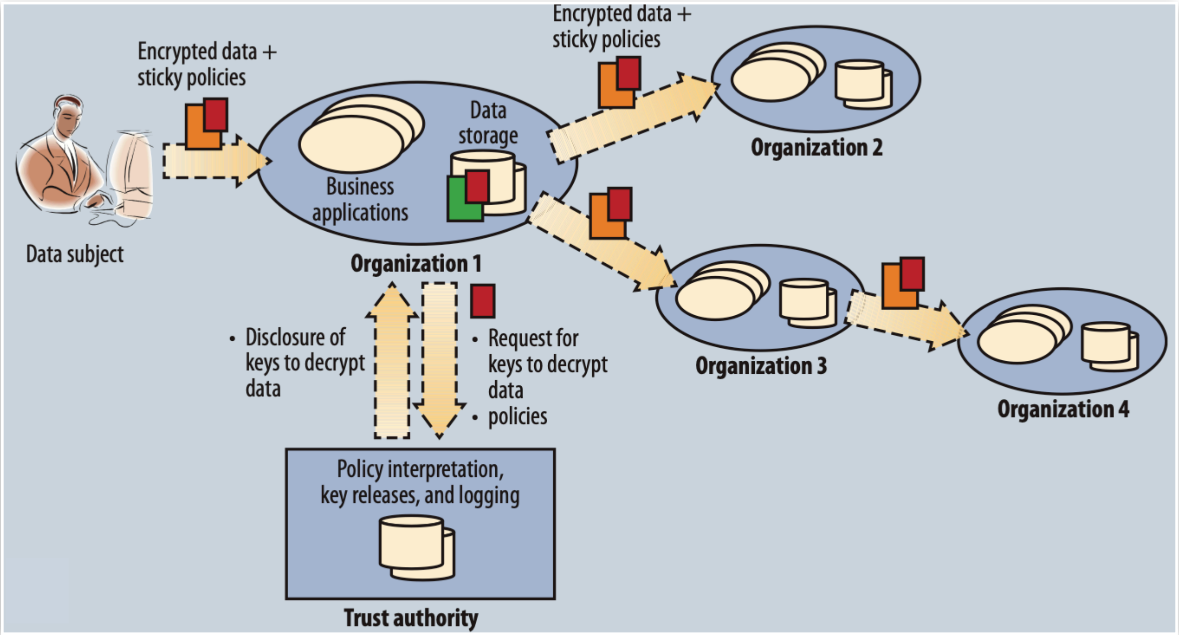
- Data controllers do not own the policies/credentials

  ➡ A trust authority (TA) is responsible for their management.

- The data subject sends them in an encrypted way along with the associated sticky policy; then the data controllers can contact the TA in order to obtain the access permissions on the received data.

- Data subjects have to “trust” the TA itself, but are protected from illegitimate behaviours carried out by data consumers to fulfil their own purposes.

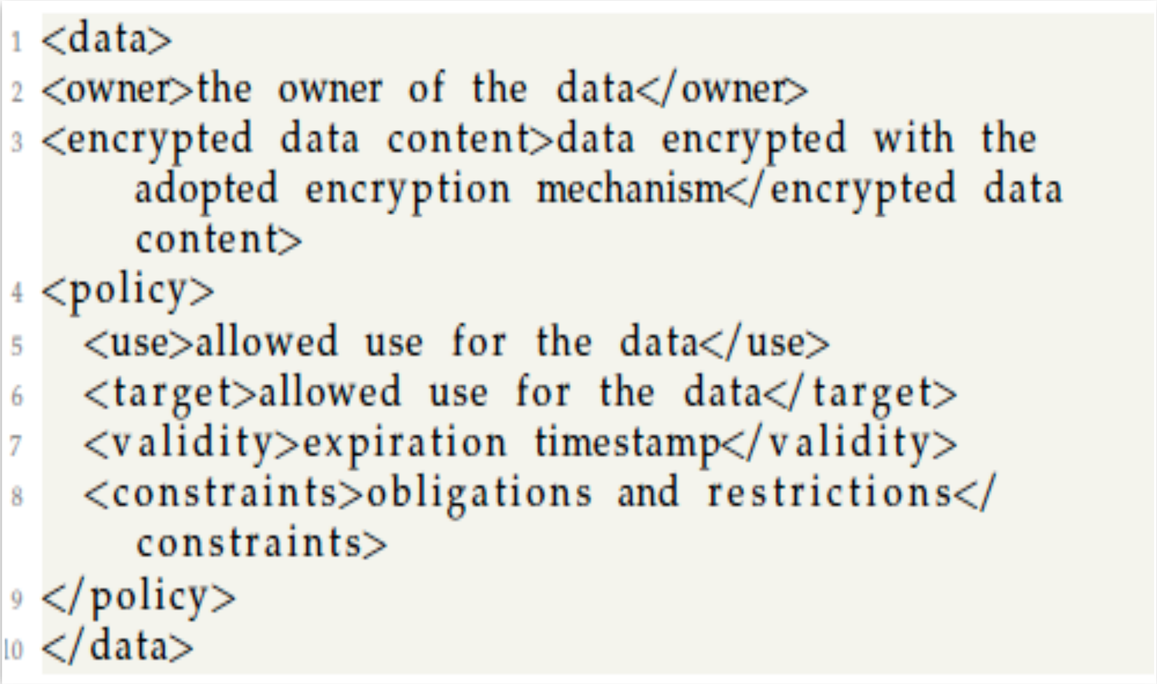
- No synchronization is required among the involved data controllers.



Sticky policy-based approach

Content & XML Example

- The data subject
- The data content
- Purpose use of the data (e.g, for research, translation processing)
- Where and when data will be available (e.g., an expiration time-stamp)
- Specific obligations and restrictions for the parties involved (third parties, people, process).
- Blacklists; notification of disclosure; and deletion or minimization of data after a certain time
- A list of trusted authorities (TAs)



```
<data>
<owner>the owner of the data</owner>
<encrypted data content>data encrypted with the
    adopted encryption mechanism</encrypted data
    content>
<policy>
  <use>allowed use for the data</use>
  <target>allowed use for the data</target>
  <validity>expiration timestamp</validity>
  <constraints>obligations and restrictions</
    constraints>
</policy>
</data>
```

XML Sticky policy example

Advantages

- **Strong enforcement policies**: the data subject has more control over their data in particular, they can establish specific rules about who can access the data and which authorisations have to be allowed, also when information flow across different realms.
- **Traceability**: the flow of the data among different parties can be controlled in a more effective way
  - the whole life-cycle of data processing can be monitored from its generation to its transmissions and disclosure.
- **Management costs reduction**: A possible reduction of the policies' management costs for the organisations, companies or businesses, since third parties will be in charge of setting up the policy enforcement system.
  - TA(s) could have the responsibility of supervising all access requests and permissions;
- **Offline management of the policies**: the possibility for data subjects to perform an offline management of the policies because they are attached to the data/resources, thus they must not be necessarily accessed in an online mode.
- **System Reliability**: an overall improvement of the reliability of the system in terms of confidentiality, preventing unauthorised data disclosure to not trusty third parties.

Disadvantages

- **No standard**: it is difficult to establish an adequate set of policies
  - ➡ Structured and well-defined policy vocabulary, in a way that the involved parties may agree on a set of standard ones.
  - ➡ Definition of a taxonomy or an ontology includes the two following aspects:
    - the classification of the policies.
    - the semantics of the policies.
- **Scalability**: policies travels across multiple parties with the data
  - ➡ An increase in the amount of information transmitted.
  - ➡ An increased computation complexity in the processing of the data chunks at each transmission step.
- **Redundancy**: many copies of the same information may be stored/cached in different hosts belonging to the system. This naturally points out the problem of policy propagation in the different parts of the system itself.

Task 2:

Interoperability

Privacy-Preservation Technologies in Information Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22

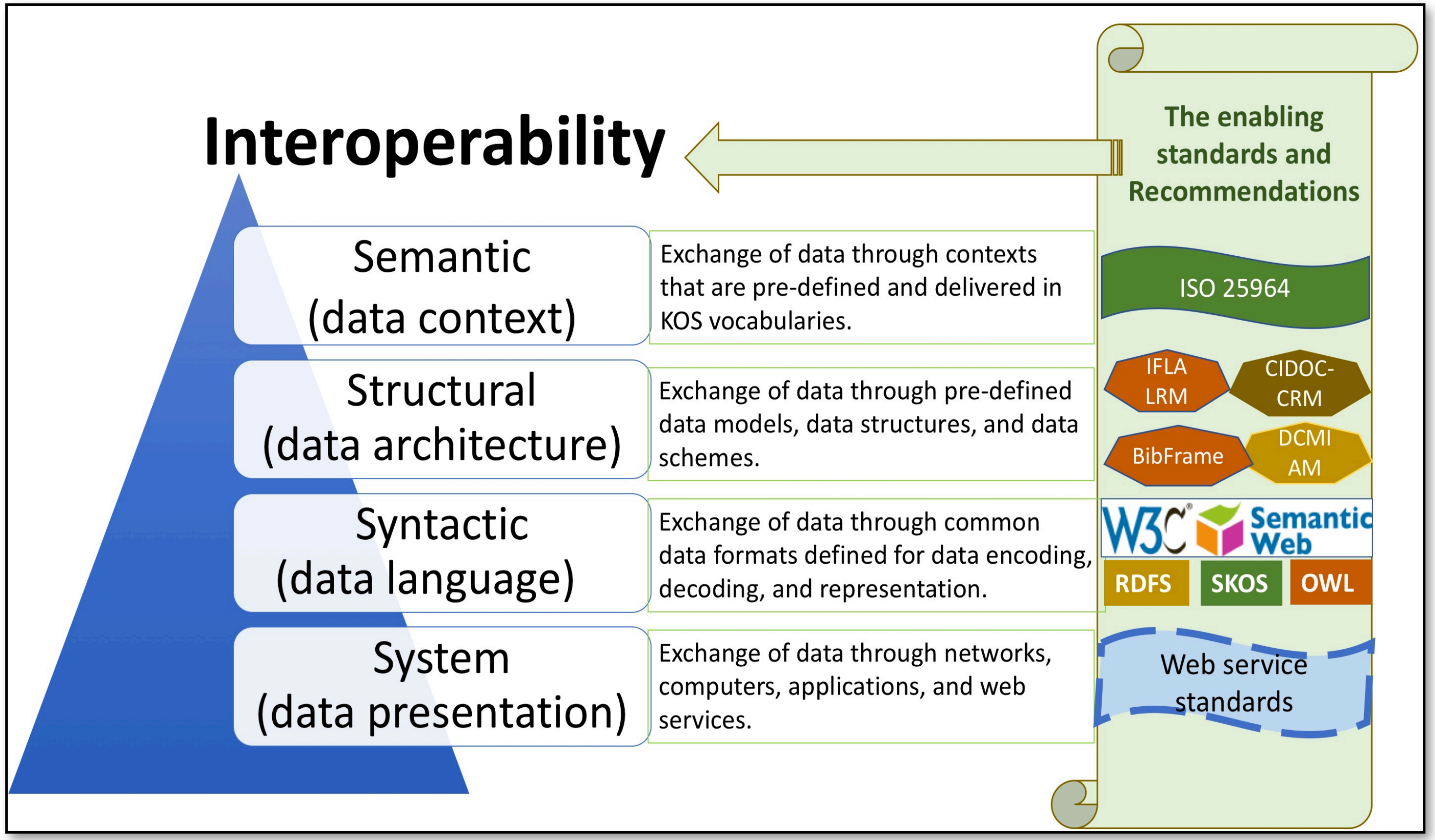
Implementation Reasons

- It's required by law if you collect personal information from users
- It's required by third-party services you may use
- Users are interested in their privacy
- Because there is too much to risk

→ Transfer of Privacy Languages between companies

Interoperability

Interoperability is the ability of two or more components or systems to exchange information and to use the information that has been exchanged.



Interoperability

Semantic (data context)	Exchange of data through contexts that are pre-defined and delivered in KOS vocabularies.
Structural (data architecture)	Exchange of data through pre-defined data models, data structures, and data schemes.
Syntactic (data language)	Exchange of data through common data formats defined for data encoding, decoding, and representation.
System (data presentation)	Exchange of data through networks, computers, applications, and web services.

Source: https://www.isko.org/cyclo/interoperability

Interoperability

- The granularity of data to be shared
- The matching of people to assure the most effective and efficient outcomes
- The continuum of participants in the flow of data
- The business workflow (to reduce the likelihood of introducing a nonautomated step into the flow of data)

Task 3:

DPV

Privacy-Preservation Technologies in Information Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22

Data Privacy Vocabulary (DPV)

Definition

DPV provides terms (classes and properties) to describe and represent information related to processing of personal data based on established requirements such as for the EU General Data Protection Regulation (GDPR).

Benefits

The DPV is useful as a machine-readable representation of personal data processing and can be adopted in relevant use-cases such as legal compliance documentation and evaluation, policy specification, consent representation and requests, taxonomy of legal terms, and annotation of text and data.

W3C

- The World Wide Web Consortium (W3C) is an international community that develops open standards to ensure the long-term growth of the Web (e.g, Data Privacy Vocabulary).
- Consists of industries, researches, politicians, etc.
- Proposed of the first draft of Data privacy Vocabulary (v02).

DPV — Custom Privacy Category

- DPV is not complete and can be updated
- It is possible to extend the PDV with custom privacy categories.

Task 4:

RDF & Turtle

Privacy-Preservation Technologies in Information Systems

Dr. Wiem Fekih Hassen / Dr. Armin Gerl / Felix Bölz

WS 21/22

Syntactical vs. Semantic Interoperability

Semantic (data context)

The data is not only exchanged between two or more systems but also understood by each system.

Syntactic (data language)

Syntactic interoperability allows two or more systems to communicate and exchange data, however, the interface and programming languages are different (e.g, RDF, RDFS and OWL).

→ Semantic interoperability is more desirable than syntactic interoperability.

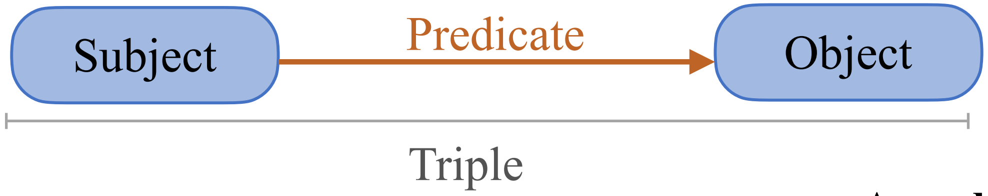
RDF — Overview I

Definition

- The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web.

- It represents metadata about Web resources, such as the title, author and modification date of a Web page, based on identifying things using Web identifiers and describing resources in terms of simple properties and property values.

Terminology



- A **triple** is an edge.
- A **subject** is the source node.
- A **predicate** is the edge name.
- An **object** is the target node.

RDF — Overview II

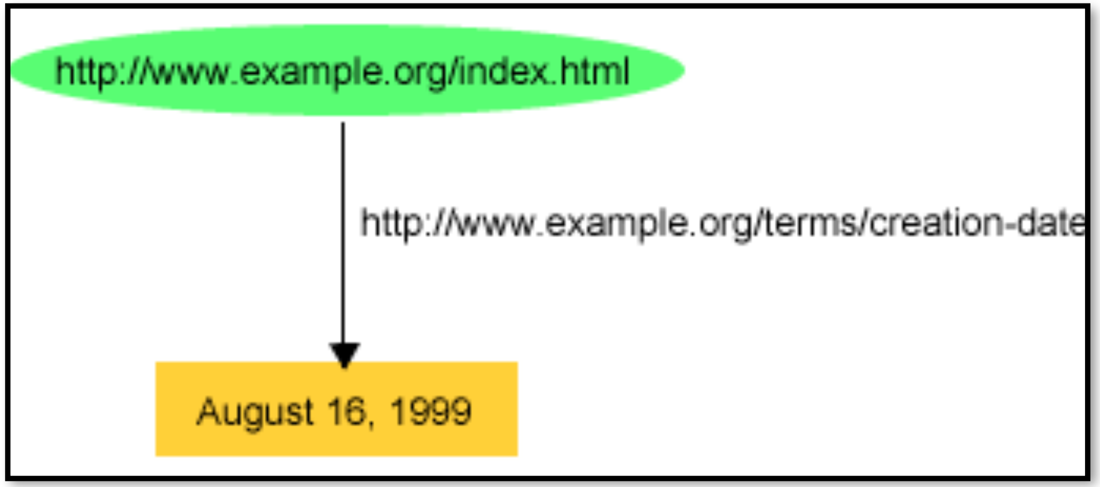
RDF Nodes

- *Uniform Resource Identifiers* (*IRI)*: an IRI is a unicode string for identifying nodes and edges in an unambiguous way. IRIs are internationalized versions of URIs which are generalizations of URLs.

- *Blank node*: Nodes without a user-visible identifier are called blank nodes. A blank node is appropriate when the node does not need to be referenced directly. Blank nodes can be reached by following its incident edges from other nodes.

- *Literal*: Literals are concrete values used to represent datatypes like strings, numbers, and dates.

RDF — Triple Example


http://www.example.org/index.html
http://www.example.org/terms/creation-date
August 16, 1999

Graph 1

Subject: `http://www.example.org/index.html`

Predicate: `http://www.example.org/terms/creation-date`

Object: `August 16, 1999`

Turtle — Graph 1

Description

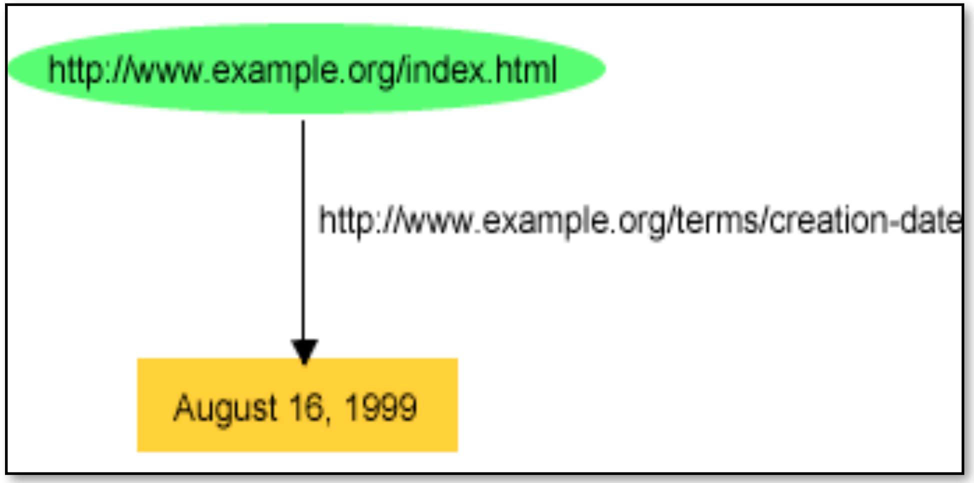
http://www.example.org/index.html has a **creation-date** whose value is **August 16, 1999**

Triple notation

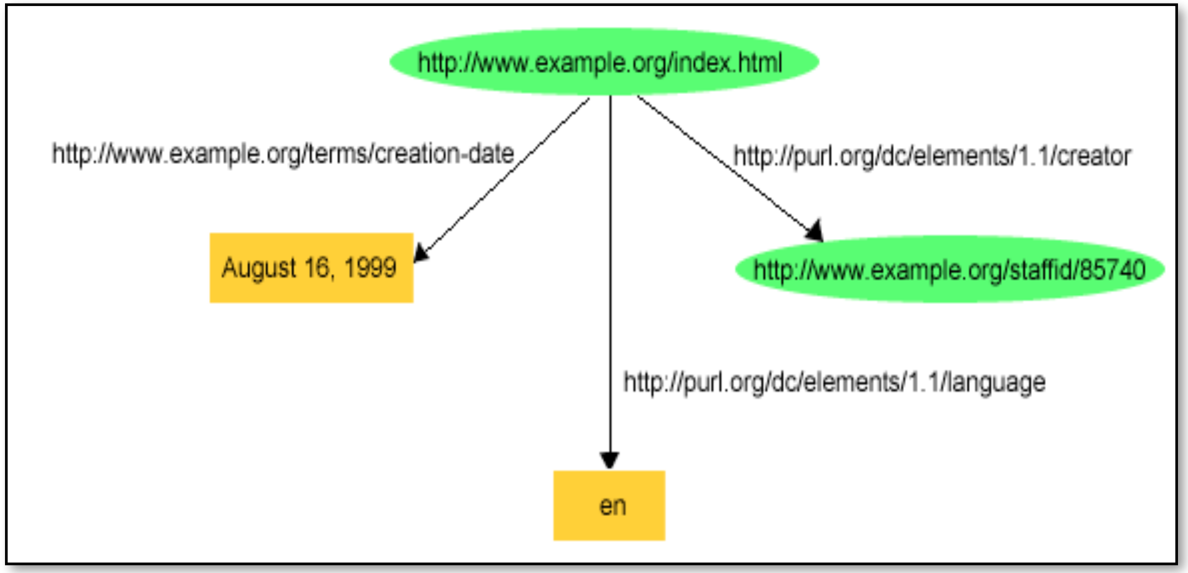
```
<http://www.example.org/index.html> <http://www.example.org/terms/creation-date> "August 16, 1999" .
```

Turtle syntax

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix exterms: <http://www.example.org/terms/>.
<http://www.example.org/index.html>
   exterms:creation-date "August 16, 1999".
```


http://www.example.org/index.html
http://www.example.org/terms/creation-date
August 16, 1999

Turtle — Graph 2


http://www.example.org/index.html
http://www.example.org/terms/creation-date
http://purl.org/dc/elements/1.1/creator
August 16, 1999
http://www.example.org/staffid/85740
http://purl.org/dc/elements/1.1/language
en

Description

http://www.example.org/index.html has a **creator** whose value is http://www.example.org/staffid/85740

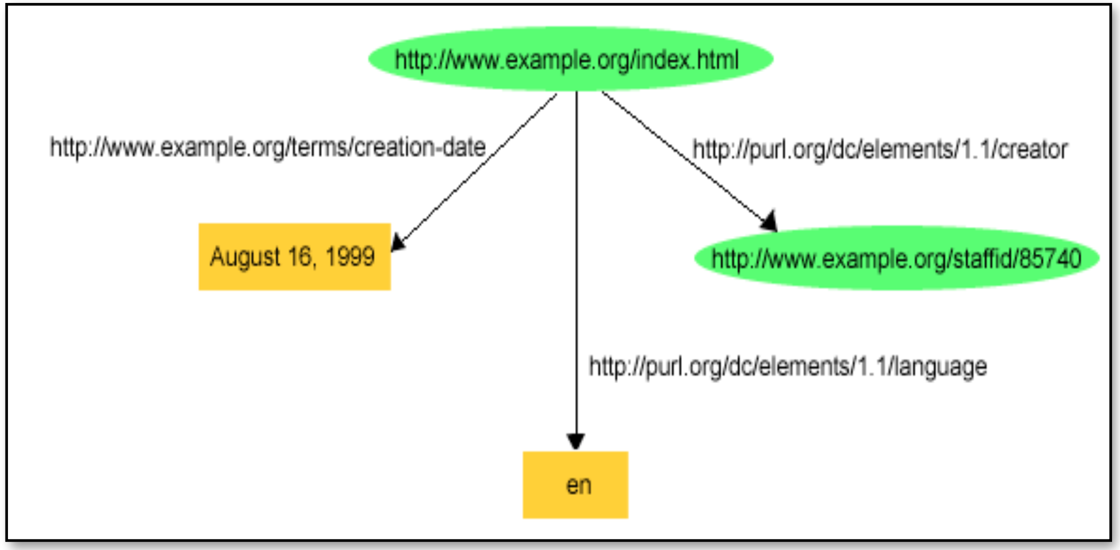
http://www.example.org/index.html has a **creation-date** whose value is **August 16, 1999**

http://www.example.org/index.html has a **language** whose value is **English**

Triple notation

```
<http://www.example.org/index.html> <http://purl.org/dc/elements/1.1/creator> <http://www.example.org/staffid/85740>
<http://www.example.org/index.html> <http://www.example.org/terms/creation-date> "August 16, 1999" .
<http://www.example.org/index.html> <http://purl.org/dc/elements/1.1/language> "en" .
```

Turtle — Graph 2

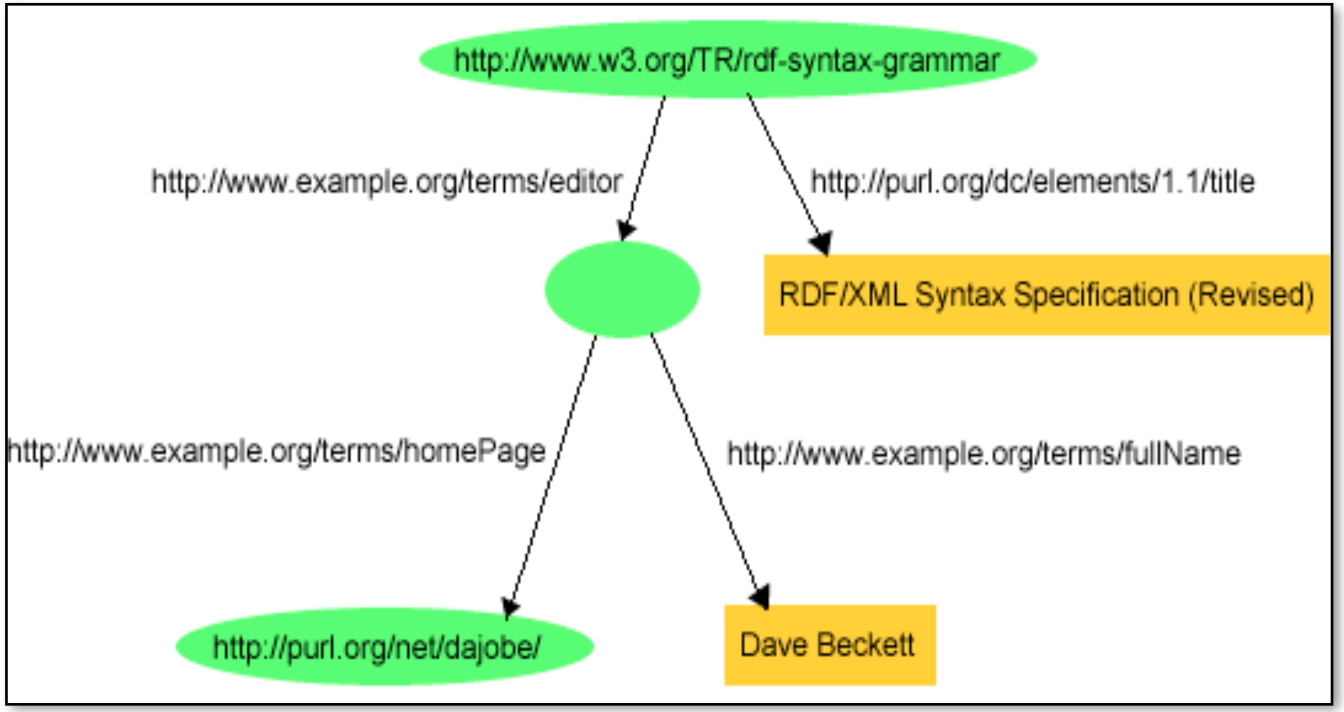

http://www.example.org/index.html
http://www.example.org/terms/creation-date
http://purl.org/dc/elements/1.1/creator
August 16, 1999
http://www.example.org/staffid/85740
http://purl.org/dc/elements/1.1/language
en

Turtle syntax

```
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix dc:      <http://purl.org/dc/elements/1.1/#>.
@prefix exterms: <http://www.example.org/terms/>.

<http://www.example.org/index.html>
    exterms:creation-date "August 16, 1999";
    dc:language "en";
    dc:creator <http://www.example.org/staffid/85740>.
```

Turtle — Graph 3


http://www.w3.org/TR/rdf-syntax-grammar
http://www.example.org/terms/editor
http://purl.org/dc/elements/1.1/title
RDF/XML Syntax Specification (Revised)
http://www.example.org/terms/homePage
http://www.example.org/terms/fullName
http://purl.org/net/dajobe/
Dave Beckett

Description

http://www.w3.org/TR/rdf-syntax-grammar has a title **RDF/XML Syntax Specification (Revised)** and has an editor, the editor has a name **Dave Beckett** and a home page **http://purl.org/net/dajobe/.**

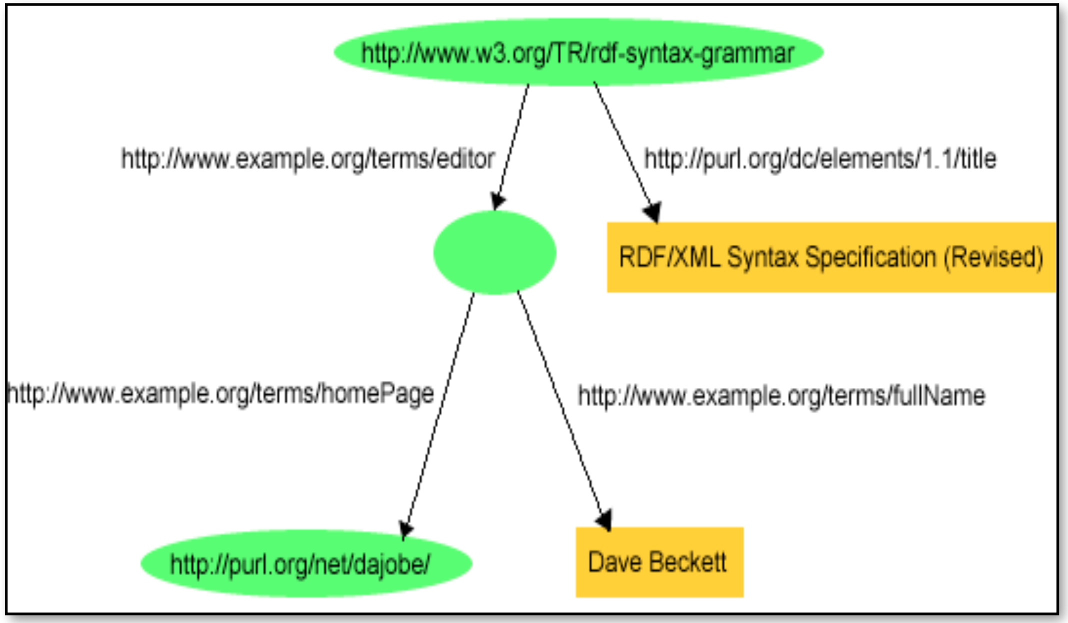
Turtle — Graph 3

Turtle syntax

```
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix dc:       <http://purl.org/dc/elements/1.1/#>.
@prefix exterms: <http://www.example.org/terms/>.

<http://www.w3.org/TR/rdf-syntax-grammar>
    dc:title "RDF/XML Syntax Specification (Revised)";
    exterm:editor _:abc.

_:abc
    exterms:fullName "Dave Beckett";
    exterms:homePage <http://purl.org/net/dajobe/>.
```


http://www.w3.org/TR/rdf-syntax-grammar
http://www.example.org/terms/editor
http://purl.org/dc/elements/1.1/title
RDF/XML Syntax Specification (Revised)
http://www.example.org/terms/homePage
http://www.example.org/terms/fullName
http://purl.org/net/dajobe/
Dave Beckett

See you next week ☺