

Lo que falta en XML...

```
<inmueble>  
  < titulo>miApartamento</titulo>  
  < precio>$200.000</precio>  
</inmueble>
```

Qué representa?

Lo que falta en XML...

```
<inmueble>  
  < titulo>miApartamento</titulo>  
  < precio>$200.000</precio>  
</inmueble>
```

Qué representa?

- Un inmueble es miApartamento más un precio de \$200.000 ?
- miApartamento es un inmueble cuyo precio es \$200.000 ?

Lo que falta en XML...

```
<inmueble>  
  < titulo>miApartamento</titulo>  
  < precio>$200.000</precio>  
</inmueble>
```

Cómo se relacionan las etiquetas?

Cuál es el elemento principal y cuáles sus atributos?

XML no es suficiente!!

Por que XML no es suficiente!

XML

- Estructura de **árbol**
- Hijos **ordenados**
- Esquemas **no permiten extensibilidad distribuída** (agregar atributos a un XML-Schema implica que todos los que estan usando ese XML-Schema esten de acuerdo con los cambios)
- El nivel de **representacion conceptual** de XML es **ambiguo**

Ejemplo:

```
<inmueble>  
  < titulo>miApartamento</titulo>  
  < precio>$200.000</precio>  
</inmueble>
```

De DATOS a CONOCIMIENTO

Si recuperamos datos de la web:

ej. Facultad de Ingeniería UdelaR

The image shows a Google search interface for 'facultad de ingeniería udelar'. The search results include a link to 'www.fing.edu.uy' with the title 'Facultad de Ingeniería: Inicio' and a snippet about building closures. Below the search bar, there are sections for 'Carreras de Grado', 'Departamento de Bedelía', 'Posgrado y Educación ...', 'Enseñanza', 'Institutos', and 'COVID-19: Comunicados e ...'. On the right side, a 'Panel de Conocimiento' (Knowledge Panel) is visible, featuring a map of Playa Ramírez, a photo of the building, and the text 'Temporarily closed'. A large black circle highlights the Knowledge Panel area, with an arrow pointing to the 'Panel de Conocimiento' label at the top.

Google

facultad de ingeniería udelar

Panel de Conocimiento

About 440,000 results (0.56 seconds)

www.fing.edu.uy Translate this page

Facultad de Ingeniería: Inicio

Udelar resuelve continuar con el cierre de los edificios. El Consejo Directivo Central de la Universidad de la República resolvió mantener hasta el 19 de mayo ...

Results from fing.edu.uy

Carreras de Grado
Ingeniería Civil - Ingeniería Eléctrica - Ingeniería Química

Departamento de Bedelía
En virtud de la Emergencia Sanitaria Nacional, la atención ...

Posgrado y Educación ...
Aviso: Inscripciones a cursos de Posgrado 2019.

Enseñanza
Carreras de Grado - Extensión - Generación de Ingreso - ...

Institutos
... de Ingeniería. Universidad de la República - Uruguay. Inicio ...

COVID-19: Comunicados e ...
... de situación. Enviado el 19 de Marzo de 2020, por Area de ...

Playa Ramírez

See photos

See outside

Temporarily closed

Facultad de Ingeniería (Universidad de la República) (Facultad de Ingeniería)

Website Directions Save

Public university in Montevideo

Address: Ave Julio Herrera y Reissig 565, 11300 Montevideo, Departamento de Montevideo

Phone: 2714 2714

Founded: 1888

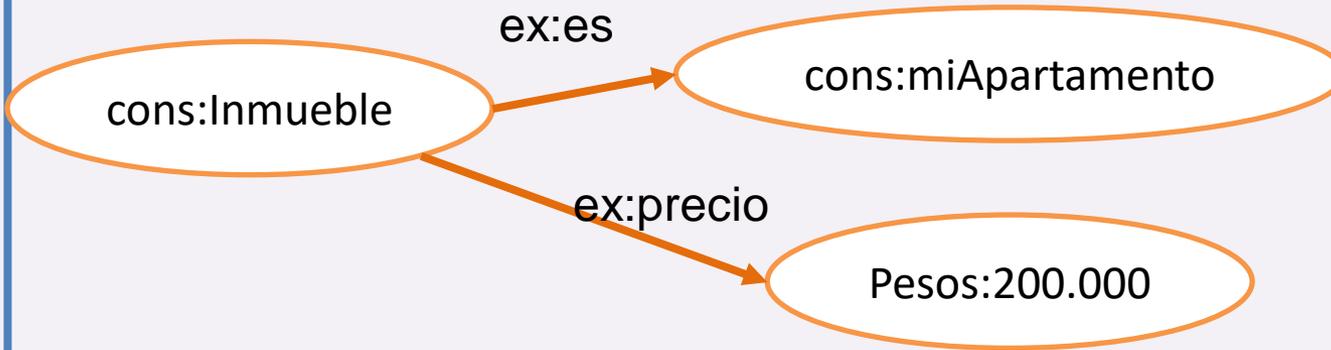
Suggest an edit · Own this business?

Grafo de Datos

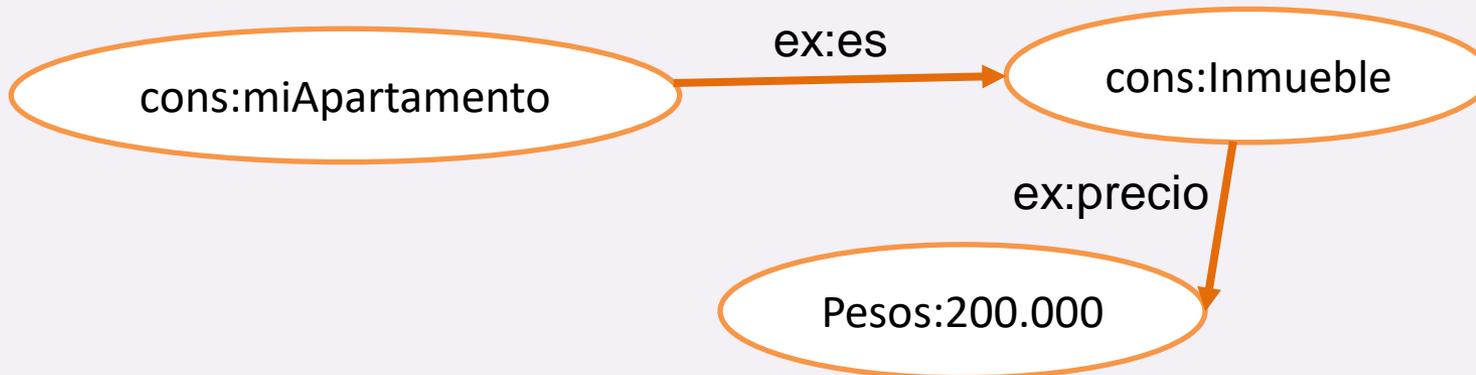
El panel de conocimiento de Google es el resultado de agrupar hechos, personas y lugares, en relación a una consulta para crear resultados de búsqueda interconectados. Un modelo sencillo que represente las relaciones que conectan a distintos recursos de la web es ***un grafo dirigido con aristas etiquetadas***.



Un inmueble es miApartamento más un precio de \$200.000



miApartamento es un inmueble cuyo precio es \$200.000



Características de los Grafos de Datos: FAIR

**FAIR se refiere a los términos en inglés
de:**

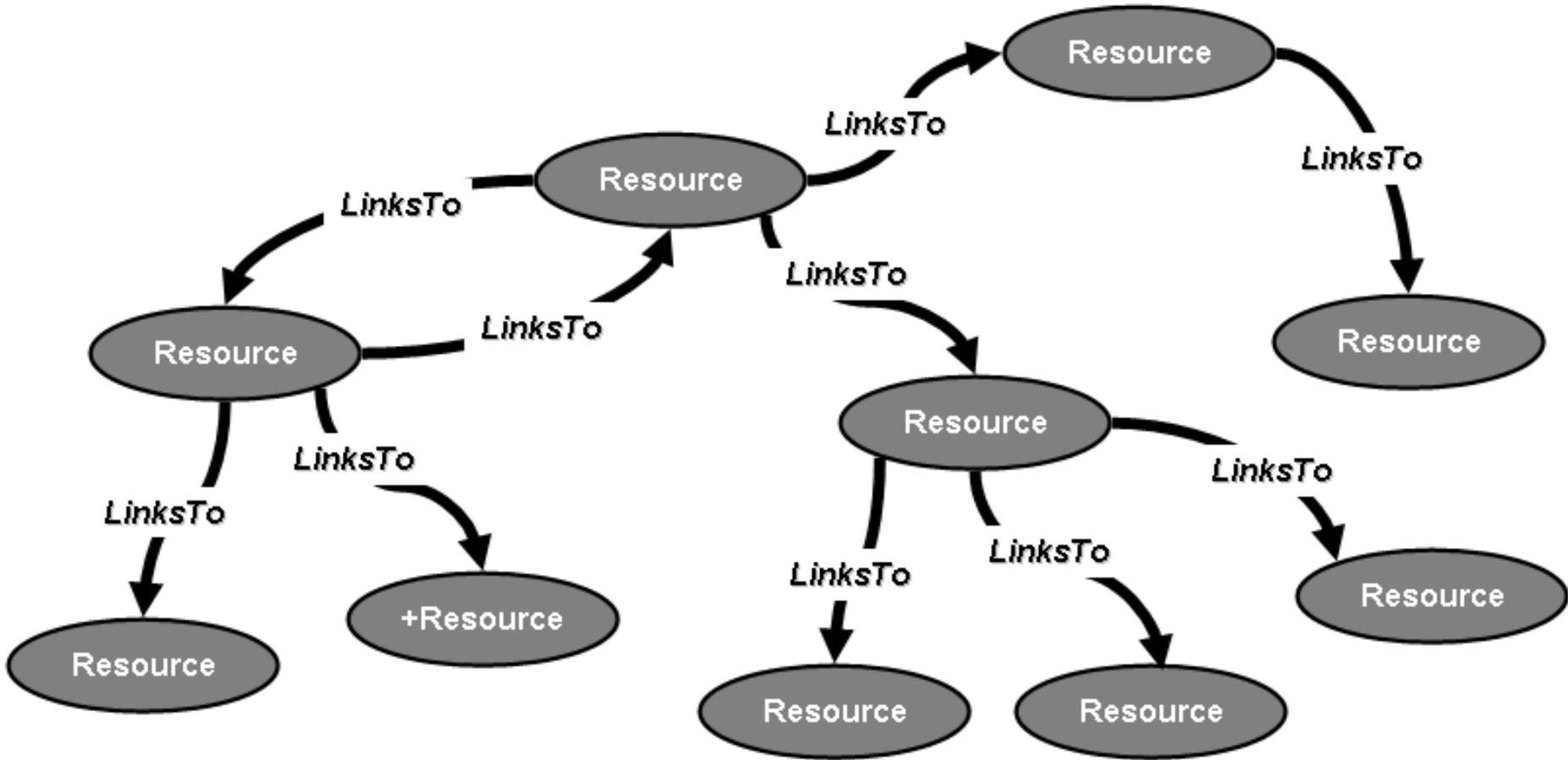
(Findable) Encontrable

Accessible

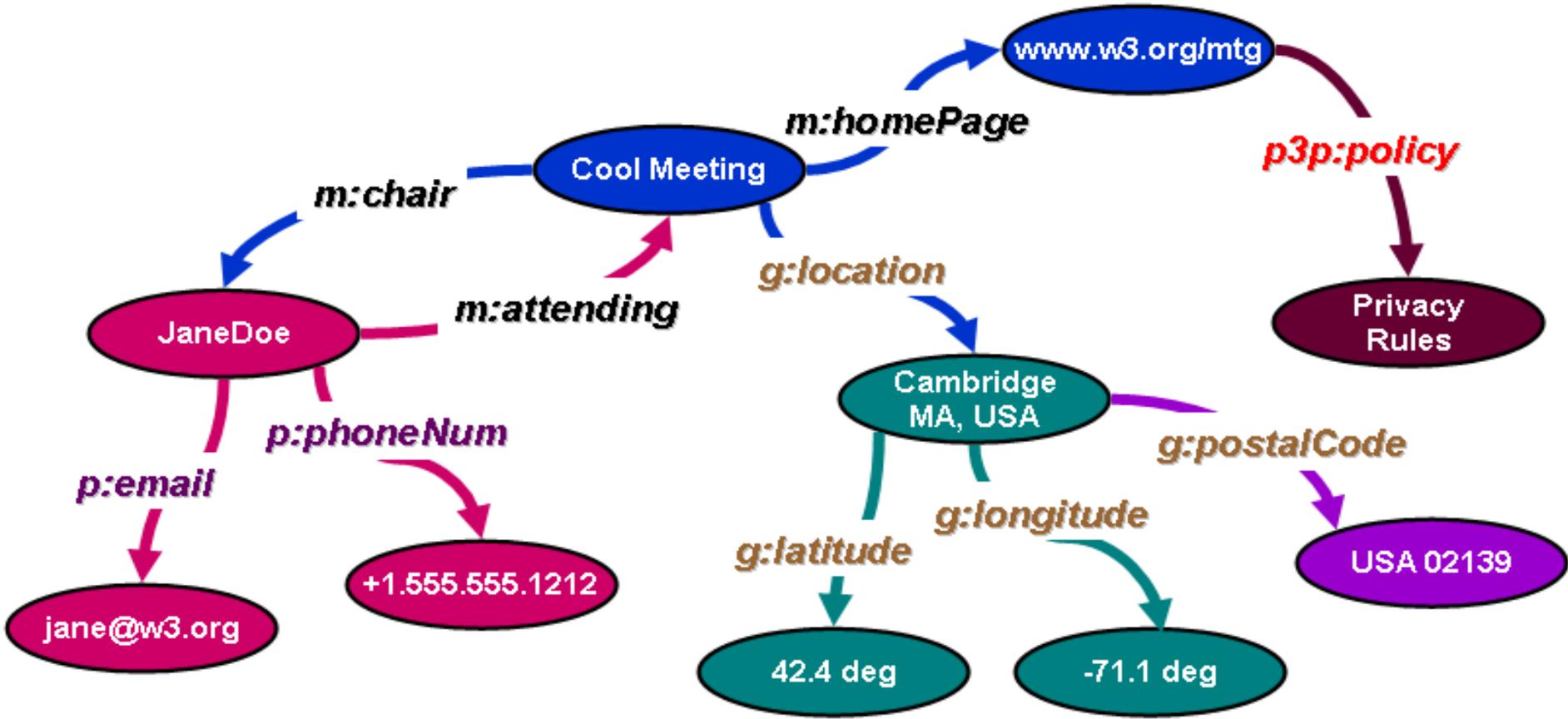
Interoperable

Reusable

Web de recursos



Web de datos



DATOS ENLAZADOS (LINKED DATA)

- **Identificación: Persistent URIs**
- **Protocolos de acceso (HTTP)**

+ **Metadata (RDF)**

Publicar Datos Abiertos

1. Disponible en la web, cualquier formato, pero con licencia abierta
2. Disponible pero en formato procesable
(ej. planilla excel en lugar de la imagen escaneada de la planilla)
3. como el nivel 2 pero no-propietario (ej. CSV en lugar de excel)
4. (3) más la utilización de estándares de W3C (RDF) para identificar “cosas”
5. Enlazar los datos con otros datos externos para dar contexto.

Wikipedia → DBpedia

Ver:

<http://dbpedia.org/page/Montevideo>

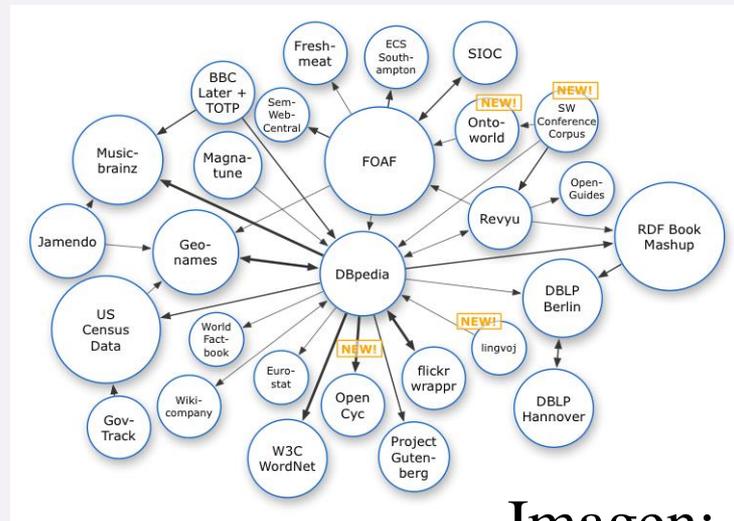


Imagen: linkeddata.org

Nuevos retos

- Legalidad
- Privacidad
- Temporalidad
- Inferencias

Buenas Prácticas

- ★ hacer que sus cosas estén disponibles en la web (en cualquier formato)
- ★★ hacer que esté disponible como datos estructurados (por ejemplo, Excel en lugar de escanear una imagen de una tabla)
- ★★★ formato no propietario (por ejemplo, csv en lugar de excel)
- ★★★★ use URL para identificar cosas, de modo que las personas puedan señalar sus cosas
- ★★★★★ vincular sus datos con los datos de otras personas para proporcionar contexto

Grafo de Datos vs. Grafo de Conocimientos

Declaraciones simples pueden acumularse como aristas en el grafo de datos. Para que el grafo de conocimientos pueda acumular declaraciones cuantificadas ("Todos los edificios tienen ubicación"), entonces es necesaria una base de conocimientos.

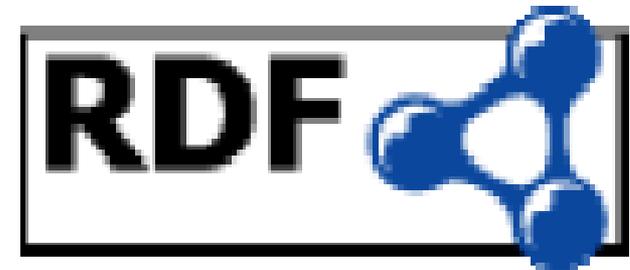


El grafo de conocimientos tiene además del grafo de datos un mecanismo que le permite hacer inferencias.

Las inferencias pueden ser :

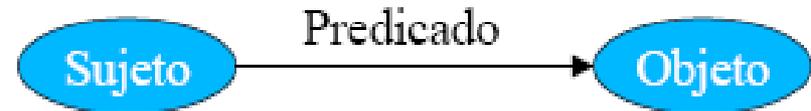
- inductivas utilizando redes neuronales
- deductivas utilizando bases de conocimiento.

Modelo de datos: Resource Description Framework



La información es un **grafo dirigido etiquetado**

- **Nodos**: recursos
- **Arcos**: propiedades



Notación de triplas: (Sujeto, Predicado, Objeto)

El Modelo de Datos RDF

- **Triplas o Sentencias:**
- **<subject, predicate, object>**
- **Describen propiedades de recursos**
- **Un recurso es una URI representando una clase de objetos:**
- **Las propiedades son también recursos (URIs)**

FUENTE: RDF Primer

<http://www.w3c.org/TR/rdf-primer/>

URIs

- URI = Uniform Resource Identifier
- URLs (Uniform Resource Locators) (usadas para acceder a la páginas de la Web)
- En RDF, URIs son normalmente URLs , apuntando a partes específicas de un documento:
- `http://www.somedomain.com/some/path/to/file#fragmentID`

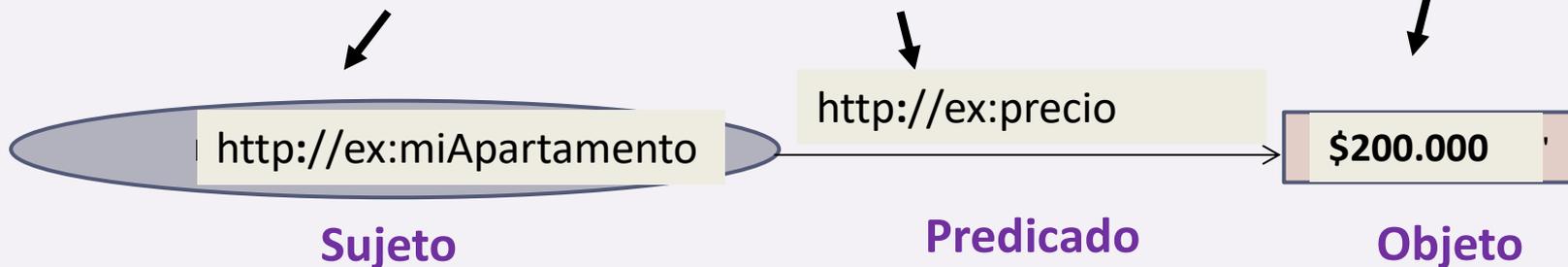
RDF: Resource Description Framework

“MiApartamento es un inmueble cuyo precio es \$200.000”

Al modelar con RDF aparecen 2 sentencias:

1. `ex:miApartamento` `df:type` `ex:Inmueble`

2. `ex:miApartamento` `ex:precio` “\$200.000”



Triplas: (Sujeto, Predicado, Objeto)

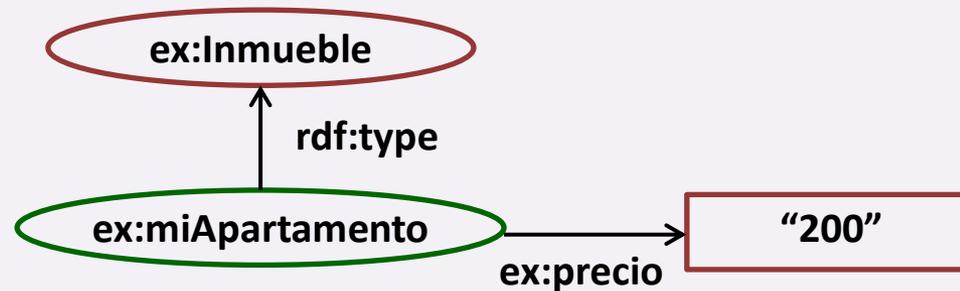
La información es un **grafo dirigido etiquetado**:

- Nodos: **recursos**
- Arcos: **propiedades** - también recursos (URIs)

Ejemplo 1- RDF: Resource Description Framework

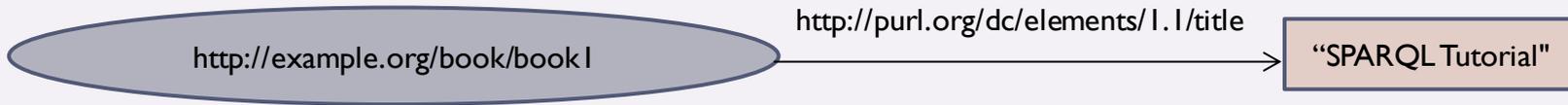
“In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean”

XML and the Second-Generation Web. Jon Bosak, Tim Bray. Scientific American, Mayo 1999



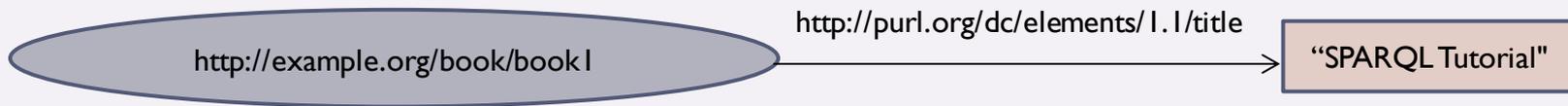
XML modela documentos, pero el mundo real NO son documentos, es una **red de objetos y relaciones** → **RDF** modela esta red

Ejemplo 2- RDF: Resource Description Framework



- En **XML**: es **ambiguo** determinar qué es el sujeto, predicado y objeto → depende de la interpretación del esquema, ya que el anidamiento de las **etiquetas no tiene un significado estándar**.
- **En el grafo** representado con el modelo RDF, siempre los elementos de la tripla tienen significado fíco: (Sujeto, Predicado, Objeto)
 - <http://example.org/book/book1> es el **sujeto**
 - <http://purl.org/dc/elements/1.1/title> es el **predicado**
 - “SPARQL Tutorial” es el **objeto**
- **RDF**: es un modelo que tiene una **semántica** consensuada.

Estructura de Ejemplo2- RDF: Resource Description Framework



Triplas:

Sujeto

Predicado

Objeto

Sujeto: Es un **nodo** en el grafo (puede ser Recurso o un nodo en blanco)

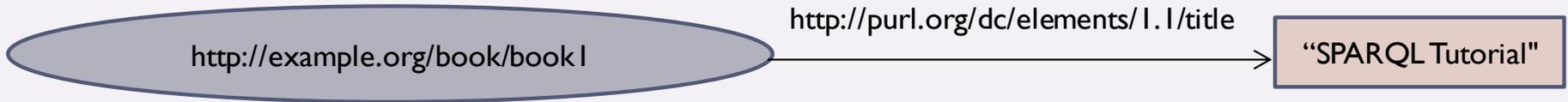
Predicado: Es una **Propiedad** (también es un recurso), corresponde a etiquetas en los arcos dirigidos. **Relación binaria entre dos recursos.**

Objeto: Es un **nodo** en el grafo, puede ser Recurso, literal o nodo en blanco .

En el grafo:

Arcos: representan **sentencias**

LITERALES en RDF: Resource Description Framework



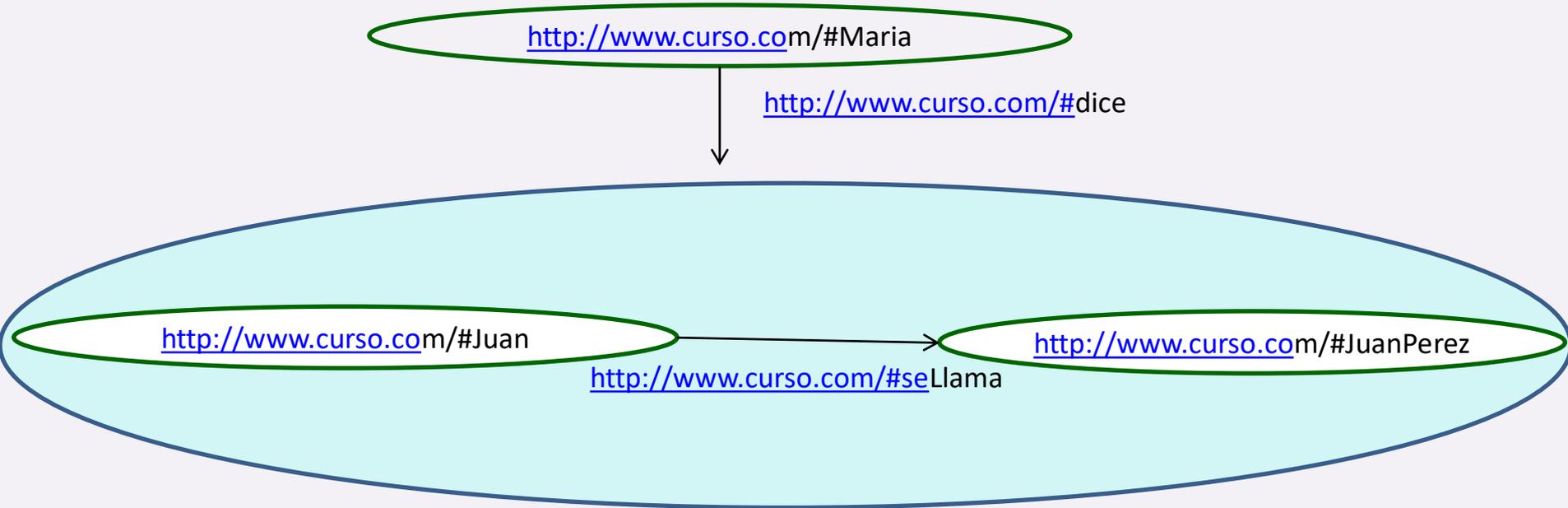
Literales

- **Simple**s – “SPARQL Tutorial”
- **Con anotaciones** - “SPARQL Tutorial”@en-GB
- **Con tipos** – “SPARQL Tutorial”^^xsd:string

Tipo de dato XML-Schema

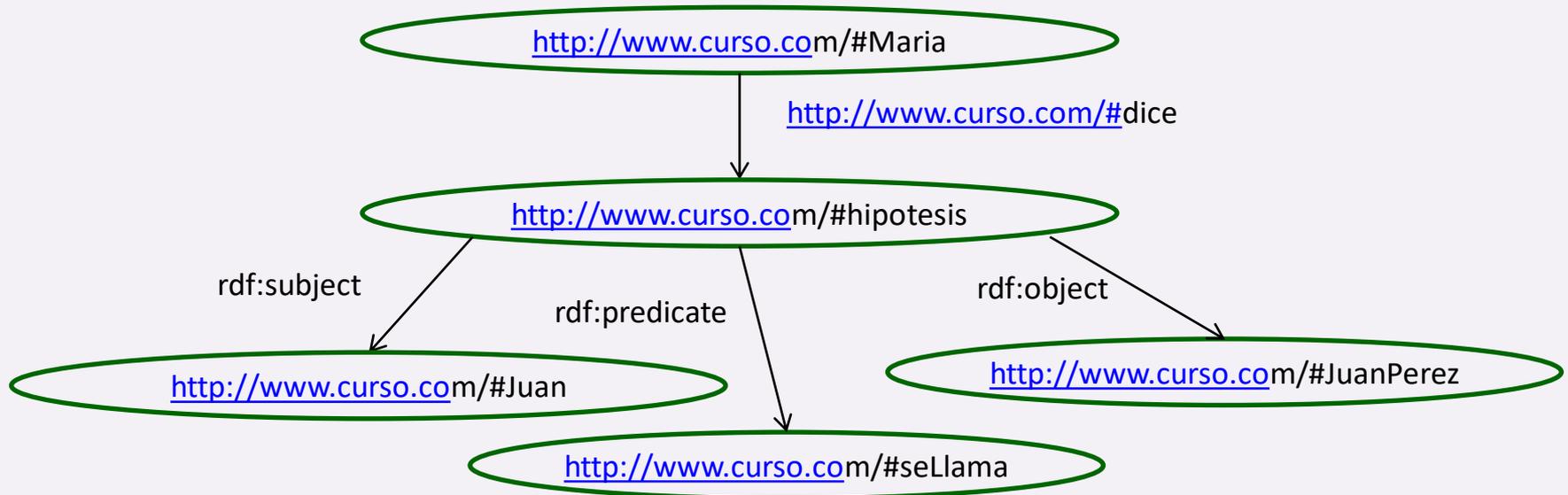
Reificación: sentencias sobre sentencias

“María dice que Juan se llama Juan Pérez”



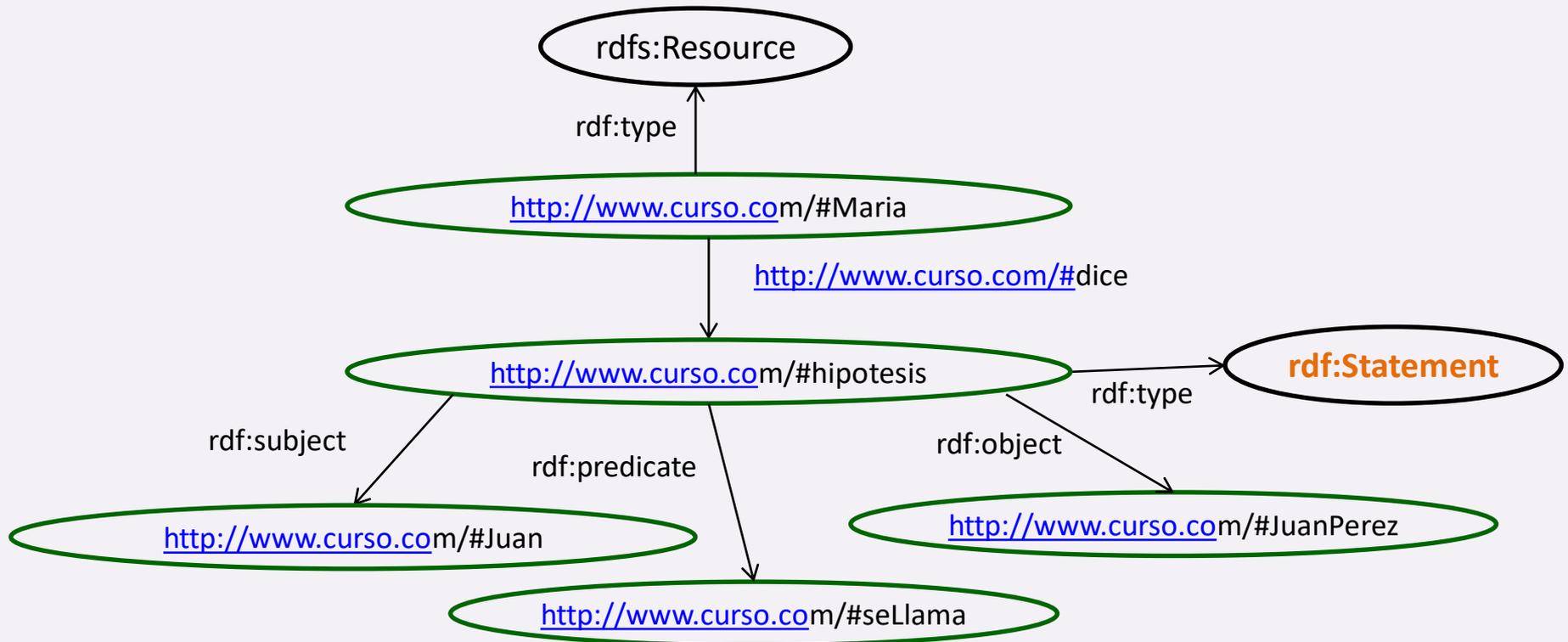
Reificación: sentencias sobre sentencias

“María dice que Juan se llama Juan Pérez”



Reificación: sentencias sobre sentencias

“María dice que Juan se llama Juan Pérez”



RDF Reificación: sentencias sobre sentencias

·
“María dice que Juan se llama Juan Pérez”

(pr:hipotesis, rdf:type, **rdf:Statement**)

(pr:hipotesis, rdf:subject, pr:Juan)

(pr:hipotesis, rdf:predicate, pr:seLlama)

(pr:hipotesis, rdf:object, pr:JuanPerez)

(pr:María, pr:dice, pr:hipotesis)

hipotesis: “Juan se llama Juan Pérez”

(pr:Juan, rdf:seLlama, ex:JuanPerez)

Vocabulario RDF

Definido en el namespace:

<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

Clases

rdf:Property
rdf:Statement
rdf:XMLLiteral
rdf:Seq
rdf:Alt
rdf:Bag
rdf>List

Propiedades

rdf:type
rdf:subject
rdf:predicate
rdf:object
rdf:first
rdf:rest
rdf:_n
rdf:value

Recursos

rdf:nil

Typing con `rdf:type`

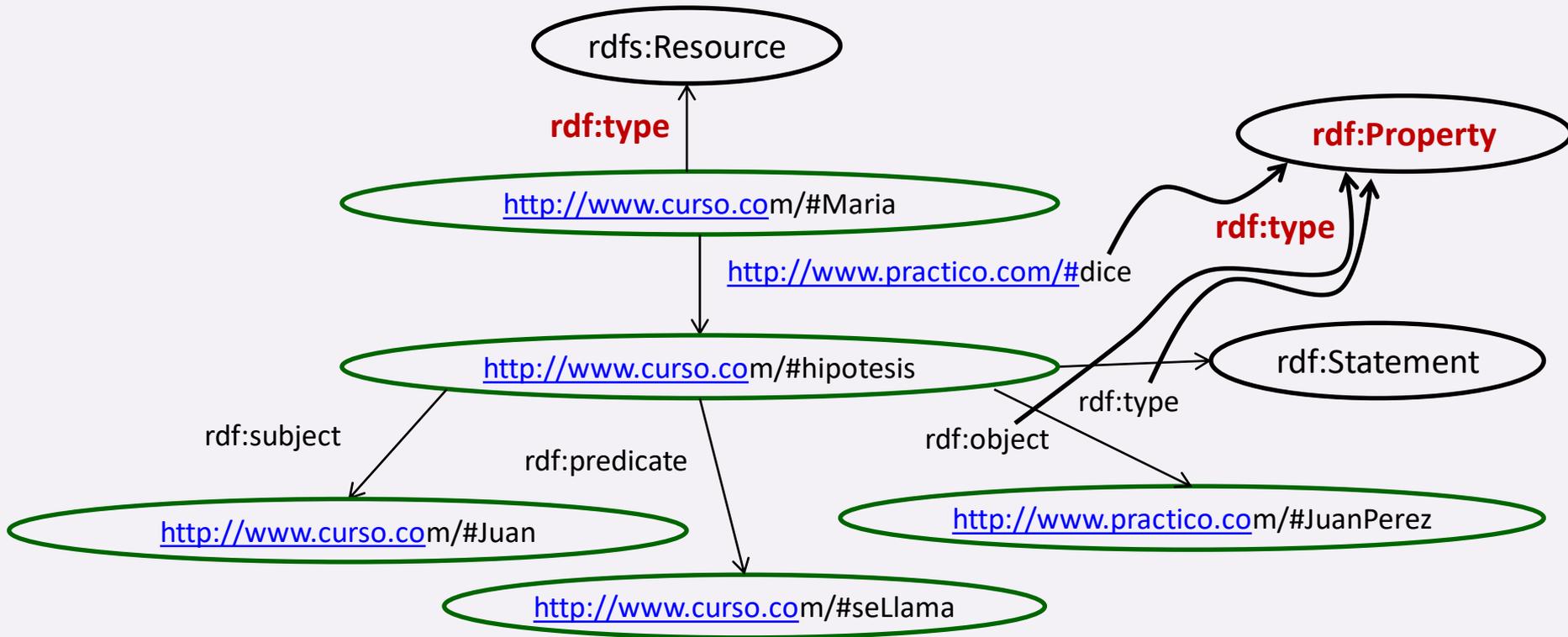
`(A, rdf:type, B)` "A pertenece a la clase B"

Todas la propiedades pertenecen a la clase `rdf:Property`:

`(P, rdf:type, rdf:Property)` "P es una propiedad"

`(rdf:type, rdf:type, rdf:Property)` "rdf:type es una propiedad"

Gráficamente Typing con rdf:type



RDF Containers

Agrupar valores

Bag (conjunto)

"The lecture is attended by John, Mary and Chris"

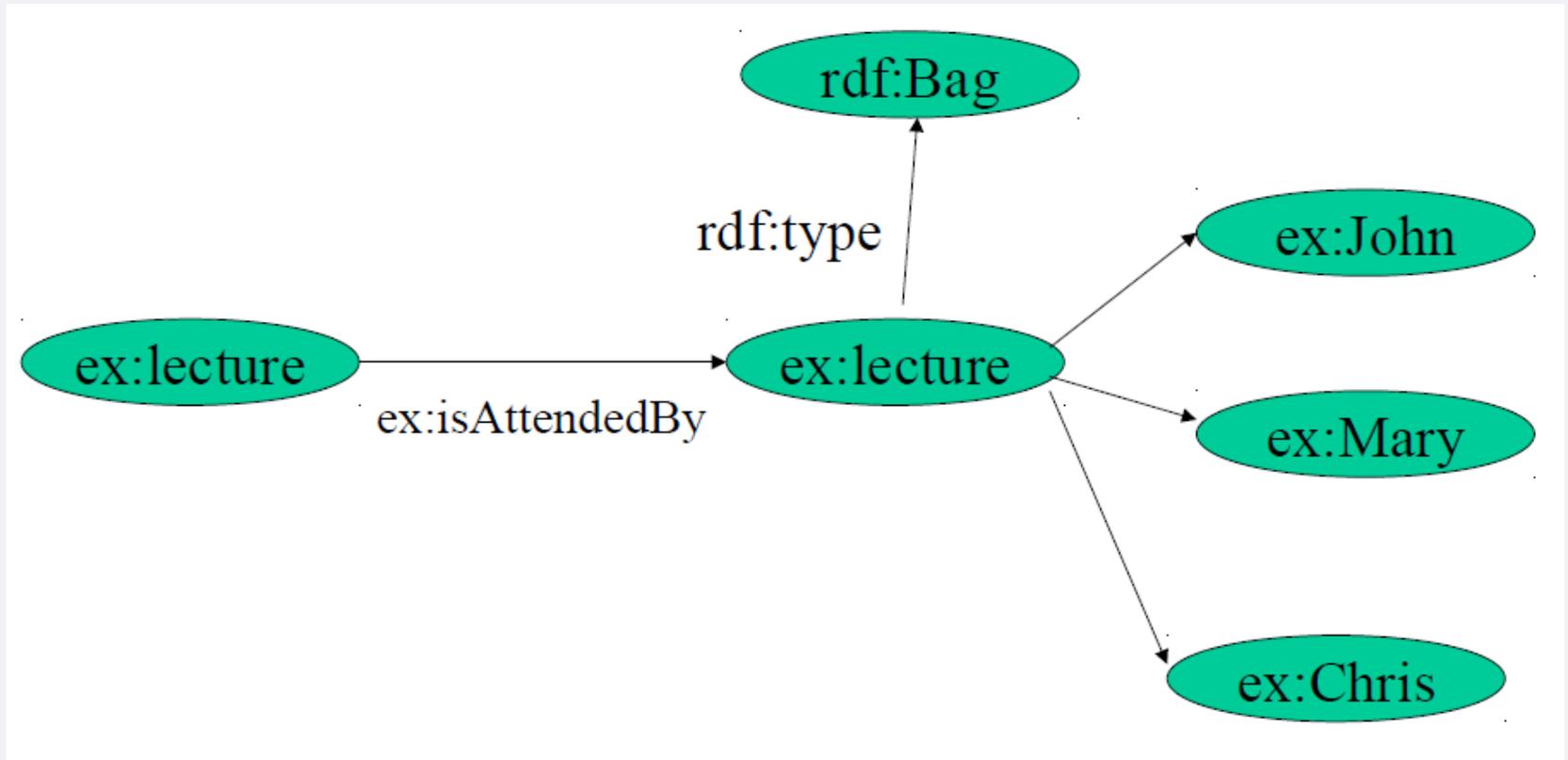
Seq (secuencia)

"[RDF-Concepts] is edited by Graham and Jeremy (in that order)"

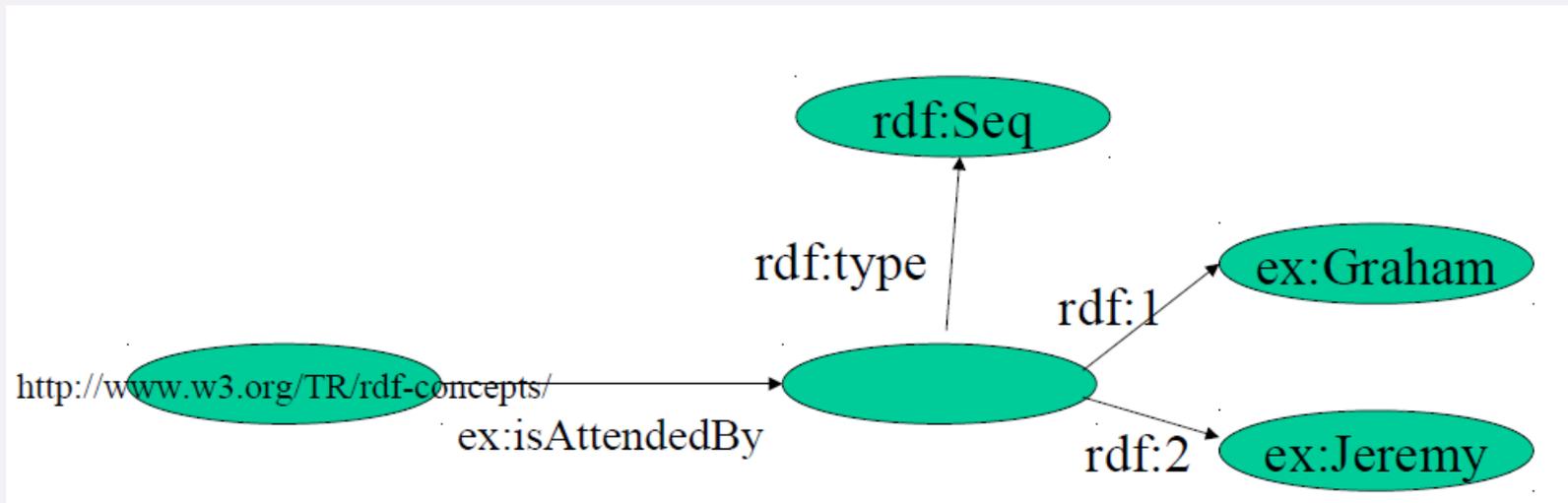
Alt (alternativas)

"The source code for the application may be found at ftp1.example.org, ftp2.example.org, ftp3.example.org"

Ejemplo Bag RDF Containers



Ejemplo Seq RDF Containers



RDF: modelo para integrar recursos

Ejemplo1

Ejemplo: Base de datos de libros

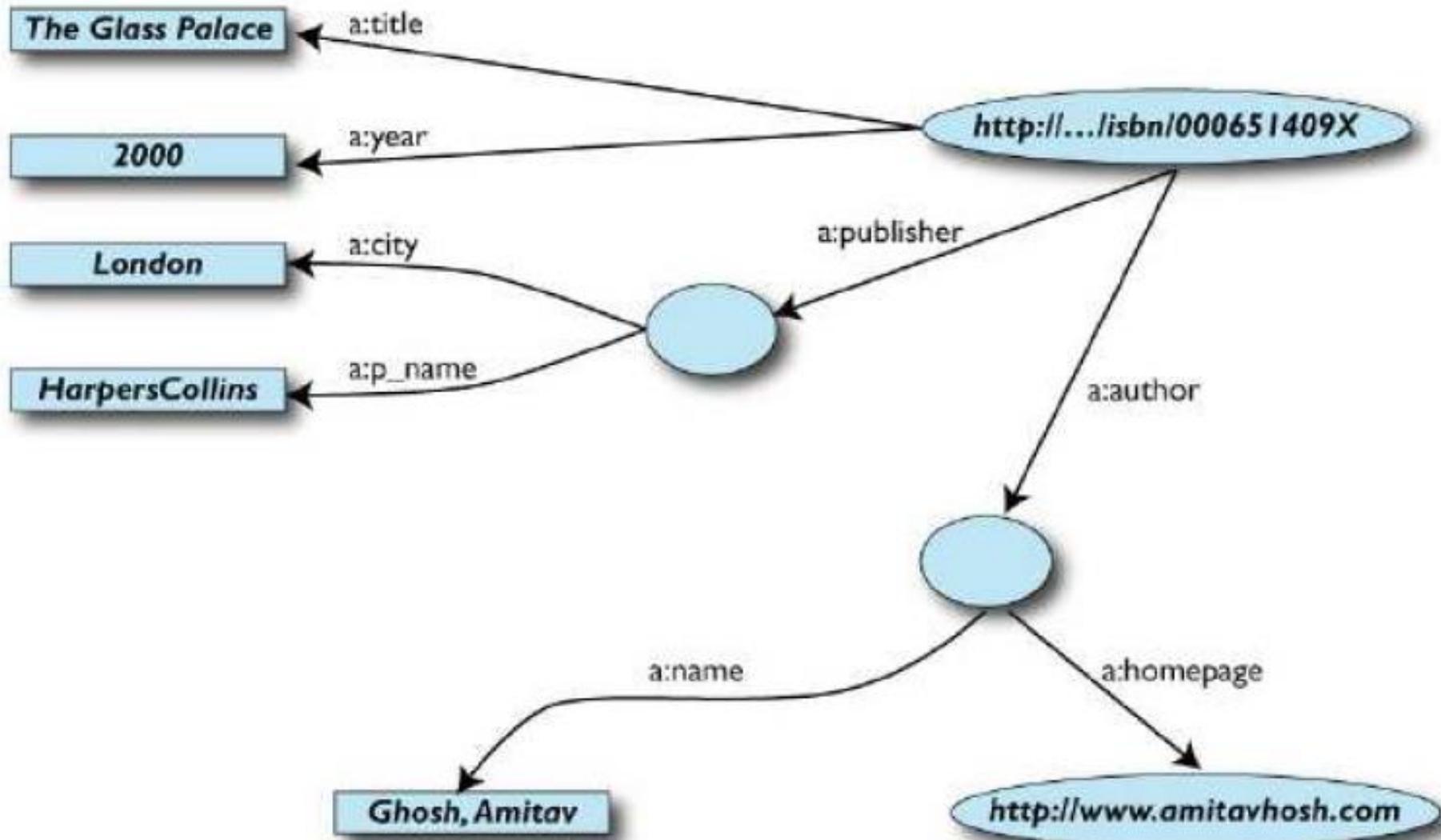
ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Home Page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publ. Name	City
id_qpr	Harpers Collins	London

RDF: Resource Description Framework

RDF del Ejemplo1



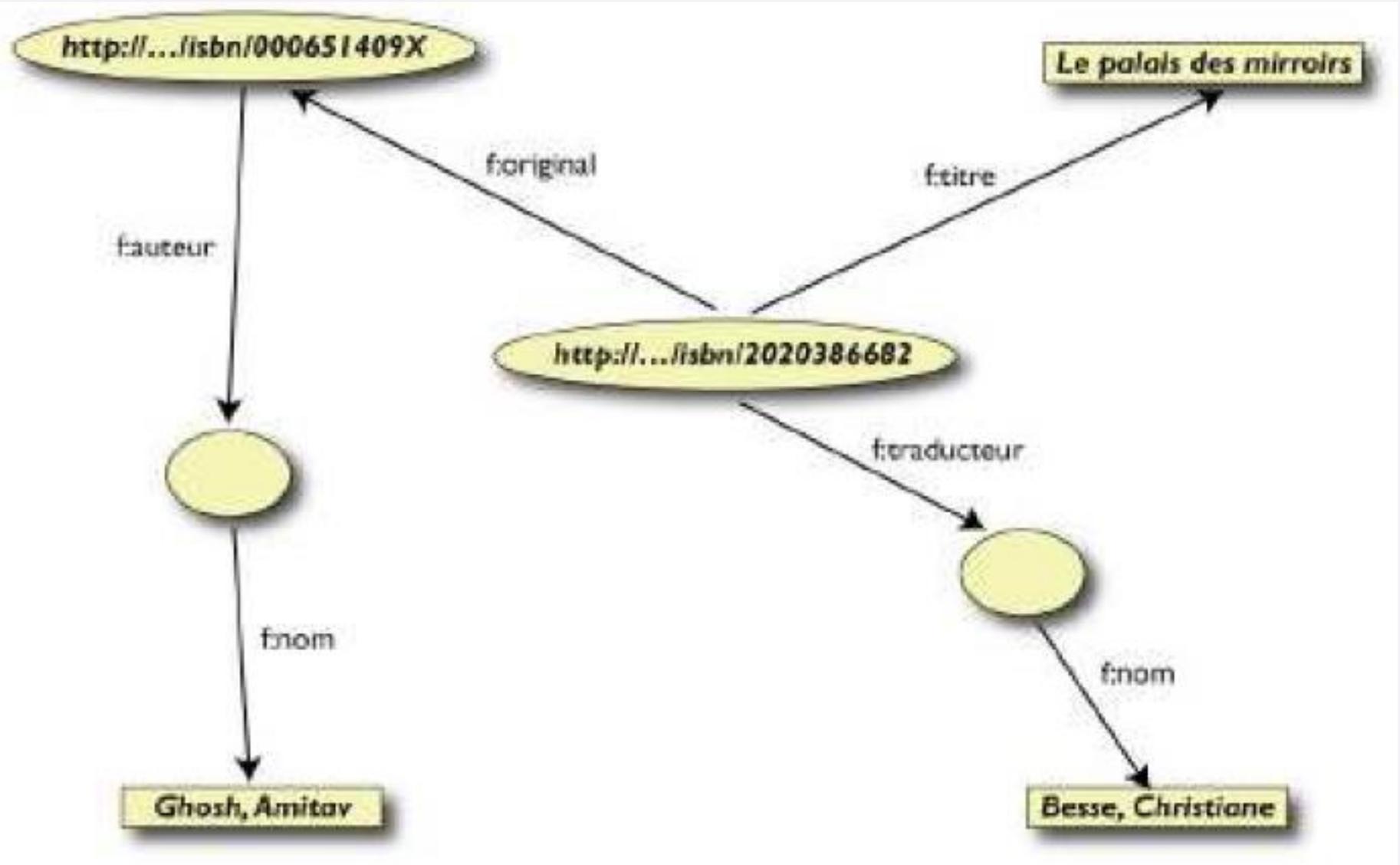
Ejemplo2

Otra base de datos de libros

ID	Titre	Auteur	Traducteur	Original
ISBN0 2020386682	Le Palais des miroirs	i_abc	id_qrs	ISBN-0-00-651409-X

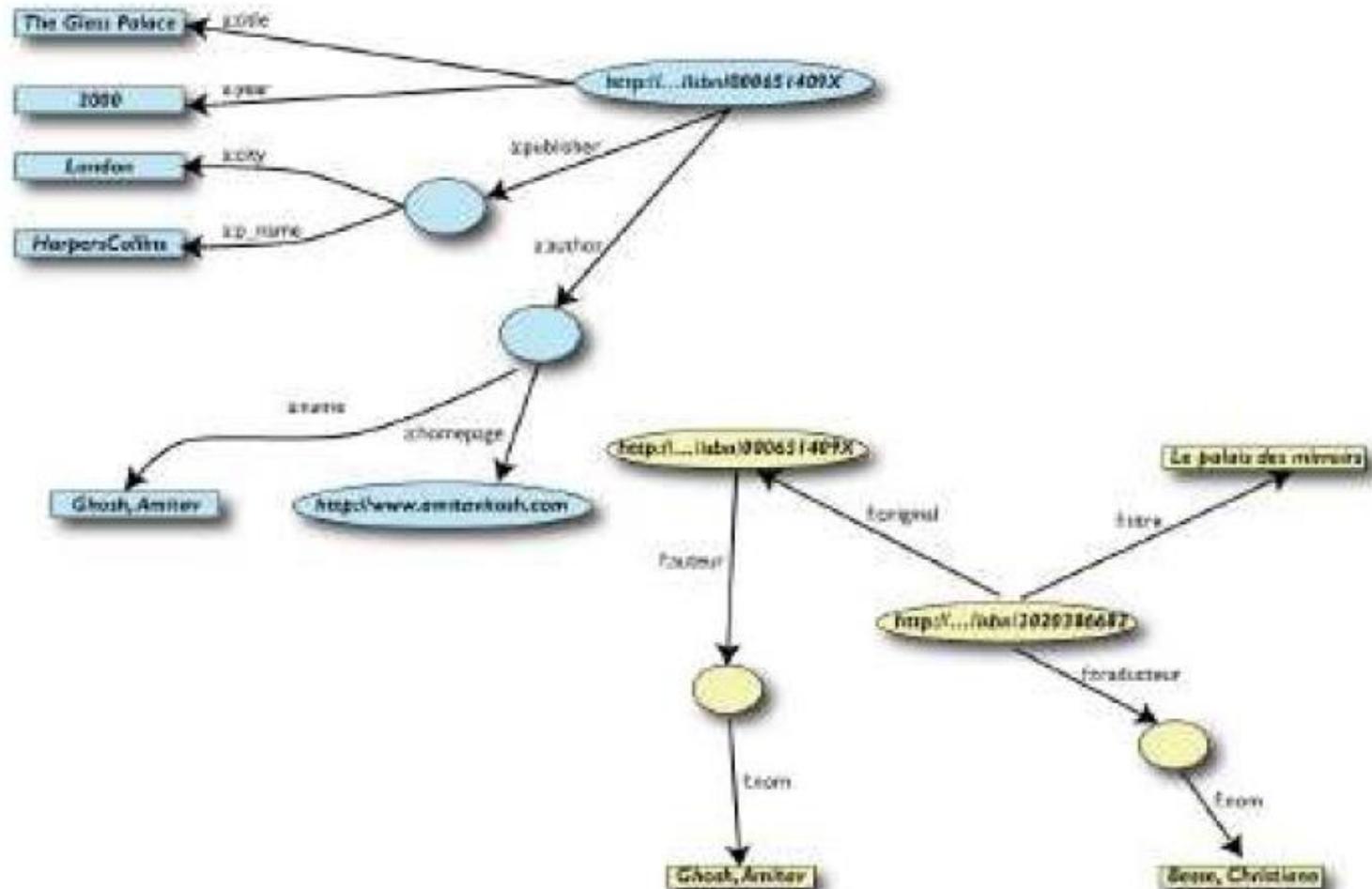
ID	Nom
id_abc	Ghosh, Amitav
id_qrs	Besse, Christiane

RDF del Ejemplo2



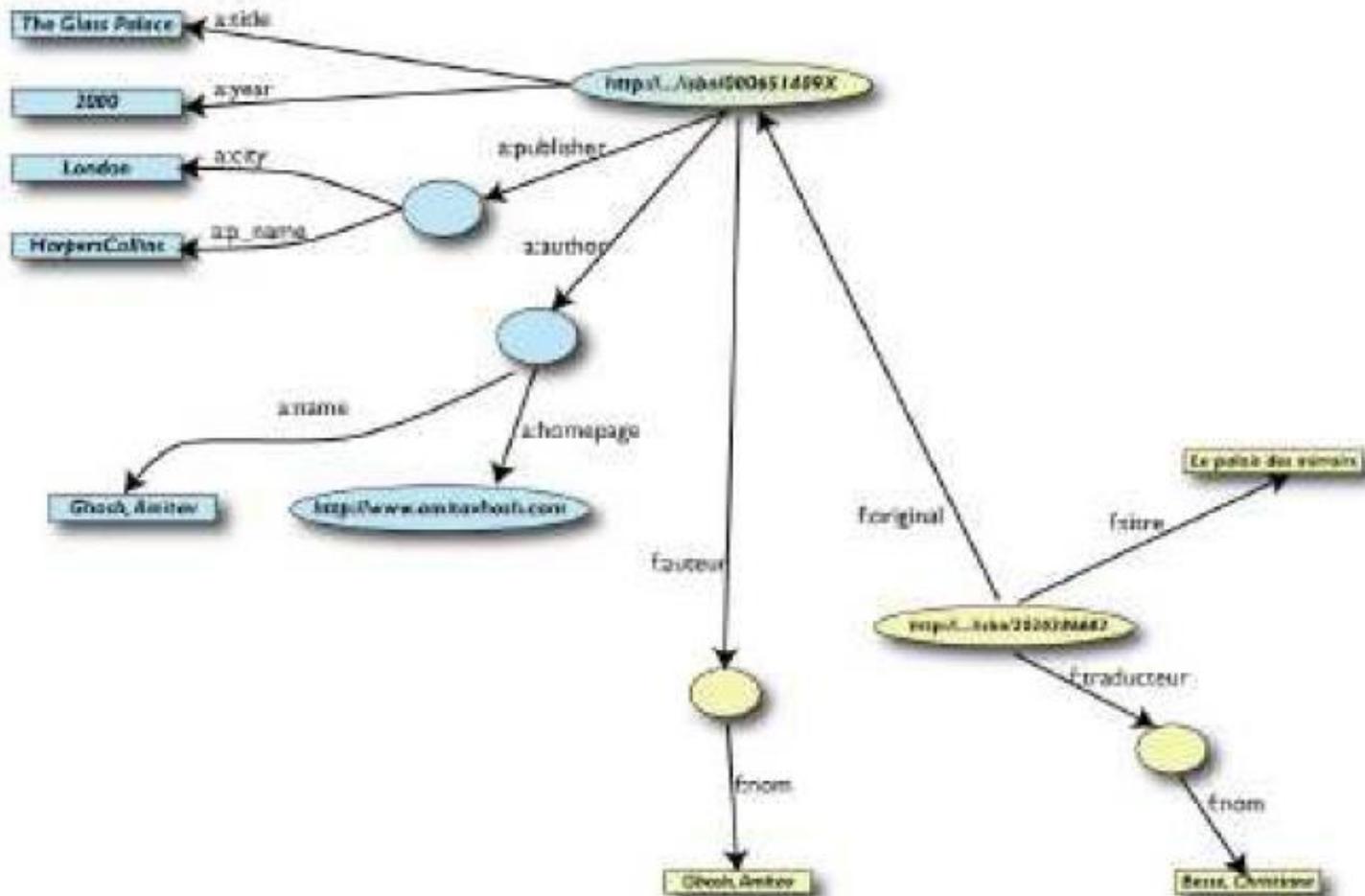
RDF Ejemplo 1 y Ejemplo2

Integrar

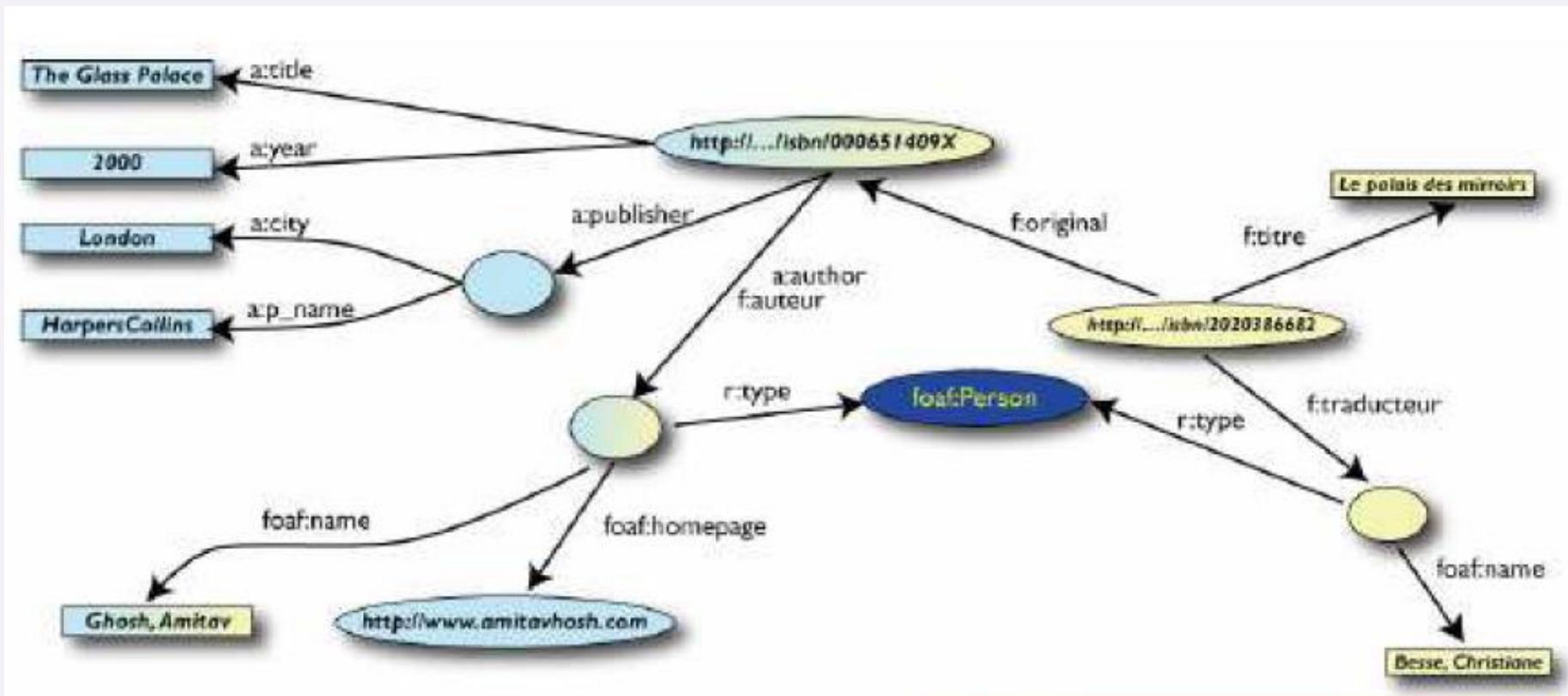


RDF Ejemplo 1 y Ejemplo2

Integrando...



RDF Ejemplo 1 y Ejemplo2 Integrados



SINTAXIS DE RDF

- RDF puede ser escrito usando varias sintaxis:
- Sintaxis XML: RDF/XML (RDF Serializado)
- Sintaxis N3
- Sintaxis Turtle
- RDFa
- JSON

Sintaxis de RDF en XML (serialización)

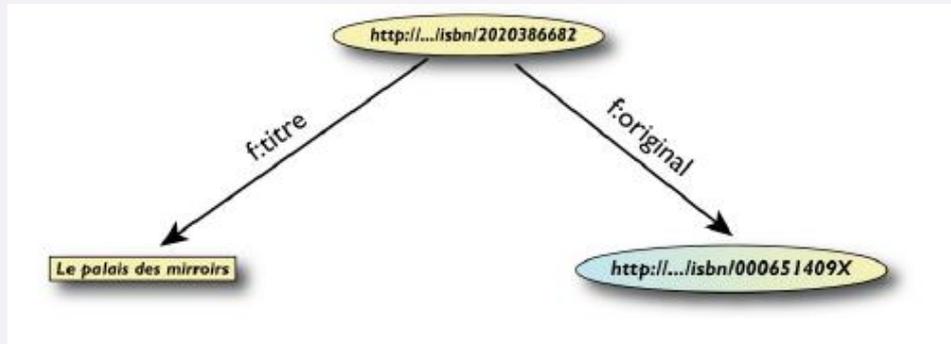
Cada elemento **Description** describe un recurso.

Los recursos son referenciados usando **URIs**

Atributos o elementos dentro de una **Description**: representan las **properties** de ese recurso

Otras sintaxis de RDF: N3, Triplas, ...

Ejemplo de RDF/XML (RDF Serialización)



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:f="http://www.editeur.fr">
  <rdf:Description rdf:about="http://.../isbn/2020386682">
    <f:original>
      <rdf:Description rdf:about="http://.../isbn/000651409X"/>
    </f:original>
  </rdf:Description>
</rdf:RDF>
```

Simplificaciones de RDF/XML

```
<rdf:Description rdf:about="http://.../isbn/2020386682">  
  <f:original rdf:resource="http://.../isbn/000651409X"/>  
  <f:titre>  
    Le palais des miroirs  
  </f:titre>  
</rdf:Description>
```

Otro Ejemplo Sintaxis de RDF en XML (serialización)

```
<?xml versión="1.0"?>
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:curso="http://eva.fing.edu.uy/course#">
  <rdf:Description
    rdf:about="http://eva.fing.edu.uy/course#encuentro4">
    <rdf:type rdf:resource="http://eva.fing.edu.uy/course#Encuentro"/>
  </rdf:Description>
</rdf:RDF>
```

Otra forma:

```
<?xml versión="1.0"?>
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:curso="http://eva.fing.edu.uy/course#">
  <curso:Encuentro
    rdf:about="http://eva.fing.edu.uy/course#encuentro4">
  </curso:Encuentro>
</rdf:RDF>
```

RDFa

RDF embebido en HTML

RDFa: Permite embeber sentencias RDF en documentos HTML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml
  xmlns:foaf=http://xmlns.com/foaf/0.1/
  xmlns:dc=http://purl.org/dc/elements/1.1/
  version="XHTML+RDFa 1.0" xml:lang="en">
  <head>
    <title>John's Home Page</title>
    <base href="http://example.org/john-d/" />
    <meta property="dc:creator" content="Jonathan Doe" />
    <link rel="foaf:primaryTopic" href="http://example.org/john-d/#me" />
  </head>
  <body about="http://example.org/john-d/#me">
    <h1>John's Home Page</h1>
    <p>My name is <span property="foaf:nick">John D</span> and I like
      <a href="http://www.neubauten.org/" rel="foaf:interest"
        xml:lang="de">Einstürzende Neubauten</a>
    </p>
    .....
```

RDF SCHEMA: RDF(S)

- Tener relaciones con semântica consensuada em RDF : “RDF Vocabulary Description Language”
- Esto es lo que se llama **RDF SCHEMA RDF(S)**!

Cuidado: No tiene nada que ver con la idea de SCHEMA de BASES DE DATOS!!

RDF Schema (RDFS)

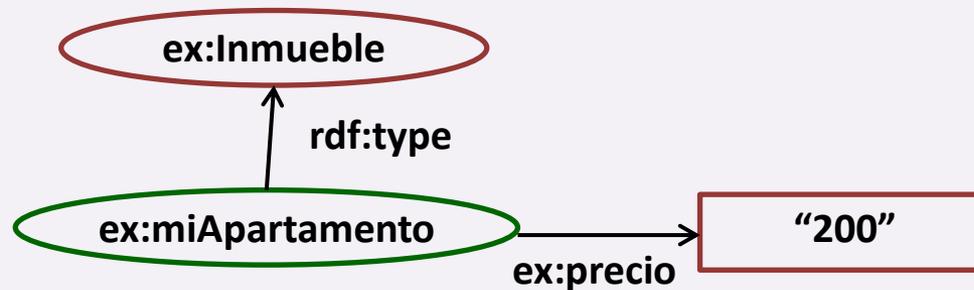
- RDF es un formalismo para metadatos escrito en XML pero no provee ningun significado especial a sus relaciones, ejemplos: `subClassOf` o `type`
- RDF Schema define “schema vocabulary” que ofrece un “significado especial” a algunas predicados y recursos. Ejemplo: `subClassOf`
 - Este “significado especial” , o semántica, especifica como un término debe ser interpretado.

RDF Schema (RDFS)

RDFS es una **recomendación de W3C** y es un lenguaje de representación de conocimiento extensible que puede ser usado para crear **vocabularios para describir clases, subclases y propiedades de recursos RDF.**

Los constructores RDFS son también clases y propiedades

RDF Schema (RDFS)



Con RDF puedo expresar que `ex:miApartamento` es un `ex:Inmueble`

¿Dónde está definido el recurso `ex:Inmueble`?

Si `ex:Inmueble` es una clase, **¿cuáles son sus superclases y subclases?**

Las comunidades de usuarios RDF necesitan definir **vocabularios** para describir **tipos o clases of recursos** específicos, que usan **propiedades** también específicas para describir estos recursos

RDFS Ejemplos

- RDF Schema términos:
 - Class
 - Property
 - type
 - subClassOf
 - range
 - domain
- Estos términos son usados (constructores) para crear vocabularios:

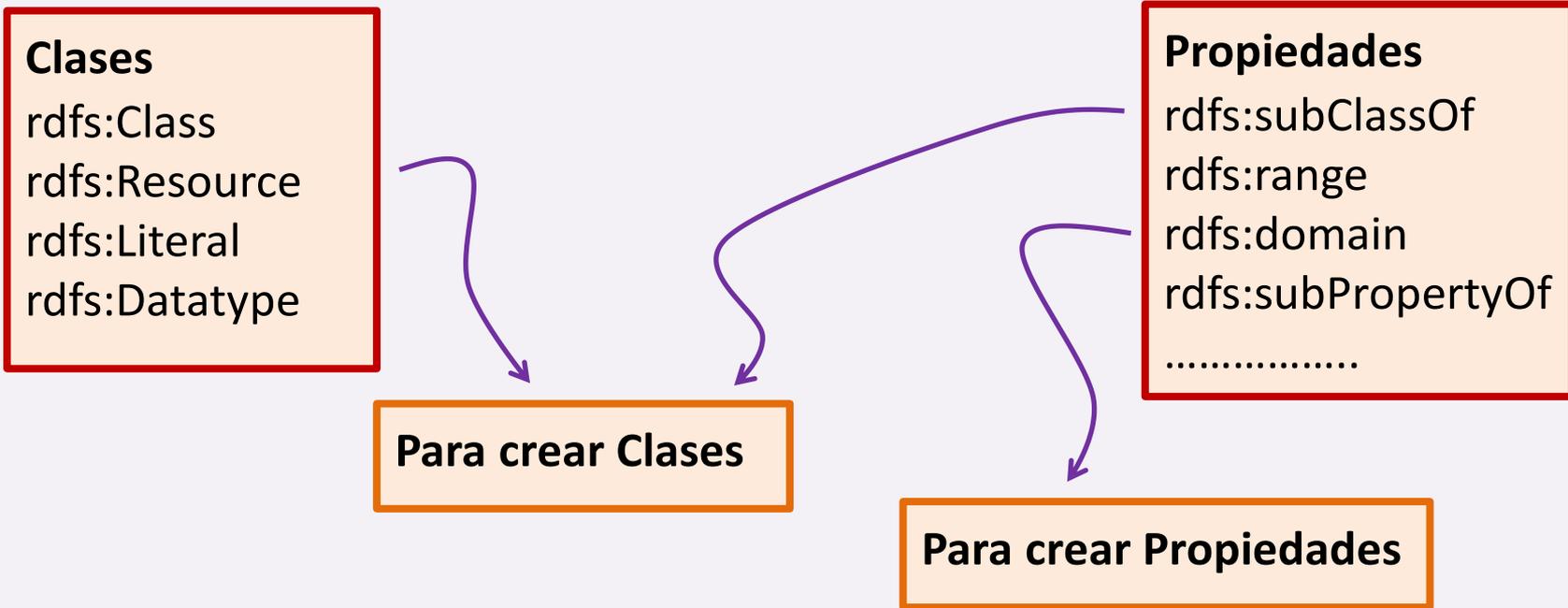
RDF Schema (RDFS)

Vocabulario RDFS

Definido en el namespace:

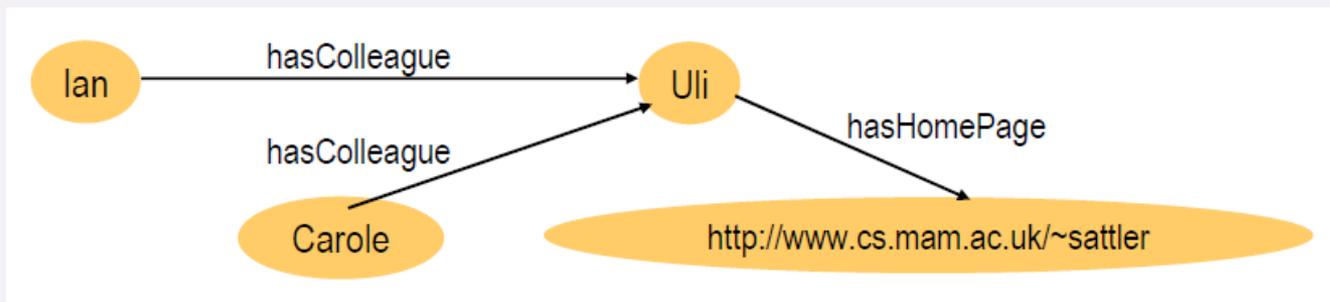
<http://www.w3c.org/2000/01/rdf-schema#>

Para definir vocabularios específicos (clases y propiedades) RDFS extiende el vocabulario definido en RDF.



RDF Schema (RDFS)

Vocabulario RDFS



<Person, type, Class>

<hasColleague, type, Property>

<Professor, subclassOf, Person>

<Carole, type, Professor>

<hasColleague, range, Person>

<hasColleague, domain, Person>

Clases

rdfs:Class
rdfs:Resource
rdfs:Literal
rdfs:Datatype

Propiedades

rdfs:subclassOf
rdfs:range
rdfs:domain
rdfs:subPropertyOf
.....

RDF Schema

<Person,type,Class>

<hasColleague,type,Property>

<Professor,subClassOf,Person>

<Carole,type,Professor>

<hasColleague,range,Person>

<hasColleague,domain,Person>

RDF/RDF(S) “Flexibilidad”

- **No distinción entre clases e instancias**
 - <Lion,**type**,Species>
 - <Leo,**type**,Lion>
- **Propiedades pueden tener propiedades**
 - <hasDaughter,**subPropertyOf**,hasChild>
 - <hasDaughter,**type**,familyProperty>
- **Constructores pueden ser aplicados sobre constructores**
 - <**type**,**range**,**Class**>
 - <**Property**,**type**,**Class**>
 - <**type**,**subPropertyOf**,**subClassOf**>

Material Referencia para RDF(S)

- [RDF Vocabulary Description Language 1.0: *RDF Schema*](#)

Trabajando con RDF(S)

- Por ejemplo usar **JAVA + JENA** (HP Bristol Lab)
 - Se crea un Modelo de objetos
 - El archivo RDF(S) es parseado y el resultado se almacena en el modelo de objetos.
 - El Modelo ofrece métodos para recuperar:
 - Triplas
 - Datos de un Sujeto: (Propiedad, objeto)
 - Datos de un Objeto: (Sujeto, Propiedad)
 - Etc..
- Se sigue con programación convencional

(Hay herramientas similares en Python, PHP..)

Un RDF dataset: DBpedia

Wikipedia → DBpedia

- Wikipedia
 - Conjunto de páginas Web creadas por una comunidad distribuída.
 - Las páginas son manualmente anotadas (semántica) usando un lenguaje de marcas.
- Algunos problemas:
 - Anotación manual a gran escala
 - Usuario que agrega la página debe además anotarla...
- **DBpedia**: un **gran RDF dataset** que contiene la información estructurada de Wikipedia.
 - Se crea una colección de grafos RDF
 - Permite consultar la información contenida en Wikipedia navegando en los grafos y accediendo a otros RDF datasets en la Web.
 - Metadatos de la Wikipedia: **procesable por máquina**
 - **Semántica precisamente definida**: conjunto de ontologías que permiten encontrar más fácilmente lo que se busca.

Un RDF dataset: DBPEDIA

Ejemplo:

Página de Roger Federer:

http://en.wikipedia.org/wiki/Roger_Federer

Página correspondiente enDBpedia:

http://dbpedia.org/resource/Roger_Federer

Al final de esta página:

Raw Data in: [CSV](#) | RDF ([N-Triples](#) [N3/Turtle](#) [JSON](#) [XML](#))

http://dbpedia.org/data/Roger_Federer.rdf



```
.....  
<rdf:Description rdf:about="http://dbpedia.org/resource/Roger_Federer">  
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person" />  
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Agent" />  
  <rdf:type rdf:resource="http://dbpedia.org/class/yago/SwissTennisPlayers" />  
  <rdf:type rdf:resource="http://dbpedia.org/class/yago/WorldNo.1TennisPlayers" />  
.....
```