

---

# Modelos RDF

---

PID\_00272097

Blas Torregrosa García

---

Tiempo mínimo de dedicación recomendado: 1 hora



**Blas Torregrosa García**

Ingeniero en Informática y máster universitario en Seguridad de las Tecnologías de la Información y de las Comunicaciones (MISTIC) por la Universitat Oberta de Catalunya (UOC). Especializado en ciberseguridad. Profesor colaborador en el máster de Ciencia de Datos de la UOC y profesor asociado en la Universidad de Valladolid (UVA).

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por el profesor: Ferran Prados Carrasco (2020)

Primera edición: febrero 2020  
© Blas Torregrosa García  
Todos los derechos reservados  
© de esta edición, FUOC, 2020  
Av. Tibidabo, 39-43, 08035 Barcelona  
Realización editorial: FUOC

*Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea este eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares de los derechos.*

# Índice

<b>Introducción</b> .....	5
<b>1. Modelo de datos RDF</b> .....	7
1.1. Contenedores .....	8
1.2. Colecciones .....	9
<b>2. Representaciones RDF</b> .....	10
2.1. Notación N3 .....	10
2.2. RDF/XML .....	10
2.3. Turtle .....	11
2.4. RDFa .....	12
2.5. JSON-LD .....	13
<b>Bibliografía</b> .....	15



## Introducción

Este módulo trata del modelo de datos RDF (*Resource Description Framework*) como piedra angular de toda la web semántica y como un estándar flexible para la estructuración de la información. Veremos en qué consiste el modelo y las diferentes representaciones (serializaciones) que admite.

«El Marco de Descripción de Recursos (RDF o, en inglés, *Resource Description Framework*) es un lenguaje diseñado para respaldar la web semántica, de la misma manera que HTML ayudó a iniciar la web original. RDF es un marco para la descripción de recursos o metadatos para la web. RDF proporciona estructuras comunes que se pueden usar para el intercambio de datos interoperable con XML.»

Declaración de Actividad de web semántica del W3C (World Wide Web Consortium).

RDF se creó en 1997, y entonces XML era un formato que comenzaba a ser reconocido y que tuvo una gran influencia en la sintaxis de RDF. El formato RDF/XML resulta muy prolijo y aparecieron varias modificaciones para hacer RDF más legible.

En 2014 aparece RDF 1.1 como una versión revisada que sigue manteniendo la mayor parte del modelo de datos y ha añadido otros formatos de representación como Turtle o JSON-LD. En este módulo veremos un breve resumen del modelo de datos RDF siguiendo la versión RDF 1.1.

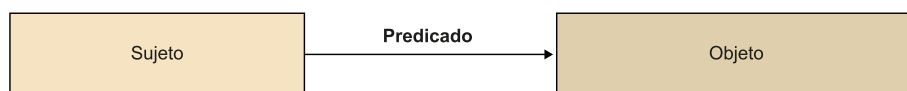


## 1. Modelo de datos RDF

El **modelo de datos RDF** se basa en definir sentencias (*statements*) sobre recursos utilizando tripletas. Cada tripeleta consta de un sujeto, un predicado y un objeto. Las **tripletas RDF** se representan como un grafo dirigido que conecta dos nodos (sujeto y objeto) mediante un arco (predicado).

El conocimiento en RDF se expresa mediante una lista de sentencias que siguen el mismo esquema (tripletas RDF).

Figura 1. Ejemplo de tripeleta RDF



Por tanto una tripeleta RDF expresa una relación binaria entre dos recursos: el **sujeto** y el **objeto**; el **predicado** representa la naturaleza de esta relación formulada de manera direccional (del sujeto al objeto) y se denomina *propiedad*. Puede haber tres tipos de nodos:

1) Un **URI/IRI** que denota un recurso. Un URI puede ser un sujeto, un objeto o un predicado. La mayoría de los formatos RDF permite definir prefijos (como un alias) para simplificar la escritura de URI extensos. Existen algunos alias frecuentemente usados como `rdf`, `xsd`, `rdfs`, `owl`, etc.

2) **Literales** denotan recursos que tienen asociado un valor como un entero o una cadena. Los literales solo pueden aparecer en el objeto de las tripletas. Todos los literales RDF tienen asociado un tipo de datos. Por ejemplo, `"33"^^xsd:integer` representa al entero 33 y `"1969-07-21"^^xsd:date` representa el 21 de julio de 1969. En caso de que no tengan declarado el tipo, se supone por defecto que es una cadena `xsd:string`.

Un tipo especial de literales son las cadenas con marcas de idioma que indican el idioma en el que está escrita esa cadena. Por ejemplo, `"hola"@es` representa literal con valor «hola» escrito en español (es).

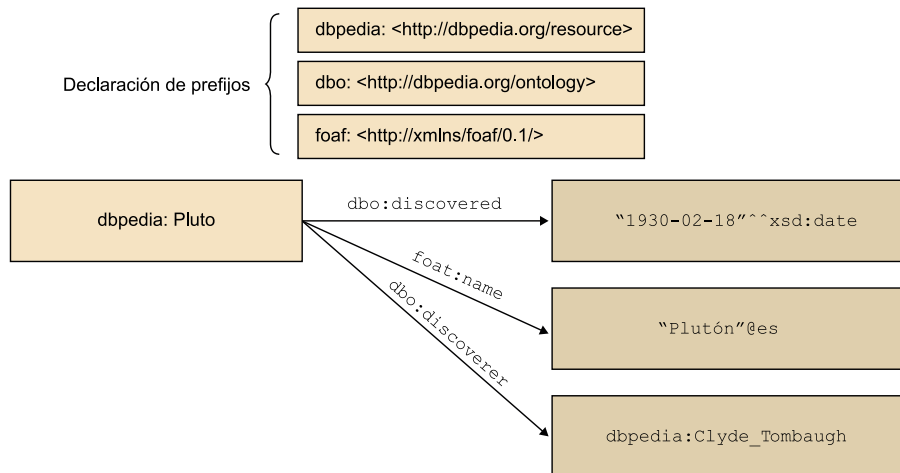
3) Nodos en **blanco** son identificadores locales que no referencian recursos. Los nodos en blanco se pueden usar como sujetos u objetos en las tripletas, especificando que existe la relación dada, sin nombrarla explícitamente. Los nodos en blanco se pueden usar para hacer afirmaciones sobre elementos cuyos URI no son conocidos.

### Enlace de interés

El servicio `prefix.cc` puede usarse para buscar URI asociados con alias conocidos.

Un **grafo RDF** es un conjunto de tripletas RDF. Los sujetos solo pueden ser URI o nodos en blanco, mientras que los objetos pueden ser URI, nodos en blanco o literales.

Figura 2. Ejemplo de tripletas que definen el hecho del descubrimiento de Plutón



Una característica importante de los grafos RDF es que dos grafos independientes se pueden fusionar para obtener otro grafo formado por la unión de sus conjuntos de tripletas. Dada la naturaleza global de los URI, los nodos con el mismo URI se pueden unir automáticamente.

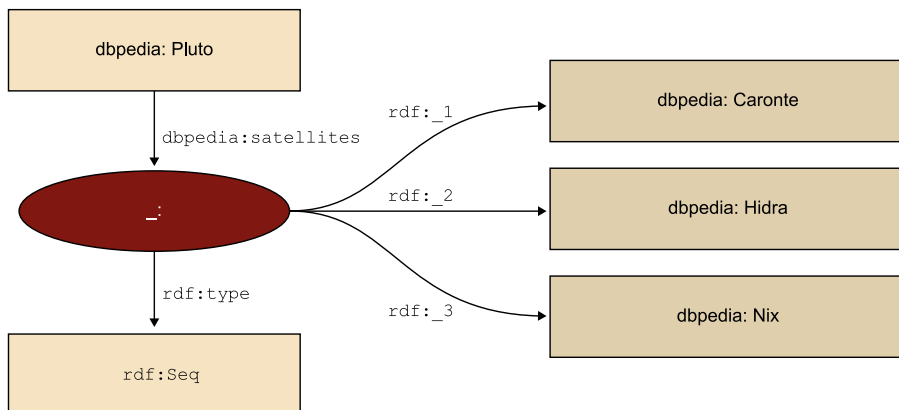
### 1.1. Contenedores

Los **contenedores** (*containers*) y las **colecciones** (*collections*) se utilizan para describir conjuntos de recursos. En RDF hay tres tipos de contenedores:

- `rdf:Bag`. Contenedor en el que se permiten duplicados y el orden de los recursos no es significativo.
- `rdf:Seq`. Contenedor secuencia en el que se permiten duplicados y el orden sí es significativo.
- `rdf:Alt`. Contenedor que guarda una lista de recursos o literales alternativos (se seleccionará uno de los miembros, por defecto el que tenga la propiedad `rdf:_1`).



Figura 3. Ejemplo de contenedores



## 1.2. Colecciones

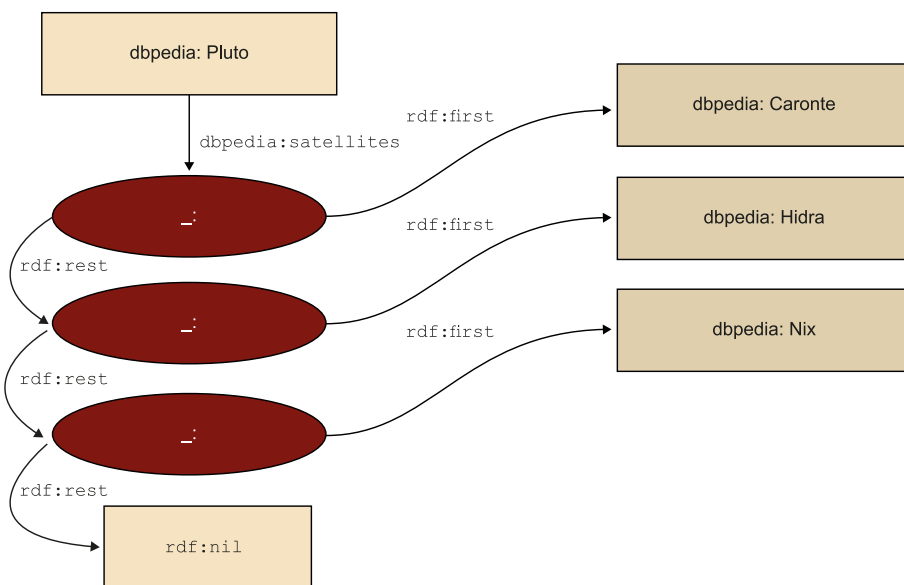
Una característica de los contenedores es que son listas abiertas, es decir, no hay ninguna limitación en el número de miembros que pueden contener. Para conseguir listas cerradas RDF dispone de las **colecciones** cuyo tipo básico es `rdf:List` y define una lista ordenada de recursos.

Se utilizan unas propiedades especiales para estructurar este tipo de listas:

- `rdf:first` para el primer elemento de la lista.
- `rdf:rest` para el resto de la lista.
- `rdf:nil` para el final de la lista.

La lista se estructura encadenando el conjunto de recursos (pueden usarse nodos en blanco) en el que cada nodo tiene una relación con un recurso y otra con el resto.

Figura 4. Ejemplo de colección



## 2. Representaciones RDF

RDF es un modelo de datos abstracto y no un formato en sí mismo, de manera que no está vinculado a una sintaxis concreta. Existen diversas representaciones sintácticas o serializaciones para el modelo RDF, algunas de ellas más adecuadas para el procesamiento por ordenador y otras más comprensibles para las personas.

### Serialización

El concepto **serialización** se refiere a la representación de datos en un formato para almacenarlos, transportarlos o cualquier otro motivo.

### 2.1. Notación N3

**Notation3 (N3)** es una forma de representar RDF con la finalidad de facilitar la lectura por personas y la compactación.

La notación N3 tiene las siguientes características:

- Los URI/IRI van entre ángulos (< >).
- Los literales van entre comillas (" ").
- Cada tripleta termina con un punto (.).

Figura 5. RDF en N3

```
@prefix dbpedia: <http://dbpedia.org/resource/>.
@prefix dbo: <http://dbpedia.org/ontology/>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

dbpedia:Pluto dbo:discovered "1930-02-18".
dbpedia:Pluto foaf:name "Plutón".
dbpedia:Pluto dbo:discovered dbpedia:Clyde_Tombaugh
```

### 2.2. RDF/XML

La primera notación utilizada fue **RDF/XML**, estandarizada por el W3C. Su principal ventaja radicaba en que los lenguajes de programación disponen de librerías para tratar con XML. Además se pueden utilizar los espacios de nombres (*Namespaces*) de XML, evitando URI completas y haciéndolo menos extenso. Sin embargo RDF/XML resulta poco legible por personas.

Como todo documento XML debe tener una única raíz y hay que indicar que es una raíz RDF junto con el espacio de nombres de W3C llamado *rdf*. Este espacio de nombres informa de que el documento es RDF y de que la etiqueta `rdf:RDF` forma parte de este.

Figura 6. RDF en XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

<!-- Cuerpo del documento -->

</rdf:RDF>
```

RDF/XML puede tener varios enunciados (tripletas), para cada uno de los cuales hay que añadir una etiqueta `rdf:Description`. El atributo `rdf:about` declara el sujeto del enunciado. Se pueden añadir varios enunciados sobre el mismo sujeto.

Las propiedades se añaden como nodos que cuelgan de `rdf:Description` y contendrán un identificador de la propiedad (etiqueta XML) así como el objeto del enunciado, que puede ser un literal o un URI/IRI. Los nodos en blanco se identifican con el atributo `rdf:nodeID`, relativo al documento en cuestión.

Figura 7. Ejemplo de RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dbo="http://dbpedia.org/ontology/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" >

  <rdf:Description rdf:about="http://dbpedia.org/resource/Pluto">
    <dbo:discovered>1930-02-18</dbo:discovered>
    <foaf:name xml:lang="es">Plutón</foaf:name>
    <dbo:discoverer rdf:resource="http://dbpedia.org/resource/Clyde_Tombaugh"/>
  </rdf:Description>
</rdf:RDF>
```

### 2.3. Turtle

**Turtle** (*Terse RDF Triple Language*) es una extensión de Notation3.

Las características de esta notación son las siguientes:

- El punto y coma (;) indica que las siguientes tripletas tienen el mismo sujeto.
- La coma (,) indica que las siguientes tripletas comparten el mismo sujeto y propiedad (lista de objetos).
- En los literales es posible indicar el tipo de datos (fecha, entero, punto flotante, etc.).
- Los nodos en blanco se indican con `_:`. Si hubiera varios pueden llevar un identificador interno para documento, es decir, `_:ID1`, `_:ID2`, etc.

Figura 8. Ejemplo de RDF en Turtle

```
@base <http://dbpedia.org/resource/>.
@prefix dbo: <http://dbpedia.org/ontology/>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.

<Pluto> dbo:discovered "1930-02-18"^^xsd:date;
        foaf:name "Plutón"@es;
        dbo:discovered <Clyde_Tombaugh>.
```

## 2.4. RDFa

**RDFa** (*Resource Description Framework in Attributes*) es un mecanismo que permite definir tripletas RDF dentro de documentos HTML de forma integrada con la página web.

El fundamento de RDFa es utilizar un conjunto de atributos para añadir información RDF a los lenguajes de marcado. Para ello se marca cada entidad con un tipo de elemento RDF. Este elemento tendrá el papel de sujeto. A continuación se definen las propiedades (equivalentes a los predicados) y los valores de las propiedades (equivalentes a los objetos).

Los atributos básicos para el marcado de datos RDF son:

- `vocab`: indica el vocabulario que se va a utilizar o de manera alternativa también se pueden indicar espacios de nombres.
- `typeof`: indica el tipo de sujeto que se declara.
- `about`: indica la URI del sujeto.
- `property`, `rel` o `rev`: indican las propiedades del sujeto, es decir, los predicados.
- `href`, `resource`, `src`: indican la URI de los recursos relacionados, es decir, los objetos. En el caso de los literales, estarán enmarcados por la etiqueta.

En el siguiente ejemplo se utiliza RDFa para extender una página web con contenido semántico. La página original es la siguiente:

Figura 9. Página web convencional

```

<div>
  <span>Plutón</span>
</div>
<div>
  Descubierta el <span> 18 de febrero de 1930 </span>.
</div>
<div>
  Descubierta por <a>Clyde Tombaugh</a>.
</div>
</div>

```

A continuación tenemos la página con marcado RDFa.

Figura 10. Ejemplo de página web con RDFa

```

<html version="XHTML+RDFa 1.1"
  xmlns="http://www.w3.org/1999/xhtml" xml:lang="es" lang="es" >
  <div xmlns:dbpedia="http://dbpedia.org/resource/"
    xmlns:dbo = "http://dbpedia.org/ontology/"
    xmlns:foaf= "http://xmlns.com/foaf/0.1/"
    about="dbpedia:Pluto">
    <span property="foaf:name">Plutón</span>
  </div>
  Descubierta el <span property="dbo:discovered"
    content="1930-02-18">18 de febrero de 1930</span>.
  </div>
  <div>
  Descubierta por <a rel="dbo:discoverer"
    href="dbpedia:Clyde_Tombaugh">Clyde Tombaugh</a>.
  </div>
</div>
</html>

```

## 2.5. JSON-LD

**JSON-LD** (*JavaScript Object Notation for Linked Data*) es un formato de serialización de tripletas RDF que tiene una representación más legible por personas.

Mediante JSON las tripletas se representan como conjuntos de pares atributo-valor.

Uno de los elementos principales de JSON-LD es la idea de **contexto**. Un contexto en JSON-LD permite que dos aplicaciones utilicen términos abreviados para comunicarse entre sí con mayor eficiencia y sin perder precisión.

Figura 11. Ejemplo de RDF en JSON-LD

```
{
"@context": { "dbpedia": "http://dbpedia.org/resource/",
              "dbo": "http://dbpedia.org/ontology/",
              "foaf": "http://xmlns.com/foaf/0.1/",
              "xsd": "http://www.w3.org/2001/XMLSchema#"
            },

"@id": "dbpedia:Pluto",

"dbo:discovered": { "@value": "1930-02-18" , "@type" : "xsd:date" },
"foaf:name": { "@value": "Plutón", "@language": "es" },
"dbo:discoverer": { "@id": "dbpedia:Clyde_Tombaugh" }
}
```

## Bibliografía

**DuCharme, R.** (2013). *Learning SPARQL* (2.<sup>a</sup> ed.). O'Reilly Media.

**Guarino, N.; Oberle, D.; Staab, S.** (2009). *What Is an Ontology?* [en línea]. [Fecha consulta: enero 2020]. Disponible en: <[https://iaoa.org/isc2012/docs/Guarino2009\\_What\\_is\\_an\\_Ontology.pdf](https://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf)>

**Kumar, A.** (2018). *Architecting Data-Intensive Applications*. Packt Pub.

**Noy, N. F.; McGuinness, D. L.** (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report (Informe SMI-2001-0880).

**Powers, S.** (2003). *Practical RDF*. O'Reilly Media.

