

# Metodologías para el diseño de ontologías Web



Elsa E. Barber, Silvia Pisano, Sandra Romagnoli, Gabriela de Pedro, Carolina Gregui, Nancy Blanco y María Rosa Mostaccio

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Investigaciones Bibliotecológicas. Ciudad Autónoma de Buenos Aires, Argentina

elsabarber.eb@gmail.com.ar / <https://orcid.org/0000-0003-2970-1356> | silvialuisapisano@gmail.com /

<https://orcid.org/0000-0003-1408-907X> | sandraer@fibertel.com.ar / <https://orcid.org/0000-0003-4479-3172> |

gabdp@yahoo.com / <https://orcid.org/0000-0001-9409-2644> | cgregui@hotmail.com / <https://orcid.org/0000-0003-2916-3543> |

nancybl@filo.uba.ar / <https://orcid.org/0000-0002-4218-6187> | mmostaccio@gmail.com / <https://orcid.org/0000-0002-9628-5914>

## Resumen

El artículo analiza las características de las ontologías Web, junto con su definición, su constitución y las tecnologías que se adoptan para su diseño e implementación en el marco de la Web Semántica. Se describe su relación con los Datos Abiertos Enlazados (LOD), con el modelo Resource Description Framework (RDF), el Web Ontology Language (OWL), el Simple Protocol and RDF Query Language (SPARQL) y los Uniform Resource Identifiers (URIs). También trata con aquellas cuestiones relacionadas con la nominación, lexicalización, localización y serialización. Se identifican las metodologías más utilizadas para desarrollar ontologías Web. Se describe la metodología propuesta por Stuart (2016). Sobre dicha base, se especifican las decisiones metodológicas adoptadas para el diseño de una ontología que permita representar las investigaciones académicas y científicas argentinas. Se destaca la importancia que tiene la resolución adecuada de la instancia metodológica en el proceso de creación de una ontología aplicada, a fin de optimizar su capacidad de inferencia automática.

## Palabras clave

Metodología  
Diseño  
Ontología  
Web

## Abstract

**Web ontologies design methodologies.** The article analyzes characteristics, definition, constitution, and technologies adopted for designing and implementing ontologies within the Semantic Web framework. Therefore, it describes relationships between Web ontologies and Linked Open Data (LOD), Resource Description Framework (RDF) model, Web Ontology Language (WQL), Simple Protocol and RDF Query Language (SPARQL) and Uniform Resources Identifiers (URIs). The study deals with naming, lexicalizing, localization and serializing. It identifies the most used methodologies for developing Web ontologies. The work describes the method proposed by Stuart (2016). According to this author, this article considers decisions about ontologies methods design representing Argentine scientific and academic research. It

## Keywords

Methodology  
Design  
Ontology  
Web

underlines the importance of methodological stage in the creation process of applied ontologies to optimizing their capacity of automatic inference.



Artículo recibido: 23-07-2018. Aceptado: 20-10-2018.

## Introducción

El término ontología, fue incorporado al campo de la Inteligencia Artificial (Artificial Intelligence, AI), según Gruber (1993a), con el fin de enunciar modelos computacionales capaces de soportar el razonamiento automático y de capturar conocimiento. Fue Gruber (1993b: 1-2) quien proporcionó una definición amplia y muy difundida del concepto:

An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what 'exists' is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a logical theory.

Con posterioridad, Guarino (1998), definió ontología como un artefacto de ingeniería y, en coincidencia con Gruber (1993b), como la representación formal de una teoría lógica:

An engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words. This set of assumptions has usually the form of a first-order logical theory, where vocabulary words appear as unary or binary predicate names, respectively called concepts and relations. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation. (Guarino, 1998: 4).

Smith, Kusnierczyk, Schober y Ceusters (2006: 61) ya la denominaron "...a representational artifact comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them..." Finalmente, para Stuart (2016:12), "...an ontology is a formal representation of knowledge with rich semantic relationships between terms..." Este autor, como los anteriores, considera que la característica distintiva de las ontologías reside en la riqueza de las relaciones que se establecen entre los términos que las constituyen. Destaca, además, la importancia de explicitar las especificaciones de los conceptos incluidos, de formalizarlas para que las computadoras puedan comunicarse entre sí con éxito y sin ambigüedad a fin de acrecentar la posibilidad de realizar

inferencias. Por otra parte, cuestiona la distinción entre metadatos y datos ya que resulta para él improductiva. En tal sentido, al producirse el desplazamiento desde los objetos del mundo real hacia los objetos digitales, TODOS son DATOS.

De acuerdo con Aguado-de-Cea, Montiel-Ponsoda, Poveda-Villalón, y Giraldo-Pasmin (2015: 152), la función de las ontologías para la Ciencia de la Información consiste en “... to capture knowledge about some domain of interest, by formally naming and defining the types, properties and attributes of the concepts that describe that domain. ...”

En la actualidad, los datos estructurados mediante la aplicación de las ontologías se publican bajo la tecnología de los datos enlazados (Linked Data, LD) que, según Berners-Lee (2006), respetan cuatro principios:

1. Usan URIs (Uniform Resource Identifiers) como nombres para las cosas
2. Usan HTTP URIs para que las personas puedan buscar dichos nombres
3. Proveen información útil al usar estándares (Resource Description Framework - RDF, SPARQL Protocol and RDF Query Language - SPARQL) cuando alguien busca un URI
4. Incluyen enlaces hacia otros URIs y así pueden descubrir otros objetos

Sin embargo, los datos enlazados pueden publicarse sin constituir necesariamente una ontología. El dinamismo y la exhaustividad son factores que permiten distinguir entre unos y otros (Stuart, 2016). Según Hedden (2010) una ontología es un tipo específico de taxonomía, por lo tanto, la formalización de los datos mediante ontologías constituye un soporte para la indización, para la recuperación y para la organización/navegación. Puede agregarse que constituye una tecnología soporte de la creación de conocimiento. Las ontologías se representan mediante grafos (vectores de bytes) constituidos por conceptos y las relaciones existentes entre ellos; pero el grafo que se obtiene como respuesta a una consulta no se construye solo con las relaciones explícitas, puede construirse mediante relaciones inferidas.

Swanson y Smalheiser (1996) rescatan la diferenciación que realiza Swanson (1986), en base a Popper, entre el Mundo 1, que corresponde al mundo físico; el Mundo 2 que abarca el conocimiento subjetivo, la experiencia, los estados y procesos mentales y el Mundo 3 que comprende el conocimiento objetivo, los problemas y las teorías que han sido comunicados. Solo el conocimiento subjetivo es abarcable por un individuo, los mundos 1 y 3, por el contrario, son inabarcables para un sujeto en particular. Desde esa perspectiva, Swanson (1986) tipifica las formas en las que el conocimiento objetivo aún no descubierto deja de contribuir a la creación de nuevo conocimiento, a saber:

- » una refutación oculta o poco visible no sirve para testear una hipótesis
- » un vínculo lógico que se ha perdido no sirve para inferir una causa (si una persona desconoce que A es causa de B y B es causa de C, entonces, no puede inferir que A causa C)
- » la acumulación oculta de test individualmente débiles impide fortalecer una hipótesis

Si bien las ontologías se orientan a superar estas carencias, tienen limitaciones debido a que manipulan los conceptos representados mediante el lenguaje expuesto a variaciones permanentes y compiten cada vez más con el procesamiento automático del lenguaje natural (*Natural Language Processing, NLP*) debido a su rigidez y al costo que demandan. Más allá de estas apreciaciones, la Web Semántica se funda en esta tecnología y en el uso de lenguajes legibles por computadora para su representación: *Resource Data Framework Schema (RDFS)*, *Simple Knowledge Organization System (SKOS)* o *Web Ontology Language (OWL)*.

En ese marco, el proceso de creación de una ontología implica múltiples tomas de decisiones, que desde el punto de vista metodológico resulta relevante sistematizar, ya que dicho proceso debe garantizar que han de respetarse aquellas cuestiones relacionadas con la manera en que las personas podrán ver y acceder a los datos. Las herramientas tecnológicas que se utilicen y la estructura de datos que se implemente reflejan, desde el punto de vista de las políticas de información que se deseen sostener, posturas diferenciadas con respecto a la visibilidad y el uso de datos estratégicos.

Por ese motivo, el grupo de investigación a cargo del proyecto UBACyT 773BA que tiene por objetivo contribuir en la elaboración de un modelo conceptual para la creación de un mapa de las investigaciones académicas y científicas en Argentina basado en las nuevas tendencias tecnológicas de Datos Abiertos Enlazados (*Linked Open Data, LOD*), se ha propuesto en este trabajo, desde el punto de vista metodológico, analizar de manera sistemática las diversas cuestiones a considerar al diseñar una ontología para que los datos sean formateados de manera apropiada, cuenten con los metadatos necesarios y las ontologías aseguren la reutilización de vocabularios existentes. Asimismo, se apunta a que posean documentación suficiente y sean publicadas bajo licencias abiertas a fin de que puedan visualizarse y reutilizarse.

### Primeras consideraciones para la aplicación de LOD y la creación de ontologías

En primer lugar, dado que una ontología no es ambigua, se requiere definir qué es posible explicitar dentro de ella mediante clases, subclases, propiedades y subpropiedades. El nivel de granularidad pauta el grado de especificidad con el que se podrá formular una interrogación al usar la ontología. Una clase es un conjunto de objetos con propiedades comunes. Las subclases heredan las propiedades de la clase y permiten establecer distinciones entre propiedades que pueden asociarse solo a un subconjunto puntual de la clase. Las propiedades son los atributos vinculados a determinadas clases/subclases.

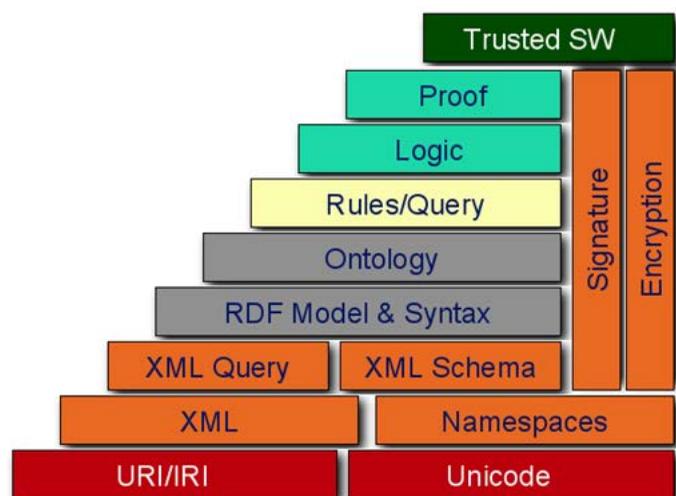
El modelo RDF (Klyne y Carroll, 2004 [2014]), permite representar información por medio de grafos dirigidos en los cuales los vértices tienen un sentido definido que constituyen triples. La estructura del triple RDF (sujeto, un recurso– predicado, una propiedad– objeto, un valor o literal) permite, sin limitaciones, que se puedan enunciar afirmaciones sobre cualquier recurso. En la descripción de los recursos cabe incluir diversos tipos de triples (Heath y Bizer, 2011):

- » Triples que describen el recurso con literales.
- » Triples que lo describen mediante un vínculo hacia otros recursos.
- » Triples que lo describen mediante un vínculo desde otros recursos.
- » Triples que describen recursos relacionados.
- » Triples que describen el *data set* más amplio del cual forma parte el recurso.

El proceso de construcción de la descripción de un recurso puede resultar dificultoso si los datos iniciales han sido estructurados en una tabla o en una base de datos relacional, a raíz de la ambigüedad que encierra el texto plano. Por ese motivo, es necesario, en segundo lugar, explicitar cada elemento del grafo RDF mediante un URI unívoco (Heath y Bizer, 2011; Stuart, 2016).

Al enunciar cada propiedad es importante indicar:

- a) cuál es su cardinalidad (el número de veces que esta puede ser asociada a una entidad en particular)



Fuente de los datos: W3C (2005). <http://www.w3c.it/talks/2005/openCulture/slide7-o.html>

Figura 1. Niveles de la web semántica

b) el tipo de objetos a los que puede restringirse (por ejemplo, a una cadena de caracteres o a un enlace hacia otra entidad)

Asimismo, hay una serie de tecnologías involucradas en la operacionalización de las ontologías. El conjunto de estas herramientas constituye la arquitectura de la Web Semántica y se representa a través de capas o niveles (*semantic web stacks* o *semantic web layer cake*) (Figura 1).

Se aplica el juego de caracteres Unicode porque RDF se halla codificado como texto. Se usan identificadores unívocos URIs, siendo los más comunes los URLs (*Uniform Resource Locators*), aunque también se incluyen URNs (*Uniform Resource Names*) dado que son identificadores independientes de la localización. Cada vez más se hace referencia a los IRIs (*Internationalized Resource Identifiers*), antes que a los URIs para no limitar los identificadores de recursos a los caracteres del alfabeto latino. Sin embargo, en el ámbito del desarrollo de las ontologías se elige aún el término URI.

Se indica el uso de URIs desreferenciables o redireccionables competentes para recuperar tanto un recurso en particular como la información asociada con dicho recurso (Heath y Bizer, 2011; Sauermann, Cyganiak y Völkel, 2006; Sauermann y Cyganiak, 2008). En el contexto de LD, los URIs se utilizan para identificar objetos del mundo real y conceptos abstractos. Las descripciones de objetos orientadas a la lectura humana se presentan en HTML, y aquellas diseñadas para su lectura mediante computadoras se representan como triples RDF. Con el propósito de satisfacer ambas demandas se implementa un mecanismo HTTP denominado negociación de contenido (*content negotiation*), mediante el cual los clientes HTTP envían cabeceras HTTP con cada interrogación para indicar qué clase de documentos prefieren, y los servidores las inspeccionan para dar lugar a la negociación que conduzca a la respuesta apropiada. Cuando el servidor debe enviar documentos RDF al responder a una solicitud, es fundamental que no exista confusión para la computadora entre el URI que identifica un objeto real (el objeto en sí mismo), y el documento Web que lo describe. Por ese motivo, a fin de desambiguar, se declaran URIs diferentes para uno y otro. Existen dos estrategias para cumplimentar este requisito: 303 URIs y *hash URIs*.

Mediante la estrategia 303 URIs, ante la demanda (*HTTP GET*) de un objeto real, el servidor envía al cliente la respuesta *HTTP code 303 See Other* y el URI del documento Web que describe dicho objeto. Este mecanismo se denomina *303 redirect*. El cliente,

entonces, por medio de *HTTP GET* desreferencia o redirecciona su pedido hacia el URI entregado por la primera respuesta del servidor, y obtiene de este, a través de una segunda respuesta *code 200 OK*, el documento Web apropiado, que describe el objeto del mundo real.

Si se adopta, en cambio, la estrategia *hash URI*, se apela a la capacidad de este para separar una parte de sí de la base que lo conforma, encabezada por el símbolo *hash* (#); esta parte se denomina fragmento identificador (*fragment identifier*). Cuando el cliente solicita recuperar un *hash URI*, el protocolo HTTP requiere separar el fragmento del resto del URI, es decir, trunca el URI, antes de formular su pregunta al servidor a través de HTTP GET. El servidor como respuesta envía el documento RDF/XML solicitado que contiene todos los triples que poseen URIs con la base incluida en la interrogación. En última instancia, el cliente *Linked Data-aware* debe inspeccionar la respuesta y hallar el URI que cuenta con el fragmento identificador de su interés.

La adopción de *hash URIs*, reduce la latencia de acceso, es decir, el tiempo que tarda en transmitirse un paquete dentro de la red y el número de consultas/respuestas entre el cliente y el servidor, o sea, los viajes ida y vuelta en el transcurso de la negociación. Su desventaja reside en el hecho de que la respuesta a la primera consulta trae todos los URIs con igual base, aunque estos sean numerosos y ello puede resultar contraproducente desde el punto de vista de la eficiencia y precisión de la respuesta. Por el contrario, en los 303 URIs el redireccionamiento puede configurarse para cada recurso por separado, mediante un único documento para todos los recursos o por medio de cualquier combinación de ambos criterios. Esta particularidad acrecienta su flexibilidad.

En consecuencia, los 303 URIs se aplican a los conjuntos de datos muy extensos (por ejemplo, la DBpedia, que incorpora más de tres millones de conceptos). Los *hash URIs*, en cambio, resultan apropiados para vocabularios restringidos, menos dinámicos. Por ese motivo, Protégé, como editor de ontologías, los adopta por defecto. De todas maneras, es importante destacar que las ventajas de ambas soluciones pueden combinarse al utilizar URIs que siguen el patrón `http://domain/resource#this`, por ejemplo, `http://www.example.com/bob#this`, para designar a la persona “Bob” con un URI combinado (Sauermann y Cyganiak, 2008).

Debe decidirse también, si los URIs serán descriptivos u opacos. Es común que la última parte de un URI incluya un nombre descriptivo, en lenguaje natural, de la entidad a identificar, denominado URI descriptivo o URI significativo. Si, por el contrario, se adoptan URIs opacos, se incluyen notas específicas o etiquetas (*labels*), a fin de proporcionar una versión legible para los seres humanos. Cuando se utilizan URIs descriptivos, esta parte significativa del URI se utiliza para lexicalizar las entidades, tanto como las notas (por ejemplo, `rdfs:label`).

Coexisten diferentes criterios para resolver las cuestiones referidas al contenido y la forma tanto del URI como del nombre, término o etiqueta de un concepto en el contexto de las ontologías Web. En los campos de la Biología y la Biomedicina, se recomienda entre las buenas prácticas la separación de cada concepto de la etiqueta elegida para designarlo, la que varía de un idioma a otro y se halla expuesta a la sinonimia y la polisemia. Tanto OWL como OBO (*Open Biological and Biomedical Ontology*) aceptan dicha separación. Las etiquetas RDFS en OWL, pueden usarse para designar el nombre sin que este se corresponda con el URI, que puede ser numérico. Incluso, los principios de OBO promueven que el ID se constituya libre de semántica. La asignación de identificadores numéricos o alfanuméricos no semánticos y constituidos automáticamente contribuye a una adecuada gestión tanto de la ontología como de los datos (Cimino, 1998; Stevens y Hull, 2010).

Montiel-Ponsoda et al. (2011), se refieren a las diversas iniciativas que han abordado esta problemática en el contexto del diseño de ontologías Web (Berners-Lee, 1998; Bizer, Cyganiak y Heath, 2008; Designing URI Sets for the UK Public Sector, 2009; Fliedl, Kop y Vöhringer, 2007; Noy y McGuinness, 2001; Schober et al., 2009; Théreaux, 2003). Estos autores resumen los antecedentes existentes sobre el tema y la experiencia que han aportado algunas implementaciones, tales como la ontología FOAF, *Friend of a Friend*, la conversión de la familia de modelos FRBR a RDF (IFLA. FRBR Review Group, 2011) y el estándar ISBD, *International Standard Bibliographic Description* (Willer, Dunsire y Bosančić, 2010). Sobre esa base, enumeran aquellas convenciones que responden a las buenas prácticas con relación a dichas ontologías, los que pueden observarse en la Tabla 1.

Guía para la Web Multilingüe	Ejemplo
Prefiera URIs opacos para los nombres locales	<a href="http://aims.fao.org/aos/agrovoc/c_330988">http://aims.fao.org/aos/agrovoc/c_330988</a>
Cuando aplique URIs descriptivos para los nombres locales, use su propia lengua con los diacríticos si correspondiera	<a href="http://geo.linkeddata.es/ontology/río">http://geo.linkeddata.es/ontology/río</a>
Cuando aplique URIs descriptivos para los nombres locales, use guión bajo o "CamelCase" como delimitadores de las palabras, según se lea con mayor facilidad en su propia lengua	<a href="http://geo.linkeddata.es/ontology/aguas_corrientes">http://geo.linkeddata.es/ontology/aguas_corrientes</a>

Tabla 1. Guía para el elemento nominación

Fuente de los datos: Montiel-Ponsoda et al. (2011)

En primer lugar, efectúan "... a clear distinction between URI local names and labels, as annotation properties or as part of external linguistic models related to ontology elements..." (Montiel-Ponsoda et al., 2011: 112). Dado que los URIs funcionan como identificadores, consideran que cuando se recurre al lenguaje natural para definirlos, deben ser breves, poco semánticos, compactos y presentarse en CamelCase (los espacios entre las palabras se reemplazan por letras mayúsculas al comienzo de cada vocablo que compone el nombre de la clase) o mediante guión bajo, según resulte conveniente en función del lenguaje utilizado.

Los URIs opacos para los nombres locales se orientan hacia la legibilidad por medio de computadoras y proporcionan estabilidad a las ontologías. Por ese motivo, no deben modificarse una vez que la ontología se publica y es adoptada por distintas comunidades de usuarios. Son especialmente útiles en el ámbito de la Web Multilingüe, ya que evitan sesgos en favor de una lengua determinada. Los URIs opacos han sido adoptados para los modelos FRBR, para las ISBD y las RDA, además, se adaptan a la evolución de los términos y facilitan la incorporación de los datos a los sistemas (Van Hooland y Verborgh, 2014). Aunque se implemente su uso, las etiquetas se usan para documentar la ontología en lenguaje natural.

Cabe normalizar el estilo de nominación de clases y propiedades. Si se sigue el adoptado por la Web Semántica en el W3C Web Ontology Language (OWL), se ha de utilizar CamelCase y un CURIE (URI compacto) para señalar en forma abreviada que un objeto proviene de una misma ontología (por ejemplo, owl:allValuesFrom; owl:maxCardinality; owl:complementOf; foaf:Agent; foaf:PersonalProfileDocument; foaf:isPrimaryTopicOf).

Los lenguajes legibles por computadora como RDFs, SKOS y OWL proporcionan los mecanismos para asociar las descripciones expresadas mediante el lenguaje natural con las entidades de las ontologías, denominándose este proceso *lexicalización* (*ontology lexicalization*). Dicho proceso resulta sumamente útil en la instancia durante la cual se vinculan las descripciones en lenguaje natural para denominar o etiquetar dichas entidades. Asimismo, los datos pueden hallarse en un idioma diferente al usado por la ontología para etiquetar los conceptos, requiriéndose que se implementen dispositivos de traducción. El proceso de adaptación de la ontología a una comunidad de usuarios particular y diferenciada se conoce como *localización* (*ontology localization*) (Aguado-de-Cea et al., 2015; Cimiano et al., 2010; Cimiano et al. 2013; Espinoza Mejía et al., 2012; García et al. 2012). Ambos procesos “...for the purposes of contributing to the linking process in the context of Linked Data or for the adoption by users from different linguistic and cultural communities, [...] have become a priority in the current ontology engineering research...” (Aguado-de-Cea et al., 2015: 152).

Aguado-de-Cea et al. (2015), tienen en cuenta las recomendaciones que han efectuado previamente distintos autores, especialmente Schober et al. (2009), con respecto a dichos procesos, y sobre esa base, resumen aquellas acciones consideradas como buenas prácticas:

» Para la lexicalización de ontologías

1. Las etiquetas de las clases deben capturar el significado de la clase a la que se refieren de la manera más breve, concisa y autoexplicativa posible, aunque se han de respetar las estructuras sintagmáticas naturales de cada idioma.
2. Debido a que, según se ha expresado, la etiqueta tiene por función describir y explicar el significado de la clase, debe evitarse el uso de palabras abreviadas.
3. Incluir aquellas etiquetas que expresen sinónimos y variantes si se utilizan como equivalentes del término adoptado en un determinado dominio.
4. Las etiquetas de las clases deben registrarse en singular. Es recomendable elegir términos que acepten sufijos para cumplir con el paradigma derivativo, por ejemplo, creador, creación, es creador de, creado por.
5. Las etiquetas para las propiedades deben adoptar una estructura sintáctica constituida al menos por un verbo acompañado por un objeto y/o una preposición con la finalidad de dar al triple una forma cercana al lenguaje natural. Por ejemplo, si se trata de una Organización, tiene sede en. Estas etiquetas también deben respetar las convenciones del idioma al que se traducen.
6. La traducción a un idioma en el que hay diferenciaciones de género debe reflejarse en las etiquetas, según se refieran a nombres femeninos o masculinos.
7. Las definiciones de los comentarios y notas de uso deben seguir el formato establecido en las UNE-ISO 1087-1 (2009) y UNE-ISO 1087-2 (2009). De acuerdo con ellas, la definición debe incluir el concepto supraordenado y las características diferenciales.

» Para la localización de ontologías

1. Antes de localizar una ontología debe consultarse la documentación relativa a las necesidades que justifican su creación, a fin de comprender su uso y propósito, así como el significado de sus clases y propiedades.
2. Debe examinarse la versión más actualizada del documento de especificación de la ontología y sus recomendaciones.
3. Es conveniente buscar otras traducciones de la ontología que puedan ayudar durante el proceso de translación, sobre todo cuando se traducen idiomas pertenecientes a la misma familia lingüística.
4. La interpretación correcta del significado y el alcance de las entidades de la ontología conlleva proporcionar descripciones en lenguaje natural de cada una de

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"
  xmlns:d="http://description.org/" >
  <rdf:Description about="http://www.etsii.us.es/user/333/docs">
    <d:Creator>
      <rdf:Description about="http://www.etsii.us.es/user/333">
        <rdf:type resource="http://www.etsii.us.es/schema/Person"/>
        <p:Name>Elena Lozano</p:Name>
        <p:Email>elozano@us.es</p:Email>
      </rdf:Description>
    </d:Creator>
  </rdf:Description>
</rdf:RDF>

```

Figura 2. Serialización de triples RDF en formato RDF/XML.

Fuente de los datos: Lozano Rosch (2007)

- ellas mediante comentarios, definiciones o notas de uso. Si estas anotaciones ya existen en la ontología que se traduce, deben leerse y traducirse antes de seleccionar el término que servirá como etiqueta en la lengua de la traducción.
5. Se aplica para la localización de la ontología lo establecido en las normas UNE-ISO 1087-1 (2009) y UNE-ISO 1087-2 (2009).
  6. Se sugiere que otros traductores y lingüistas, tanto como usuarios de la ontología, revisen el trabajo de traducción realizado.

Con respecto a la codificación, se prefiere la sintaxis XML (Extensible Markup Language) para serializar los triples RDF en un formato conocido como RDF/XML (véase Figura 2), debido a que la mayoría de los datos de la Web Semántica se hallan en este formato. Además, XML acepta el uso de CURIEs (URIs Compactos expresados en CURIE, una sintaxis abreviada para expresar Identificadores de Recursos Uniformes) con prefijo *namespace* que permite referenciar elementos comunes y reutilizar conceptos ya desarrollados (Figura 2).

Es importante diferenciar los datos en sí mismos de la forma en que estos han de ser codificados, es decir, de cómo han de ser serializados. Tal como se ha destacado, RDF/XML es la serialización o codificación por defecto para RDF en la Web Semántica, aunque no es la más amigable. Se requiere, por tanto, validarla mediante el W3C RDF Validator, o, cuando se trata de archivos RDF extensos, *on the desktop*, por ejemplo, mediante Apache-Jena (Apache-Jena Java Framework RIOT package permite validar un archivo RDF desde línea de comando y ARQ package, interrogarlo).

Asimismo, hay serializaciones alternativas, populares y más amigables, entre otras, Turtle, N-Triples, N-Quads, JSON-LD (Beckett, 2014; Beckett, Berners-Lee et al., 2014; Davis, Steiner y Le Hors, 2013). Cada una de ellas posee ventajas y tiene desventajas. Es común que los mismos datos se encuentren disponibles en distintas serializaciones. RDF-Translator permite cambiar de una serialización a otra. También hay herramientas que trabajan por línea de comandos: RDF2RDF; Apache-Jena Java framework RDFcat.package (Stuart, 2016).

Por otra parte, para ser útiles, los triples no deben carecer de estructura, siendo necesario que cuenten con determinadas restricciones. Por ello, en el siguiente nivel de estructuración de los datos se emplean taxonomías, ontologías, lenguajes de interrogación y reglas (Arp, Smith y Spear, 2015; Heath y Bizer, 2011; Stuart, 2016):

- a) La taxonomía RDF Schema (RDFS) proporciona el vocabulario para definir un primer nivel de restricciones, a fin de proveer ya no triples sueltos sino conjuntos de grafos (Allemang y Hendler, 2011). Habilita, también, un nivel simple de inferencia al disponer de un vocabulario común para enunciar rangos y dominios de las propiedades. Sin embargo, al contar con limitaciones debe utilizarse OWL para establecer todas las restricciones que requiere el buen funcionamiento de la ontología.
- b) La ontología OWL, actualizada en OWL 2, ofrece tres perfiles, OWL 2 EL, OWL 2 QL y OWL 2 RL. Permite formular reglas de razonamiento formales que favorecen la producción de conocimiento por inferencia, entre las que se mencionan: que algo es miembro de una clase y no es miembro de otra; que una clase es unión de otras dos; que una propiedad es inversa de otra; que un individuo es distinto de otro, aunque su nombre sea el mismo.
- c) SPARQL (Simple Protocol and RDF Query Language; SPARQL 1.0, SPARQL 1.1) es el lenguaje de interrogación, donde cada interrogación representa un grafo con variables conocidas y desconocidas, a ser contrastado frente a otros grafos por un motor de interrogación. Se han diseñado variantes de este lenguaje para satisfacer situaciones específicas (Malik, Goel y Maniktala, 2010), razón por la cual conviene revisar la documentación al consultar un repositorio de triples.
- d) Otros lenguajes de reglas y restricciones (RIF, *Rule Interchange Format*; SWRL, *Semantic Web Rule Language*; SPIN, SPARQL Inferencing Notation) pueden usarse para mejorar los lenguajes de ontologías. A través de ellos es posible describir reglas y relaciones que no pueden codificarse. Por ejemplo, en OWL, se requiere para su implementación el uso de motores de inferencia o razonamiento (Rattanasawad et al., 2013).

## Pasos metodológicos para el diseño e implementación de ontologías

Tonkin, Pfeiffer y Hewson (2010) identifican tres tipos de metodologías para desarrollar ontologías. En primer lugar, las auto-reflexivas en las que, dadas las características acotadas del dominio, una sola persona aborda el diseño sobre la base de los conocimientos que posee. En segundo lugar, las colaborativas, convenientes cuando debe abordarse un dominio extenso que exige considerar múltiples facetas y consensuar puntos de vista de individuos con conocimientos especializados y diferentes. Por último, las empíricas que toman como punto de partida los datos registrados en documentos de interés en función del dominio elegido.

Arp, Smith y Spear (2015), autores del proyecto *Basic Formal Ontology* (BFO), una ontología de nivel superior, indican que los pasos a seguir para diseñar una ontología de dominio son los siguientes:

1. Demarcar el alcance, la materia y el nivel de granularidad de la realidad a representar
2. Reunir información, identificar los términos generales más usados en el dominio, verificar su existencia en ontologías de referencia y eliminar las redundancias
3. Ordenar los términos resultantes jerárquicamente, de lo general a lo específico
4. Normalizar el resultado para asegurar:
  - a. Coherencia lógica, filosófica y científica
  - b. Coherencia y compatibilidad con ontologías vinculadas
  - c. Comprensión por parte de los seres humanos a través de definiciones consistentes
5. Formalizar el artefacto representacional normalizado en un lenguaje de computación, de tal manera que el resultado pueda implementarse en un entorno automatizado.

Por otra parte, diversas metodologías han sido reconocidas y aplicadas en el ámbito del diseño de las ontologías Web (Aguado-de-Cea et al., 2015; Alvarado, 2010; Bautista-Zambrana, 2015a, 2015b; Dhingra y Bhatia, 2015; Fonseca Marangon et al., 2016; Kanoh et al. 2015; Kyriaki-Manessi y Dendrinou, 2014; López Rodríguez, Hidalgo Delgado y Silega Martínez, 2016; Malik et al., 2010; Malik, Prakash y Rizvi, s. f., 2011; Malviya, Mishra y Sahu, 2011). Entre las más difundidas, se encuentran las expuestas por Grüninger y Fox (1995), Methontology (Fernández-López, Gómez-Pérez y Juristo, 1997), Noy y McGuinness (2001) y Uschold y King (1995), quienes claramente reconocen pasos a seguir. A pesar de los matices propios que cada una aporta, las cuatro brindan, dentro del proceso total, mayor importancia a la fase de construcción de la ontología. Stuart (2016) enriqueció el modelo al combinar las cuatro metodologías que han servido como antecedente y agregar dos pasos más. Propone, así, un método constituido por doce pasos que incorpora la identificación del software apropiado y la sustentabilidad de la ontología.

### *Uschold y King (1995)*

1. Identificar el propósito
2. Construir la ontología
  - a. Capturar la ontología
  - b. Codificar la ontología
  - c. Integrarla con ontologías existentes
3. Evaluar
4. Documentar

### *Grüninger y Fox (1995) – Metodología TOVE*

1. Determinar la competencia de la ontología
2. Definir la terminología de la ontología
3. Especificar las definiciones y restricciones de la terminología
4. Probar la competencia de la ontología para demostrar la integridad de las teorías

### *Fernández-López, Gómez-Pérez y Juristo (1997) – Methontology*

1. Especificación
2. Adquisición de conocimiento
3. Conceptualización
4. Integración
5. Implementación
6. Evaluación
7. Documentación

### *Noy y McGuinness (2001) – Simple knowledge-engineering methodology*

1. Determinar el dominio y alcance de la ontología
2. Considerar la reutilización de ontologías existentes
3. Enumerar términos importantes en la ontología
4. Definir las clases y las jerarquías
5. Definir las propiedades de las clases
6. Definir las facetas
7. Crear instancias

### Stuart (2016)

1. Alcance de la ontología
2. Reutilización de la ontología
3. Identificación del software apropiado
4. Adquisición de conocimiento
5. Identificación de términos importantes
6. Identificación de términos adicionales, atributos y relaciones
7. Especificación de las definiciones
8. Integración con ontologías existentes
9. Implementación
10. Evaluación
11. Documentación
12. Sustentabilidad

Si bien Stuart (2016) destaca que la construcción es un proceso básicamente iterativo, reconoce diversas cuestiones a considerar en cada uno de los pasos enunciados por él. Con relación al alcance es necesario tener en claro desde un primer momento el propósito, los usuarios y el contenido de la ontología. Se debe tener en cuenta en el diseño si la ontología se ha de utilizar para la indización de recursos o para la navegación y la recuperación de los mismos. La reutilización de ontologías existentes reduce los costos y facilita la interoperabilidad de los datos. Para facilitar la reutilización es conveniente incluir información relacionada con el estatus de los términos que han sido contemplados, ejemplos de uso, etc., mediante notas o anotaciones incorporadas como propiedades (*annotation properties*). La *Information Artifact Ontology* (IAO) contiene 57 propiedades destinadas a tal fin y puede consultarse para determinar cuáles resultan pertinentes en la ontología que se diseña.

Con respecto al software es fundamental saber que, posiblemente, un único software no reuna todos los requerimientos que se exigen para el desarrollo e implementación de la ontología. Puede que se use una aplicación para la ingesta de datos, otra para el proceso central de desarrollo y una tercera para la consulta. También, si se adopta una herramienta de uso libre como, por ejemplo, *Protégé* o un producto comercial como *Top Braid Composer*, entre otros. Se evaluará su popularidad, extensibilidad, interface (textual o gráfica), lenguaje y si es monousuario o multiusuario, en caso de construcciones colaborativas; si se trata de un editor orientado a un dominio específico, si permite la simple ingestión de ontologías existentes y de datos estructurados y si logra ajustar datos capturados a través del procesamiento de lenguaje natural (Stuart, 2016).

La adquisición o captura de conocimiento puede realizarse de dos maneras, mediante elicitación (u obtención), que facilita el traspaso fluido de conceptos asociados por medio del lenguaje, a través de personas involucradas con tales conceptos, o por medio del descubrimiento, al extraer la información necesaria de los documentos existentes. Shabolt y Smart (2015), identifican métodos naturales (entrevista, análisis de protocolos, decisión crítica) e inventados (clasificación de conceptos, mapeo de conceptos y procesos) de elicitación. Fernández-López, Gómez-Pérez y Juristo (1997), reconocen varias técnicas posibles de adquisición a partir de la realización de entrevistas y del análisis de texto: las entrevistas no estructuradas o estructuradas con expertos, el análisis de texto informal o formal mediante técnicas de procesamiento de lenguaje natural (*Natural Language Processing*, NLP) o la minería de datos textuales no estructurados (*text mining*).

Una vez adquirido el conocimiento necesario, el siguiente paso exige identificar los términos importantes en correspondencia con el alcance asignado a la ontología, previamente documentado. Este paso, demanda evaluar si, de acuerdo con el dominio elegido, es posible considerar términos sin la evidencia de que existan instancias que

los requieran o, por el contrario, deben establecerse restricciones en este sentido e incluir solo aquellos términos para los que existen instancias, de acuerdo con lo que recomiendan Arp, Smith y Spear (2015). Además, dado que no se trata de construir un vocabulario controlado sino un modelo de datos es necesario asociar las entidades seleccionadas a los atributos y las relaciones que permitan enriquecer dicho modelo.

El valor agregado de la ontología depende en gran parte de la definición formal de los conceptos ya que ello asegura un uso consistente de la misma. Arp, Smith y Spear (2015) recomiendan considerar la BFO, ontología destinada a servir como herramienta para resolver adecuadamente esta tarea de acuerdo con principios ontológicos filosóficos que tienen en cuenta las clases (universales) vs. las instancias (particulares), las jerarquías (es un, es parte de, tiene una parte), los solapamientos, las dependencias (es inherente a, es portador de). Así mismo, la integración con otras ontologías, clave de la Web Semántica, puede expresarse mediante propiedades de amplio uso tales como OWL: `equivalentClass`, `rdfs: subclassOf`, `owl: equivalentProperty`, `rdfs: subPropertyOf`, `OWL: sameAs`, `skos: broader`, `skos: narrower`. Para garantizar la correcta resolución del nivel lógico se requiere validar relaciones y propiedades, con esa finalidad la *Ontology Alignment Evaluation Initiative* (2018), OAEL, gestiona un foro que ofrece distintos abordajes a ser testeados y tiene por objetivo evaluar las fortalezas y debilidades de los sistemas de alineación/coincidencia y comparar su rendimiento para mejorar las técnicas de evaluación. Por su parte, Köhler et al. (2006) proponen un método computacional para controlar automáticamente la calidad de los términos y de sus definiciones en ontologías y taxonomías.

La implementación, según Stuart (2016), es el estadio donde se integran las distintas partes que se han desarrollado, recién entonces es posible iniciar un proceso de evaluación continua. Vrandečić (2010) menciona ocho criterios mediante los cuales puede evaluarse una ontología: exactitud de la representación del dominio; adaptabilidad y grado de escalabilidad; claridad, en términos de facilidad de comprensión para los usuarios; completitud, con referencia a cobertura y granularidad; eficiencia computacional, con respecto a rapidez y facilidad de razonamiento automático; concisión, como la capacidad para excluir axiomas e instancias irrelevantes; consistencia/coherencia, desde el punto de vista lógico; y aptitud organizacional, que indica con qué facilidad una ontología puede desplegarse en una organización. Asimismo, Vrandečić (2010) identifica seis aspectos a evaluar:

- » Vocabulario, tanto se hayan elegido URIs o literales para representar el conjunto de nombres de la ontología;
- » Sintaxis, especialmente relacionada con la normalización en la serialización;
- » Estructura, se evalúa el grafo RDF que describe la ontología Web;
- » Semántica, una ontología consistente puede ser interpretada por un conjunto infinito de posibles modelos, las semánticas son las características comunes de todos ellos;
- » Representación, se refiere a la relación estructura/semánticas. Se evalúa por comparación de métricas calculadas sobre el grafo RDF con las características de los posibles modelos, según se especifica en la ontología;
- » Contexto, tiene en cuenta los aportes de la ontología comparados con otros dispositivos en su entorno, por ejemplo, con relación a restricciones semánticas adicionales, a una manera diferente de representar los datos o a los requerimientos que se han formalizado.

En paralelo con la implementación, proporcionar la documentación es fundamental para garantizar el uso y la reutilización de la ontología. Ella misma es el núcleo a documentar, por lo tanto, aunque hay distintas maneras de cumplir con este requisito, la publicación en el sitio Web de un archivo de texto que la contenga en RDF/XML

resuelve básicamente dicha instancia. Sin embargo, es conveniente complementar la información codificada con documentos narrativos que proporcionen una introducción a la ontología y el mapeo hacia las ya existentes (Véase, por ejemplo, <http://www.ontobee.org/ontology/bfo> y <http://basic-formal-ontology.org/>).

Por último, la sustentabilidad exige contar con los recursos necesarios para actualizar la ontología a fin de asegurar a través del tiempo la vigencia de su funcionalidad, así como su significación y utilidad dentro del dominio que abarca.

Allemang y Hendler (2011: 311) afirman que el arte de modelar en la Web Semántica consiste en "...combining the building blocks in useful ways to create a dynamic system through which the data of the Semantic Web can flow.". Asimismo, consideran que los peligros a evitar se sintetizan en cinco categorías o antipatrones (Allemang y Hendler, 2011):

1. El clasismo sin límite (*rampant classism*) que tiende a representar todo mediante clases, en lugar de introducir propiedades, relaciones e instancias en el modelo;
2. El supuesto de exclusividad (*exclusivity*) por el cual se asume que, si todos los miembros de una subclase pertenecen a la misma superclase en un momento dado, los miembros futuros también pertenecerán a ella;
3. La cosificación (*objectification*) es contraria a los tres supuestos que constituyen el contexto de la Web Semántica: AAA (*Anyone can say Anything about Any topic*), visión abierta del mundo (*Open World*) y ausencia del requisito de nominación única (*Nonunique Naming*). Asimismo, esta basada en la creencia errónea de que una clase es una plantilla para crear instancias, cuando una instancia puede ser miembro de múltiples clases y las propiedades existen independientemente de dichas clases;
4. El uso de identificadores para las clases (*managing identifiers for classes*) es un requerimiento común en algunos modelos. Sin embargo, la utilización de una propiedad, tal como un identificador, para describir la clase, entraña el riesgo de generar confusión con respecto a si se describe la clase o un individuo que pertenece a la misma.
5. La conceptualización exagerada (*creeping conceptualization*) que va más allá de los requerimientos del dominio y de los objetivos de la ontología.

Una vez diseñadas e implementadas, las ontologías pueden ser interrogadas. Según Stuart (2016) los motivos para interrogarlas pueden responder a la necesidad de reutilizarlas, de consultar una base de conocimiento o de comprender su uso. Para explorar una ontología las tecnologías apropiadas son los editores, los visualizadores o los motores de razonamiento, mientras que para consultar una base de conocimiento se emplean motores de búsqueda o asistentes personales cuando se trata de preguntas simples. En cambio, ante preguntas complejas es imprescindible, recurrir a lenguajes de interrogación formales (SPARQL Protocol and RDF Query Language) (Bonifati, Ciucanu y Lemay, 2015); si se desea indagar sobre el uso de una ontología se utilizan los motores de búsqueda, las librerías de ontologías y los rastreadores semánticos que proporcionan métricas útiles para conocer sus particularidades en cuanto a conformación e impacto.

## Metodología de diseño adoptada por el proyecto UBACyT 773BA

Una vez examinadas diversas cuestiones tecnológicas relacionadas con la implementación de LOD (Berners-Lee, 2006; Heath y Bizer, 2011; Sauermann, Cyganiak y Völkel, 2006; Stuart, 2016) y diferentes opciones metodológicas para la creación de una ontología (Arp, Smith y Spear, 2015; Fernández-López, Gómez-Pérez y Juristo, 1997;

Grüninger y Fox, 1995; Noy y McGuinness, 2001; Stuart, 2016; Uschold y King, 1995), el equipo de investigación resolvió aplicar el modelo expuesto por Stuart (2016) y registrar las decisiones iniciales a tomar en cada una de las etapas incluidas en su propuesta:

### *Alcance de la ontología*

Tipo de ontología: aplicada

Propósito: navegación y recuperación de recursos

Usuarios: catálogos y repositorios científicos, académicos y técnicos

Contenido: representación de la producción científico-técnica argentina

### *Reutilización de la ontología*

Para facilitar la reutilización se incluye información relacionada con el estatus de los términos que han sido contemplados, en relación con otras ontologías existentes.

### *Identificación del software apropiado*

Se adopta *Protégé*, en función de su popularidad y de que cuenta con un razonador semántico, piezas de software destinadas a inferir las consecuencias de las afirmaciones de la ontología, dado que la información al ingresarse en este editor tiene una estructura explícita que puede ser comprendida como un axioma. El motor de razonamiento usa reglas lógicas para obtener conclusiones de los axiomas y de las afirmaciones. En consecuencia, puede determinar nuevas relaciones entre los universales o clases de la ontología y utilizarse para verificar la consistencia de la información ingresada. Incluso, se ha desarrollado para este software un plugin que implementa el procedimiento de árboles de decisión para evaluar la consistencia de la ontología (Seyed y Shapiro, 2012). Por otra parte, permite importar términos (clases) relevantes de otras ontologías, como, por ejemplo, BFO e IAO.

### *Adquisición de conocimiento*

Dada la naturaleza de la ontología a diseñar, se decide que el grupo de investigación constituye una fuente de conocimiento adecuada para encarar la primera selección de los términos que la conforman.

### *Identificación de términos importantes*

La ontología se estructura en torno a la columna vertebral conformada por la jerarquía *is\_a*. Esta cualidad se testea mediante la validación de la afirmación *A is\_a B*, donde cada instancia de A es una instancia de B, cuando A y B refieren a universales o a clases definidas. Se asegura una sola herencia para cada término; ya que una única cadena de relaciones padre/hijo favorece la etapa de elaboración de definiciones de acuerdo con el criterio aristotélico y la capacidad del razonador automático para efectuar inferencias correctas. Con respecto a la terminología, solo se representan aquellas clases para las que existen instancias, que cuenten con evidencia de que se poseen miembros y se seleccionan los nombres y frases comunes preferidos en el dominio, usándose para los tipos de entidades más importantes aquellos términos preferidos por los grupos de científicos más influyentes. Se asegura el significado unívoco de los términos adoptados, aunque múltiples términos pueden referir a uno preferido con el cual han sido asociados. Se usan nombres y frases comunes que refieren a un universal o a una clase definida, en singular para que los razonadores de las aplicaciones trabajen correctamente. Se adhiere a la convención "CamelCase" a fin de normalizar los nombres constituidos por frases, dado que en idioma español la aplicación de esta convención no genera inconvenientes. Se evita el uso de acrónimos

y abreviaturas; cada término, para satisfacer los requerimientos computacionales, se asocia con un único identificador; se usan nombres que pueden cuantificarse; no se crean términos a través de combinaciones lógicas para los universales, se evitan los términos negativos, los universales son en todos los casos positivos.

### *Identificación de términos adicionales, atributos y relaciones*

Con relación a la estructuración de los datos se adopta el RDF Schema (RDFS) para definir un primer nivel de restricciones. RDFS permite la representación de dominios y rangos, por ejemplo, la propiedad *authored\_by*, con dominio *document* y rango *person*. Introduce además, un número de predicados que aumentan la expresividad de la ontología (*rdfs:label*, *rdfs:comment*, *rdfs:seeAlso*, *rdfs:subClassOf*, *rdfs:domain*, *rdfs:range*). Aun así, tal como se indica en el presente trabajo, RDFS no puede expresar las relaciones transitivas, simétricas, reflexivas, inversas ni restricciones de existencia o cardinalidad. Por ello, se recurre a OWL 2 (Web Ontology Language 2), que permite formular otras reglas de razonamiento formales y acepta expresiones tales como: cuantificación universal ( $\forall$ ), mediante la restricción *owl:allValuesFrom*; cuantificación existencial ( $\exists$ ), a través de las restricciones *owl:someValuesFrom*, *owl:hasValue*; cardinalidad, *owl:cardinality*, *owl:minCardinality*, *owl:maxCardinality*; operadores booleanos, *owl:intersectionOf*, *owl:unionOf*, *owl:complementOf*; equivalencia, *owl:equivalentClass*, *owl:equivalentProperty*. Asimismo, admite declarar propiedades para establecer relaciones (inversa, funcional, inversa funcional, transitiva, simétrica, asimétrica, reflexiva, irreflexiva). Se garantiza, de esta manera, la unicidad de las expresiones relacionales. Al enunciar cada propiedad se indica cuál es su cardinalidad (el número de veces que una propiedad puede ser asociada a una entidad en particular) y el tipo de objetos a los que puede restringirse.

### *Especificación de las definiciones*

De acuerdo con las pautas que proporcionan Arp, Smith y Spear (2015), se proveen definiciones para todos los términos; es decir, una mención del conjunto de condiciones necesarias y suficientes para describir el término. Al mencionar las condiciones se usan términos más fáciles de comprender y más simples desde el punto de vista lógico que el término a definir. Se establecen características esenciales (aquellos elementos que todas las instancias del universal deben poseer, sin los cuales una cosa deja de ser el tipo de cosa que es). Se usan definiciones aristotélicas,  $E = \text{def. un } G \text{ que } Ds$ , donde  $G$  (género) es el orden inmediatamente superior de  $E$  (especie) en la ontología y  $D$  (diferencia) indica la distinción que justifica la existencia de la especie dentro del género y se enmarca el término definido dentro de la jerarquía *is\_a* (la taxonomía provee el elemento inicial para formular la definición). Siempre que sea posible, los términos usados para manifestar las diferencias se extraen de alguna ontología donde ellos mismos son definidos. Cada definición, al descomponerse, refiere al nodo raíz de la ontología a la que pertenece, eliminándose automáticamente la circularidad de las definiciones. Todos los nodos en el grafo se vinculan al nodo raíz mediante una única cadena de relaciones *is\_a*. Se recurre a una ontología de nivel más alto, para definir el término raíz y se definen los términos desde los niveles superiores de la ontología hacia los niveles inferiores. Las definiciones son sustituibles por los términos definidos sin cambios en el significado, con la finalidad de preservar el criterio de verdad a través de los procesos de inferencia automática.

### *Integración con ontologías existentes*

El valor agregado de la ontología depende en gran parte de la definición formal de los conceptos, ya que ello asegura un uso consistente de la misma. Para cumplimentar este requisito se recurre a BFO, ontología de nivel superior que tiene en

cuenta las clases (universales) vs. las instancias (particulares), las jerarquías (es un, es parte de, tiene una parte), los solapamientos y las dependencias (es inherente a, es portador de). La integración con otras ontologías, clave de la Web Semántica, se expresa mediante propiedades de amplio uso tales como owl : equivalentClass, rdfs : subclassOf, owl : equivalentProperty, rdfs : subPropertyOf, owl : sameAs, skos : broader, skos : narrower.

También se tiene en cuenta, la IAO, como ontología de entidades de información, que constituyen un subtipo dentro de las entidades de BFO (continuos genéricamente dependientes). Representa los materiales portadores de información (libros, discos duros, fotografías, etc.), las entidades de información en sí mismas, los procesos que involucra la producción y el uso de entidades que contienen información, incluso, las relaciones entre todos ellos (*is\_about, denotes, is\_translation\_of, etc.*). Se consultan, asimismo, ontologías bibliográficas, tales como: *Bibliographic ontology* <<http://bibliontology.com>>; CiTO (*Citation Typing Ontology*) <<https://sparontologies.github.io/cito/current/cito.html>>; *RDA/ONIX framework* <<http://www.rdaregistry.info/rgAbout/rof.html>>; y FaBIO <<http://purl.org/spar/fabio>>. Esta última provee un vocabulario para representar registros bibliográficos en la Web Semántica de acuerdo con los principios de FRBR y FRBRoo que expresa las relaciones FRBR según el modelo CIDOC-CRM.

### Implementación

Se elige el juego de caracteres Unicode dado que RDF se halla codificado como texto. Se adoptan como identificadores: URIs desreferenciables mediante el uso de “hash URIs”, en principio, y de “303 URIs”, de ser necesario; URIs opacos; URNs (*Uniform Resource Names*) en la medida que son identificadores independientes de localización. En la nominación de clases y propiedades se usan CURIEs (URIs compacto) para señalar en forma abreviada que un objeto proviene de una misma ontología (foaf:Agent; foaf:PersonalProfileDocument; foaf:isPrimaryTopicOf). Para la serialización se prefiere la sintaxis XML (*Extensible Markup Language*) en la codificación; es decir, serializar los triples RDF en un formato conocido como RDF/XML. Para garantizar la correcta resolución del nivel lógico, se validan relaciones y propiedades mediante el motor de razonamiento de *Protégé*. Además, se evalúa el uso de W3C RDF Validator, o, de Apache-Jena y la consulta de la *Ontology Alignment Evaluation Initiative* (OAEI). Esta iniciativa, como ya se ha expresado, gestiona un foro que ofrece distintos abordajes a ser testeados. Se opera con el lenguaje de interrogación SPARQL (*Simple Protocol and RDF Query Language; SPARQL 1.0, SPARQL 1.1*).

### Evaluación

Para evaluar la ontología se adoptan tanto los criterios enunciados por Vrandečić (2010), como el método propuesto por Seyed y Shapiro (2012).

### Documentación

En paralelo con la implementación, se proporciona la documentación, la que se publica en RDF/XML en el sitio Web del archivo de texto que la contiene, complementándose con documentos narrativos.

### Sustentabilidad

Con el objetivo de asegurar a través del tiempo la vigencia de su funcionalidad, de su significación y de su utilidad dentro del dominio abarcado, se adhiere a una visión abierta del mundo (*Open-World Assumption, OWA*) que asume la posibilidad de

extensión y corrección constantes de la ontología, ya que su objetivo último consiste en dar soporte al trabajo de los científicos de clasificar particulares. Se adhiere, también, a la regla de objetividad, es decir, a describir lo que existe en la realidad y no lo que se conoce sobre lo que existe en la realidad.

## **Conclusión**

En función de lo expuesto en el presente trabajo, se concluye que las cuestiones metodológicas constituyen un aspecto crítico en el proceso de construcción de las ontologías. Adquieren especial relevancia a partir del paradigma de los Datos Abiertos Enlazados (LOD) que promueve la recuperación de información en entorno multilingüe y se orienta a visibilizar todas aquellas relaciones que contribuyan a la creación de conocimiento. Con respecto al diseño del modelo conceptual encarado por el proyecto UBACyT 773BA, conformar la documentación y llevar a cabo cada uno de los pasos establecidos para desarrollar las ontologías de acuerdo con las pautas metodológicas, facilita la resolución de la instancia de elaboración y, asimismo, garantiza el funcionamiento de la inferencia automática. Finalmente, se destaca también, la importancia de dar el tratamiento adecuado, en el marco de aplicación de la metodología, a los aspectos lógicos y lingüísticos.

## Referencias bibliográficas

- » Aguado-de-Cea, Guadalupe; Elena Montiel-Ponsoda; María Poveda-Villalón y Olga Giraldo-Pasmin. Ximena. 2015. Lexicalizing Ontologies: The Issues Behind The Labels. En *Procedia. Social and Behavioral Sciences*. No. 212, 151-158. <[https://www.researchgate.net/publication/286541385\\_Lexicalizing\\_Ontologies\\_The\\_Issues\\_Behind\\_the\\_Labels](https://www.researchgate.net/publication/286541385_Lexicalizing_Ontologies_The_Issues_Behind_the_Labels)> [Consulta: 20 abril 2017].
- » Allemang, Dean y James Hendler. 2011. *Semantic web for the working ontologist: Effective modeling in RDFS and OWL*. Burlington, MA: Morgan Kaufmann.
- » Alvarado, Rubén Darío. 2010. *Metodología para el desarrollo de ontologías*. <<https://es.slideshare.net/lceman1976/metodologia-para-ontologias>> [Consulta: 30 marzo 2016].
- » Arp, Robert; Barry Smith y Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. Cambridge, MA: The MIT Press.
- » Bautista-Zambrana, María Rosario. 2015a. Creating corpus-based ontologies: a proposal for preparatory work. En *Procedia - Social and Behavioral Sciences*. No. 212, 159-165.
- » Bautista-Zambrana, María Rosario. 2015b. Methodologies to build ontologies for terminological purposes. En *Procedia - Social and Behavioral Sciences*. No. 173, 264-269.
- » Beckett, David. 2014. *A line-based syntax for an RDF graph*. <<http://www.w3.org/TR/n-triples/>> [Consulta: 20 abril 2017].
- » Beckett, David; Tim Berners-Lee; Eric Prud'hommeaux y Gavin Carothers. 2014. *Terse RDF Triple Language*. <<http://www.w3.org/TR/turtle/>> [Consulta: 10 marzo 2016].
- » Berners-Lee, Tim. 1998. *Cool URIs don't change*. <<https://www.w3.org/Provider/Style/URI.html.en>> [Consulta: 5 mayo 2016].
- » Berners-Lee, Tim. 2006. *Linked data*. <<http://www.w3.org/DesignIssues/LinkedData.html>> [Consulta: 5 mayo 2016].
- » Bizer, Chris; Richard Cyganiak y Tom Heath. 2008. *How to Publish Linked Data on the Web*. <<http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedData-Tutorial/>> [Consulta: 13 marzo 2017].
- » Bonifati, Angela; Radu Ciucanu y Aurélin Lemay. 2015. Learning Path Queries on Graph Databases. Trabajo presentado al *18th International Conference on Extending Database Technology (EDBT), Mar 2015, Bruxelles, Belgium*. <<http://edbticdt2015.be/>> [Consulta: 20 marzo 2017].
- » Cimiano, Philipp; et al. 2013. Multilingual Question Answering over Linked Data (QALD-3): lab overview. En P. Forner; H. Müller; R. Paredes; P. Rosso y B. Stein, eds. *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*. Berlin, Heidelberg: Springer. p. 321-332.
- » Cimiano, Philipp; Elena Montiel-Ponsoda; Paul Buitelaar; Mauricio Espinoza y Asunción Gómez-Pérez. 2010. A Note on Ontology Localization. En *Applied Ontology*, No. 0, 1-9. (127-137). <[http://oa.upm.es/5178/1/A\\_Note\\_on\\_Ontology\\_Localization.pdf](http://oa.upm.es/5178/1/A_Note_on_Ontology_Localization.pdf)> [Consulta: 21 mayo 2017].

- » Cimino, James. J. 1998. Desiderata for Controlled Medical Vocabularies in the TwentyFirst Century. En *Methods of Information in Medicine*. Vol. 37, no. 4-5, 394-403. <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415631/pdf/nihms-396702.pdf>> [Consulta: 23 abril 2016].
- » Davis, Iam; Thomas Steiner y Arnaud J. Le Hors. 2013. *JSON Alternate Serialization (RDF/JSON)*. <<https://www.w3.org/TR/rdf-json/>> [Consulta: 21 mayo 2017].
- » *Designing URI Sets for the UK Public Sector: A report from the Public Sector Information Domain of the CTO Council's cross-Government Enterprise Architecture* 2009 October. Version 1.0. 2009. London: Chief Technology Officer Council.
- » Dhingra, Vandana y Komal Kumar Bhatia. 2015. Development of Ontology in Laptop Domain for Knowledge Representation. En *Procedia Computer Science*. No. 46, 249-256.
- » Espinoza Mejía, Mauricio; Elena Montiel-Ponsoda; Guadalupe Aguado de Cea y Asunción Gómez-Pérez. 2012. Ontology localization. En Suárez-Figueroa, Mari Carmen; Asunción Gómez-Pérez; Erico Motta y Aldo Gange-mi, eds. *Ontology Engineering in a Networked World*. Berlin-Heidelberg: Springer. p 171-191.
- » Fernández-López, Mariana; Asunción Gómez-Pérez y Natalia Juristo. 1997. Methontology: From ontological art towards ontological engineering. En Farquhar, A. y M. Grüninger, eds. *Ontological Engineering: Papers from the AAAI Spring Symposium*. Palo Alto, CA: AAAI Press. p. 33-04.
- » Fliedl Günder; Chistian Kop y Jünger Vöhringer. 2007. From OWL Class and Property Labels to Human Understandable Natural Language. En Kedad Zubida; Lamari Nadira; Elisabeth Métais; Farid Meziane y Yacine Rezugui, eds. *Natural Language Processing and Information Systems*. NLDB. Berlin: Springer.
- » Fonseca Marangon, Jose; et al. 2016. *ontoAGA: Ontology to Support Educational Systems Interoperability*. ONTOBRAS. <<https://www.semanticscholar.org/paper/ontoAGA%3A-Ontology-to-Support-Educational-Systems-Marangon-Campos/fee1d4cfd912ed176f373e8552eeaf3b779f85fb>> [Consulta: 20 octubre 2017].
- » García, Jorge; et al. 2012. Challenges for the multilingual web of data. En *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*. No. 11, 63-71. <http://doi.org/10.1016/j.websem.2011.09.001>
- » Gruber, Thomas. R. 1993a. A translation approach to portable ontology specifications. En *Knowledge Acquisition*. Vol. 5, no. 2, 199-220.
- » Gruber, Thomas. R. 1993b. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. En Guarino, N. y R. Poli, eds. *Formal Ontology in Conceptual Analysis and Knowledge Representation*. London: Kluwer Academic Publishers <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.43.6200&rep=rep1&type=pdf>> [Consulta: 30 agosto 2016].
- » Grüninger, Michael y Mark. S. Fox. 1995. Methodology for the design and evaluation of ontologies. En *Workshop on Basic Ontological Issues in Knowledge Sharing: IJCAI-95*. Montreal: IJCAI. <<http://www.eil.utoronto.ca/wp-content/uploads/enterprise-modelling/papers/gruninger-ijcai95.pdf>> [Consulta: 15 mayo 2017].
- » Guarino, Nicola. 1998. Formal Ontology and Information Systems. En Guarino, Nicola., ed. *Formal Ontology in Information Systems. Proceedings of FOIS'98 (6-8 June 1998: Trento, Italy)*. Trabajos presentados. Amsterdam, IOS Press. p. 3-15.
- » Heath, Tom y Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Florida: Morgan & Claypool.

- » Hedden, Heather. 2016. *The accidental taxonomist*. 2a. ed. Medford, NJ: Information Today.
- » IFLA. FRBR Review Group. 2011. *Functional requirements: the FRBR family of models*. <<http://www.ifla.org/node/2016>> [Consulta: 15 mayo 2017].
- » Kanoh, Hiroko; Kouji Kozaki; Motohiro Hasegawa y Takaaki Hishida. 2015. Development of ontology for information literacy. En *Procedia Computer Science*. No. 60, 170-177.
- » Klyne, Graham y Jeremy J. Carroll. 2014 [2004]. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. <<https://www.w3.org/TR/rdf-concepts/>> [Consulta: 14 septiembre 2016].
- » Köhler, Jacob; Katherine Munn; Alexander Rüegg; Andre Skusa y Barry Smith. 2006. Quality control for terms and definitions in ontologies and taxonomies. En *BMC Bioinformatics*. 7, 212. <http://doi.org/10.1186/1471-2105-7-212>.
- » Kyriaki-Manessi, Daphne y Markos Dendrinou. 2014. Developing Ontology for the University Archives: The Domain of Technological Education. En *Procedia. Social and Behavioral Sciences*. No. 147, 349-359. <http://doi.org/10.1016/j.sbspro.2014.07.111>
- » López Rodríguez, Yoan A.; Yusniel Hidalgo Delgado y Nemury Silega Martínez. 2016. Método para la integración de ontologías en un sistema para la evaluación de créditos. En *Revista Cubana de Ciencias Informáticas*. Vol. 10, no. 4, 97-111.
- » Lozano Rosch, Elena. 2007. *Portal de colaboración con capacidades semánticas*. Sevilla: Universidad de Sevilla. <<https://www.rediris.es/ptyoc/res/store/dl18/ptyoc.pdf>> [Consulta: 14 agosto 2016].
- » Malik, Shaily. K.; Anisha Goel y Saurabh Maniktala. 2010. A comparative study of various variants of SPARQL in semantic web. En *Computer Information Systems and Industrial Management Applications (CISIM): 2010 International Conference*. New York: IEEE. p. 471-474.
- » Malik, Sanjay Kumar; Nupur Prakash y SAM Rizvi. 2011. Ontology Creation towards an Intelligent Web: Some Key Issues Revisited. En *International Journal of Engineering and Technology*. Vol. 3, no. 1, 44-52.
- » Malik, Sanjay Kumar; Nupur, Prakash y SAM Rizvi. [s. f.]. *Ontology Design and Development: Some aspects: An overview*. <<https://pdfs.semanticscholar.org/b285/4360e68127e0ae82d1631b6c477bc4801b5d.pdf>> [Consulta: 15 mayo 2017].
- » Malviya, Naveen; Nishchol Mishra y Santosh Sahu. 2011. Developing University Ontology using protégé OWL Tool: Process and Reasoning. En *International Journal of Scientific & Engineering Research*. Vol. 2, no. 9, 1-8.
- » Montiel-Ponsoda, Elena, et al. 2011. Style Guidelines for Naming and Labeling Ontologies in the Multilingual Web. En *International Conference on Dublin Core and Metadata Applications (Dublin)* <<http://dcevents.dublincore.org/IntConf/dc-2011/paper/view/47/15>> [Consulta: 30 junio 2016].
- » Noy, Natalya. F. y Deborah. L. McGuinness. 2001. *Ontology development 101: A Guide to Creating Your First Ontology*. <[http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf)> [Consulta: 10 agosto 2016].
- » *Ontology Alignment Evaluation Initiative*. 2018. <<http://oaei.ontologymatching.org/>> [Consulta: 20 marzo de 2018].
- » Rattanasawad, Thanyalak; Kanda Runapongsa Saikaew; Marut Buranarach y Thepchai Supnithi. 2013. A review and comparison of rule languages and rule-

- based inference engines for the semantic web. En *International Computer Science and Engineering Conference ICSEC (2013)*. Trabajos presentados. New York: IEEE, p. 1-6.
- » Sauermann, Leo y Richard Cyganiak. 2008. *Cool URIs for the Semantic Web*. <<https://www.w3.org/TR/cooluris/>> [Consulta: 16 junio 2016].
  - » Sauermann, Leo; Richard Cyganiak y Max Völkel. 2006. *Cool URIs for the Semantic Web*. <[https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25142/1/TM\\_07\\_01.pdf](https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25142/1/TM_07_01.pdf)> [Consulta: 30 marzo 2017].
  - » Schober, Daniel; et al. 2009. Survey-based naming conventions for use in OBO-Foundry ontology development. En *BMC Bioinformatics*. Vol. 10, no. 125. <https://doi.org/10.1186/1471-2105-10-125>
  - » Seyed, A. Patrice y Stuart C. Shapiro. 2012. *A Method for Evaluating Ontologies. Introducing the BFO-Rigidity Decision Tree Wizard*. FOIS. DOI: 10.3233/978-1-61499-084-0-191
  - » Shabolt, Nigel R. y Paul R. Smart. 2015. *Knowledge elicitation: Methods, tools, and techniques*. <<https://eprints.soton.ac.uk/359638/1/Knowledge%2520Elicitationv7.pdf>> [Consulta: 3 febrero 2017].
  - » Smith, Barry; Walcaw Kusnierczyk; Daniel Schober y Werner Ceusters. 2006. Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. En Bodenreider, O., ed. *Proceedings of the 2nd International Workshop on Formal Biomedical Knowledge Representation (2nd; 2006)*. Trabajos presentados. Baltimore, MD: KR-MED Publications, vol. 222. p. 57-66. <<http://www.informatik.unitrier.de/~ley/db/conf/krmed/krmed2006.html>> [Consulta: 10 agosto 2016].
  - » Stevens, Robert y Duncan Hull. 2010. *Separating Concepts from Labels* [Mensaje de blog]. <<http://ontogenesis.knowledgeblogger.org/818>> [Consulta: 13 mayo 2017].
  - » Stuart, David. 2016. *Practical ontologies for information professionals*. London: Facet Publishing.
  - » Swanson, Don R. 1986. Undiscovered public knowledge. En *Library Quarterly*. Vol. 56, no. 2, p. 103-118. <<http://www.sciencengines.com/NPHS/Documents/UndiscoveredPublicKnowledge.pdf>> [Consulta: 14 Agosto 2017].
  - » Swanson, Don R. y Neil R. Smalheiser, 1996. Undiscovered public knowledge: A ten-year update. En *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD-96, AAAI*. <<https://aaai.org/Papers/KDD/1996/KDD96-051.pdf>> [Consulta: 14 agosto 2017].
  - » Théreaux, Oliver. 2003. *Common http implementation problems: W3C note*. <<https://www.w3.org/TR/chips/>> [Consulta: 16 mayo 2016].
  - » Tonkin, Emma; Heather. D. Pfeiffer y Andrew Hewson. 2010. An evidence-based approach to collaborative ontology development. En *Workshop on Matching and Meaning (2010)*. Trabajos presentados. <<http://opus.bath.ac.uk/18033/1/tonkinpfeiffer.pdf>> [Consulta: 19 junio 2017].
  - » UNE-ISO 1087-1. 2009. *Trabajos terminológicos. Teoría y aplicación*. AENOR: Madrid.
  - » UNE-ISO 1087-2. 2009 *Trabajos terminológicos. Aplicaciones informáticas*. AENOR: Madrid.

- » Uschold, Mike y Martín King. 1995. Towards a methodology for building ontologies. En *Workshop on Basic Ontological Issues in Knowledge Sharing*. <<http://www.aiai.ed.ac.uk/publications/documents/1995/95-ont-ijcai95-ont-method.pdf>> [Consulta: 10 marzo 2016].
- » Van Hooland, Seth y Ruben Verborgh. 2014. *Linked data for libraries, archives and museums: How to clean, link and publish your metadata*. London: Facet Publishing.
- » Vrandečić, Denny. 2010. *Ontology evaluation*. <[http://simia.net/download/ontology\\_evaluation.pdf](http://simia.net/download/ontology_evaluation.pdf)> [Consulta: 10 marzo 2017].
- » Willer, Mirna; Gordon Dunsire y Boris Bosančić. 2010. ISBD and the Semantic Web. En *Italian Journal of Library, Archives and Information Science*. Vol. 1, No. 2, 213-236. <http://dx.doi.org/10.4403/jlis.it-4536>

