



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA - ESCUELA DE LETRAS - BIBLIOTECAS UC

# **INTEGRACIÓN DE TECNOLOGÍAS DE LA WEB SEMÁNTICA EN EL BUSCADOR DE PATENTES DE INAPI**

**DANILO ANDRÉS REYES LILLO**

Actividad de graduación para optar al grado de  
Magister en Procesamiento y Gestión de la Información

PROFESOR SUPERVISOR: CLAUDIA GUTIERREZ

Santiago de Chile, Julio 2017

© MMXVII, DANILLO REYES



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA - ESCUELA DE LETRAS - BIBLIOTECAS UC

# **INTEGRACIÓN DE TECNOLOGÍAS DE LA WEB SEMÁNTICA EN EL BUSCADOR DE PATENTES DE INAPI**

**DANILO ANDRÉS REYES LILLO**

Miembros del Comité:

PROFESOR SUPERVISOR: CLAUDIA GUTIERREZ

PROFESOR INVITADO: CÉSAR AGUILAR

PROFESOR INVITADO: DENIS PARRA

DIRECTOR PROGRAMA: MAURICIO ARRIAGADA BENÍTEZ

Actividad de graduación para optar al grado de  
Magister en Procesamiento y Gestión de la Información

Santiago de Chile, Julio 2017

© MMXVII, DANILO REYES

*A quienes siempre me dan su apoyo  
y amor para lograr mis objetivos.*

## **AGRADECIMIENTOS**

Agradezco a Claudia Gutiérrez por su guía, no solo en esta investigación sino en el plano del desarrollo profesional y personal.

Agradezco a Mauricio Arriagada, Director del Magister en Procesamiento y Gestión de la Información, por su colaboración y compromiso para que esta investigación llegue a buen puerto.

Agradezco a Rafael Castillo por una de sus grandes virtudes: compartir desinteresadamente lo mucho que sabe.

Agradezco a Sergio Fredes, quien con su manejo técnico me brindó una importante colaboración y de quien siempre aprendo algo nuevo.

A todos y todas quienes de una u otra manera son parte de esto.

## INDICE GENERAL

AGRADECIMIENTOS . . . . .	IV
INDICE DE FIGURAS . . . . .	VII
INDICE DE TABLAS . . . . .	IX
RESUMEN . . . . .	X
ABSTRACT . . . . .	XI
1. INTRODUCCIÓN . . . . .	1
2. ANTECEDENTES GENERALES . . . . .	4
2.1. Las patentes y su relevancia para la I+D+i . . . . .	4
2.2. Bases de datos de patentes . . . . .	7
2.3. Contextualización y formulación del Problema . . . . .	9
2.4. Objetivos . . . . .	11
2.4.1. Objetivo General . . . . .	11
2.4.2. Objetivos Específicos . . . . .	11
2.5. Metodología . . . . .	11
2.5.1. Análisis de la base de datos de patentes de INAPI . . . . .	11
2.5.2. Definición de Ontología(s) reusables para el modelo de datos . . . . .	12
2.5.3. Modelamiento de datos (R2RML) . . . . .	12
2.5.4. Definición y publicación de Tripletas en RDF . . . . .	12
2.5.5. Descripción de usos y aplicaciones de datos en RDF . . . . .	12
2.5.6. Resultados esperados . . . . .	12
3. ESTADO DEL ARTE . . . . .	14
3.1. Las patentes como fuente de información . . . . .	14
3.2. Aproximaciones a la web semántica y linked data . . . . .	21
3.2.1. Uniform Resource Identifier (URI) . . . . .	24

3.2.2.	Ontologías . . . . .	26
3.2.3.	RDF . . . . .	27
3.3.	Integración de tecnologías de la web semántica y Linked Data (LD) . . . . .	31
3.4.	Integración de tecnologías de la web semántica en propiedad industrial . . . . .	36
4.	METODOLOGÍA - INTEGRACIÓN DE TECNOLOGÍAS SEMÁNTICAS . . . . .	39
4.1.	Análisis de bases de datos de INAPI . . . . .	39
4.1.1.	Análisis del modelo de datos de INAPI . . . . .	40
4.1.2.	Levantamiento de base de datos en MySQL . . . . .	42
4.2.	Definición de Ontologías . . . . .	44
4.3.	Mapping / Modelamiento de datos . . . . .	49
4.3.1.	Input o integración de ontología y fuente de datos. . . . .	52
4.3.2.	Definición de URIs . . . . .	53
4.3.3.	Vinculación de clases de la ontología con la fuente de datos . . . . .	57
4.3.4.	Construcción del grafo . . . . .	60
4.3.5.	Refinamiento del modelo . . . . .	62
4.3.6.	Publicación del modelo . . . . .	63
4.4.	Definición de tripletas . . . . .	65
4.5.	APLICACIONES . . . . .	69
4.5.1.	Optimización de búsqueda de patentes . . . . .	70
4.5.2.	Vinculación de datos con otras fuentes. . . . .	71
4.5.3.	Aplicación de estándares de datos abiertos y enlazados (LOD). . . . .	72
5.	CONCLUSIONES . . . . .	75
	BIBLIOGRAFIA . . . . .	80
	ANEXO A. POTENCIALES DATASETS ENLAZABLES CON INAPI . . . . .	84

## INDICE DE FIGURAS

2.1. Buscador de Patentes. . . . .	9
2.2. Buscador de Patentes de Dominio Público. . . . .	10
3.1. Documento de Patente. . . . .	17
3.2. Registro bibliográfico de un documento de Patente. . . . .	20
3.3. Arquitectura de la web semántica. . . . .	23
3.4. URI, recurso y representación. . . . .	25
3.5. Tripletas en RDF (URI-Propiedad-Literal). . . . .	29
3.6. Tripletas en RDF (URI-Propiedad-URI). . . . .	30
3.7. Grafo en RDF. . . . .	30
3.8. Representación textual de tripletas en RDF. . . . .	31
3.9. Diferencias entre OAI-PMH y Linked Data. . . . .	33
3.10. Transformación vs Mapping. . . . .	34
4.1. Diseño metodológico de Integración Semántica. . . . .	39
4.2. Modelo de datos de la base de datos de INAPI. . . . .	40
4.3. Réplica de base de datos en MySQL. . . . .	42
4.4. Descripción de elementos de la ontología de AKSW (Class, object property y data property). . . . .	46
4.5. Visualización gráfica de la ontología de AKSW. . . . .	48
4.6. Propiedades de objeto (Object property) de la clase Country. . . . .	49
4.7. Proceso de Karma para modelar fuentes de datos. . . . .	51
4.8. Importación de base de datos a Karma. . . . .	53
4.9. Uri de documento de patente. . . . .	54
4.10. Uri del inventor. . . . .	54

4.11. Uri del propietario. . . . .	55
4.12. Relación de URIs con otras fuentes. . . . .	56
4.13. Potencial relación entre datos de una patente con DBPedia (Ejemplo de Codelco). . . . .	56
4.14. Interfaz para establecer clase, propiedad y tipo de literal. . . . .	59
4.15. Interfaz de Karma para elaborar el grafo. . . . .	61
4.16. Construcción del grafo en Karma. . . . .	62
4.17. Herramientas de refinamiento de Karma. . . . .	63
4.18. Publicación de Modelo R2RML. . . . .	64
4.19. Registro de base de datos de INAPI y su traducción a un registro en RDF. . . . .	66
4.20. Registro en RDF de una patente. . . . .	66
4.21. Tripletas de una patente y su título. . . . .	67
4.22. Tripletas de una patente y su inventor. . . . .	68
4.23. Tripletas de una patente y su solicitante. . . . .	68
4.24. Grafo de relaciones Inventor – Patente - Solicitante. . . . .	69
4.25. Datos potencialmente obtenidos de DBPedia mediante datos vinculados. . . . .	72
4.26. Modelo de 5 estrellas de datos abiertos y enlazados. . . . .	74
5.1. Modelo de traducción de base de datos a RDF. . . . .	77

## INDICE DE TABLAS

2.1. Principales bases de datos de patentes del mundo . . . . .	8
4.1. Tabla de clases, propiedades y tipos de literales . . . . .	60
A.1. Datasets potencialmente enlazables con INAPI . . . . .	84

## RESUMEN

El presente estudio aborda la integración de tecnologías de la web semántica en el contexto de las patentes, específicamente en el buscador de patentes del Instituto Nacional de Propiedad Industrial (INAPI). La creciente integración de tecnologías basadas en datos enlazados a nivel mundial, la escasa incorporación de éstas en el contexto de la propiedad industrial y su gran potencial para fortalecer la búsqueda y recuperación de información tecnológica hacen necesaria la integración de servicios semánticos en el ámbito de las patentes.

En primera instancia, esta investigación efectúa un estudio bibliográfico acerca del estado del arte en bases de datos de patentes, el desarrollo de tecnologías de la web semántica y su incorporación al ámbito de la propiedad industrial.

En segunda instancia, alineado con el principal objetivo de este estudio, se analizan los elementos fundamentales para determinar una estrategia procedimental para la incorporación de tecnologías de la web semántica en los servicios de búsqueda de información de patentes de INAPI. Lo anterior se plantea desde una metodología basada en el análisis del modelo de datos existente, la definición de una ontología usando el software Protégé, el modelamiento de los datos para traducir la base de datos relacional a un modelo semántico en RDF usando el software Karma y la publicación de tripletas en RDF en el ámbito de la propiedad industrial para plasmar un portal de datos enlazados de patentes (datos.inapi.cl).

Finalmente, el estudio plantea una serie de aplicaciones que ofrecen las tecnologías de la web semántica y linked data, las cuales potencialmente podrían ser aplicadas a los servicios de INAPI para fortalecer la búsqueda y recuperación de información de patentes.

**Palabras Claves:** Web Semántica, Linked Data, Propiedad Industrial, Patentes de Invención, Búsqueda semántica de Patentes.

## ABSTRACT

This study addresses the integration of semantic web technologies in the context of patents, specially on the INAPI patents browser. The world wide increasing integration of technologies based on linked data, the limited integration of these technologies into industrial property context and its great potencial to reinforce the technological information search and retrieval makes it necessary the integration of semantic services in patents context.

In the first instance, this study carried out a bibliographic review about state of the art on patent databases, web semantic technologies development, and its incorporation to the industrial property field.

Secondly, in line with the main objective of this study, fundamental elements are analysed to determine a procedural strategy to incorporate semantic web technologies into INAPI patents information browser. This incorporation is approached from a methodology based on the analysis of the existent data model, an ontology definition using the software Protegé, the data modeling to translate the relational data base into a semantic model on RDF using the software Karma and the RDF triple publication in the context of industrial property to implement a patents linked data portal (datos.inapi.cl).

Finally, this study raises a number of applications of web semantic and linked data technologies, which could potentially be applied to INAPI services, to reinforce the patents information search and retrieval.

**Keywords:** Semantic Web, Linked Data, Industrial Property, Patents, Semantic Patent Search.



## 1. INTRODUCCIÓN

El crecimiento en el desarrollo e integración de tecnologías de la web semántica en distintos servicios digitales es una realidad, incluso para muchos organismos e instituciones de gobierno a nivel mundial. Por otra parte, el interés de distintos actores como empresas, universidades, instituciones gubernamentales o centros de investigación por el mundo de la propiedad intelectual y las patentes también ha sido creciente debido a que son fuentes de información fundamentales para promover la I+D+i. Sin embargo, si bien la Organización Mundial de Propiedad Intelectual (OMPI)<sup>1</sup> declara identificar algunos servicios semánticos en bases de datos de patentes (OMPI, 2016), estos no están relacionados con el contexto de la web semántica. Adicionalmente, si bien las investigaciones científicas han empezado a incursionar este potencial nexo entre la web semántica y las patentes, en general, no han sido llevadas a un plano más aplicado a servicios concretos.

Dentro de las principales iniciativas elaboradas en este plano, es posible destacar a Go-Patents, un prototipo de búsqueda de patentes apoyado en el uso de ontologías de diversos dominios para potenciar la relación entre los datos (Eisinger, Mönnich, y Schroeder, 2014). De igual manera, la iniciativa NEST ("Newly Emerging Science and Technology") ha incorporado tecnologías de la web semántica para apoyar la identificación de innovaciones tecnológicas (Zhang, Zhou, Porter, y Gomila, 2014). Asimismo, el grupo de investigación AKSW (Agile Knowledge Engineering and Semantic Web) impulsaron una iniciativa denominada USPatents, cuyo objetivo fue publicar y enlazar datos de la Oficina de Patentes y Marcas de Estados Unidos (USPTO), generando los registros de las patentes en RDF, en base a una ontología desarrollada a partir de la estructura de datos de la USPTO. Dicha ontología resulta uno de los insumos más relevantes para esta investigación.

A pesar de las iniciativas descritas, el desarrollo o la integración de tecnologías semánticas no ha sido una tendencia con el crecimiento necesario en el ámbito de la propiedad

---

<sup>1</sup>La Organización Mundial de la Propiedad Intelectual (OMPI) es el organismo mundial de las Naciones Unidas, que cuenta con 189 Estados miembros y se encarga de los servicios, políticas, cooperación e información en materia de Propiedad Intelectual.

industrial, la búsqueda y recuperación de patentes, aún considerando su tremendo potencial para fortalecer estos servicios. Dicho potencial se plasma en la posibilidad que ofrecen las tecnologías de la web semántica para mejorar la organización de la información en base a ontologías, enriquecer los datos y promover su reutilización por parte de empresas u organizaciones que requieran de su uso (García Moreno, 2015).

Por otra parte, una gran parte de los datos disponibles que utilizan los diversos sitios web están albergados en bases de datos relacionales (Hert, Reif, y Gall, 2011) que pueden ofrecer servicios de consultas, pero no están potenciados con vocabularios apropiados para un contexto de la web semántica y linked data como ontologías o RDF. Lo anterior no permite que estos datos sean comprendidos por las máquinas y, por lo tanto, no se pueden enlazar con otras fuentes de datos para proveer de contexto y nutrir semánticamente dichos datos.

Bajo la perspectiva previamente expuesta, el presente estudio tiene por objetivo determinar una estrategia procedimental para la integración de tecnologías de la web semántica a la base de datos de patentes del Instituto Nacional de Propiedad Industrial (INAPI). La finalidad de lo anterior es fortalecer la búsqueda y recuperación de información de patentes, junto con proporcionar a la base de datos la posibilidad de enlazar y ser enlazada por otras fuentes de datos que hagan más potentes sus registros, a través de un portal de datos enlazados ([datos.inapi.cl](http://datos.inapi.cl)). Lo anterior se transforma en una herramienta que permite un mejor acceso y comprensión de la información de las patentes para las empresas, centros de investigación, universidades, inventores y otros actores que intervienen en la investigación, desarrollo e innovación (I+D+i). Asimismo, este insumo resulta fundamental para fortalecer la gestión de la innovación mediante el uso del conocimiento descrito en los documentos de patentes.

De esta manera, este trabajo plantea esta nexos entre bases de datos de patentes (fundamentalmente relacionales) y tecnologías de la web semántica, planteando un mecanismo de traducción y se fundamenta en cuatro apartados fundamentales:

- **Antecedentes generales:** se plantean los principales antecedentes de la investigación, el problema a tratar, los objetivos, la metodología y los resultados esperados.
- **Estado del arte:** expone una revisión documental sobre los principales elementos abordados en esta investigación, dividiendo el apartado en 1) Las patentes como fuentes de información, 2) Aproximaciones a la web semántica y Linked Data, 3) Integración de tecnologías de la web semántica y Linked Data y, finalmente, 4) Integración de tecnologías de la web semántica en propiedad industrial.
- **Metodología - Integración de tecnologías semánticas:** se plantea una metodología para traducir una base de datos relacional de patentes a tripletas en RDF describiendo cinco pasos fundamentales: Análisis de la base de datos de INAPI, definición de ontologías, mapping / modelamiento de datos, definición de tripletas y aplicaciones.
- **Conclusiones:** finalmente, se establecen las principales conclusiones que se desprenden de la revisión bibliográfica y la aplicación de la metodología.

Cabe destacar que el presente estudio busca abrir e incentivar nuevas investigaciones en el campo de las patentes y la web semántica en Chile, junto con ofrecer una estrategia procedimental de traducción de bases de datos relacionales a RDF que potencialmente podría ser aplicada a otros contextos.

## **2. ANTECEDENTES GENERALES**

### **2.1. Las patentes y su relevancia para la I+D+i**

Desde hace algunos años, la Investigación, el Desarrollo y la Innovación (I+D+i) están resultando un pilar fundamental tanto para el Estado, las universidades, los centros de investigación y las empresas (Etzkowitz y Leydesdorff, 2000). En este contexto, la propiedad intelectual se transforma en un elemento sustancial para las personas y organizaciones que buscan generar nuevo conocimiento o nuevos desarrollos tecnológicos.

La propiedad intelectual es un cuerpo legal que se refiere a la protección de todas las creaciones del intelecto humano. En el Convenio que establece la creación de la Organización Mundial de Propiedad Intelectual (OMPI), en el año 1967, se indica la materia que puede ser susceptible a protección mediante derechos de propiedad intelectual:

- Obras literarias, artísticas y científicas.
- Interpretaciones y ejecuciones de obras, junto con fonogramas.
- Invenciones de cualquier campo de la actividad humana.
- Descubrimientos científicos.
- Diseños industriales.
- Marcas y denominaciones comerciales y no comerciales.
- Protección contra competencia desleal.
- Todos los derechos relativos a la actividad intelectual del hombre, en materia industrial, artística, literaria y/o científica.

En términos concretos, el derecho de propiedad intelectual fundamenta dos razones para explicar la necesidad de los países para promulgar leyes relacionadas con protección intelectual: Por un lado, amparar a los creadores, artistas e inventores con derechos en relación a sus creaciones, equilibrando estos derechos con el interés público de poder acceder a dichas creaciones o invenciones; por otro lado, fomentar la creatividad y la innovación para fortalecer el desarrollo social y económico de una nación.

Asimismo, la propiedad intelectual se divide en dos ramas: el derecho de autor o “copyright” y la propiedad industrial. En síntesis, el derecho de autor se encarga de la protección de obras artísticas y literarias como libros, obras musicales, pinturas, entre otras manifestaciones artísticas; la propiedad industrial, por su parte, protege invenciones de aplicación industrial en base a patentes, modelos de utilidad, diseños industriales, topografías de circuitos integrados, indicaciones geográficas y denominaciones de origen (OMPI, 2016).

Si bien, ambas ramas de la propiedad intelectual resultan fundamentales para la promoción de la I+D+i, es en el ámbito de la propiedad industrial donde las patentes de invención toman especial relevancia, debido a que el sistema de patentes tiene como finalidad incentivar la innovación y el crecimiento económico mediante dos mecanismos:

- Entregando protección de la creatividad y una recompensa para los inventores a raíz de las fuertes inversiones efectuadas en una nueva invención. Las patentes protegen por un período de tiempo determinado y de manera territorial los derechos del inventor sobre su desarrollo.
- Publicando y divulgando toda la información técnica referida a una nueva invención a través de un documento de patente. Dicho documento se publica luego de un plazo de tiempo determinado, una vez presentada la solicitud y da cuenta de todos los detalles técnicos de la invención para que la comunidad lo pueda consultar (OMPI, 2013).

Complementando lo anterior, es preciso señalar que la protección legal que otorga una patente es de carácter territorial, es decir, se limita a un determinado país o grupo de países. Sin embargo, la información técnica que divulga la patente se visibiliza de manera mundial, es decir, se puede consultar desde cualquier lugar del mundo. Si la patente no está protegida en un país, se considera de dominio público.

Las patentes son importantes fuentes de información técnica, legal y de negocios y son un recurso fundamental para investigadores, emprendedores e inventores. Adicionalmente, son documentos que poseen información exclusiva, por lo tanto, no ha sido publicada en artículos científicos o papers de conferencias. Se estima que alrededor del 70 u 80 por ciento

de la información contenida en una patente no ha sido publicado en ningún otro lugar (Singh, Chakraborty, y Vincent, 2016). Por otro lado, de acuerdo a la norma ST.9 de la OMPI, que define los datos bibliográficos que deben estar contenidos en un documento de patente, es que la información que disponen dichos documentos está altamente estructurada y normalizada, lo que facilita la posibilidad de hacer análisis y procesamiento masivo de patentes.

Todo lo anterior hace de las patentes un recurso de información altamente atractivo no sólo para investigadores o inventores, sino también para otros actores involucrados en el entorno de la propiedad industrial como abogados, especialistas de marketing o estrategias comerciales. Sin embargo, la naturaleza técnica y jurídica de las patentes hace muy compleja, no sólo su lectura, sino también las maneras de buscar y acceder a ellas (Adams, 2012).

De acuerdo a (Hunt, Nguyen, y Rodgers, 2012), las búsquedas de patentes se pueden tipificar, de acuerdo a la necesidad y los objetivos que busca cubrir. De esta manera, es factible identificar los siguientes tipos de búsqueda:

- **Búsqueda de patentabilidad**, se utiliza para determinar si un invento puede o no ser patentado y si se pudiera, qué otras patentes o documentos pudieran ser relevantes de analizar.
- **Búsqueda de validación**, se utiliza para determinar si un invento tiene novedad absoluta.
- **Búsqueda de infracción**, se utiliza para determinar si una patente vigente afirma lo mismo que determinado concepto o invento no patentado. Para esta búsqueda se incluyen sólo patentes que no han expirado.
- **Búsqueda de autorización** (búsqueda de derecho a uso o libertad de operar), se utiliza para determinar si un concepto inventivo tiene autorización legal para ser utilizado, vendido o realizado. Dicha autorización se establece cuando una patente no ha sido infringida o ha expirado, es decir, es de dominio público.

- **Búsqueda de estado del arte o estado de la técnica**, La búsqueda de estado de arte es una búsqueda exhaustiva de todas las patentes disponibles y documentos no patentados, sin enfocarse sólo en un invento, sino que reúne todo lo relacionado a un campo técnico. Todos estos documentos reunidos reflejan el “estado de arte”. La extensión y profundidad de la búsqueda dependerá de qué tan acotada fue la definición de la tecnología.
- **Búsqueda de actividad de patentamiento** (Patent Landscape), es un análisis más profundo de las referencias, tanto patentadas como no, después de la realización del estado de arte. Este estudio resulta, por lo general, en la categorización de patentes, en descubrimientos fundamentales versus mejoras incrementales, una visualización de las patentes en el tiempo, la historia del desarrollo de la tecnología, e incluso análisis de las colaboraciones de algún inventor.

En base a la categorización de búsqueda de patentes antes expuesta, es posible concluir que cada búsqueda de documentos de patentes dependerá del objetivo del usuario. Sin embargo, las patentes como fuente de información resultan útiles para diferentes objetivos, ya sea desde el punto de vista legal, económico, comercial, técnico y inventivo. Una patente es un documento que ofrece muchos datos explotables para contextos diversos.

## **2.2. Bases de datos de patentes**

Para la búsqueda de información de patentes, la principal fuente y herramienta de búsqueda son las bases de datos de patentes. En base al principio de territorialidad anteriormente expuesto, cada Estado que tenga una legislación vinculada a propiedad industrial debe tener un organismo que disponga en acceso abierto las patentes para su consulta. Una base de datos de patentes es constituida por una colección de documentos de patentes y/o elementos de información concernientes a dichos documentos, como los campos correspondientes a los metadatos de registro, reivindicaciones, abstract y especificaciones de la patente (Paranjpe, 2012). Cada base de datos de patente, correspondiente a la oficina de

cada país o de organismos internacionales, ofrece una interfaz de búsqueda distinta. Entre las principales bases de datos de patentes, es posible mencionar:

Principales Bases de Datos de Patentes	
País o Institución	URL
IP Australia	<a href="https://www.ipaustralia.gov.au/">https://www.ipaustralia.gov.au/</a>
Canadá	<a href="http://www.ic.gc.ca/opic-cipo/cpd/eng/search/basic.html">http://www.ic.gc.ca/opic-cipo/cpd/eng/search/basic.html</a>
India	<a href="http://www.ic.gc.ca/opic-cipo/cpd/eng/search/basic.html">http://www.ic.gc.ca/opic-cipo/cpd/eng/search/basic.html</a>
Japón	<a href="https://www.j-platpat.inpit.go.jp/web/all/top/BTmTopEnglishPage">https://www.j-platpat.inpit.go.jp/web/all/top/BTmTopEnglishPage</a>
China	<a href="http://www.chinatrado.com/">http://www.chinatrado.com/</a>
Estados Unidos	<a href="https://www.uspto.gov/patent">https://www.uspto.gov/patent</a>
European Patent Office	<a href="https://worldwide.espacenet.com/">https://worldwide.espacenet.com/</a>
Google	<a href="https://patents.google.com/">https://patents.google.com/</a>
WIPO	<a href="https://patentscope.wipo.int/search/es/search.jsf">https://patentscope.wipo.int/search/es/search.jsf</a>
LatiPat	<a href="http://lp.espacenet.com/">http://lp.espacenet.com/</a>

TABLA 2.1. Principales bases de datos de patentes del mundo

Dichas bases de datos presentan características distintas, dependiendo de la entidad que las administre. Es posible distinguir tres tipos de bases de datos: bases de datos nacionales, que cuentan con el respaldo de la oficina de propiedad industrial de cada país; bases de datos internacionales, que son respaldadas por organismos internacionales las que recogen documentos de varios países; y las bases de datos comerciales que, a diferencia de las anteriores, tienen un costo económico y ofrecen funcionalidades más complejas (Manglano y Zulueta, 2008).

Los sistemas de búsqueda y tecnologías de recuperación de información (RI) que ofrecen las bases de datos de patentes suelen variar unas a otras. Por un lado, las funciones básicas que están ofreciendo estas herramientas son la búsqueda rápida (quick search), donde se incluyen campos predeterminados con los que el usuario puede efectuar sus consultas y la búsqueda experta (advanced search), donde el usuario puede construir una estrategia de búsqueda a través de comandos y operadores. Asimismo, se incluye la utilización de operadores booleanos y, en casos destacados, es posible visibilizar el desarrollo de funciones de búsqueda y clasificación más sofisticadas como búsqueda federada de patentes (Salampasis y Hanbury, 2014), herramientas de traducción (Magdy y Jones, 2014) e incorporación de consultas utilizando Procesamiento de Lenguaje Natural (PLN).

### 2.3. Contextualización y formulación del Problema

En el contexto chileno, el Instituto Nacional de Propiedad Industrial (INAPI) – Organismo del Estado encargado de la administración y atención de servicios de propiedad industrial en Chile – dispone de una serie de herramientas para fomentar el acceso, gestión y educación en materia de propiedad industrial.

Entre los servicios ofertados por INAPI, destaca su propia base de datos de patentes, que permite efectuar búsquedas sobre patentes vigentes en Chile mediante campos correspondientes a la descripción que se pueden extraer de un documento de patente (Figura 2.1). El buscador dispone de operadores booleanos para estructurar la sintaxis de búsqueda.

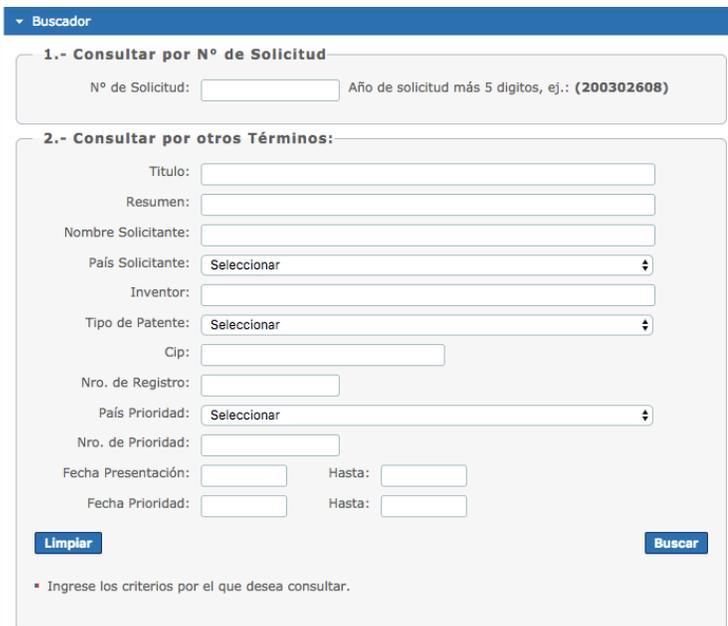


FIGURA 2.1. Buscador de Patentes.

Adicionalmente, bajo una interfaz y una tecnología de recuperación de información similar, INAPI dispone de un buscador de patentes de dominio público en Chile, por lo tanto, su plazo de protección legal ha expirado y pueden ser utilizadas con toda libertad por cualquier persona o institución. Ver Figura 2.2

Conoce patentes solicitadas en Chile cuyos plazos de protección han expirado y por lo tanto son parte del dominio público.

**Buscador**

Título:  ?

Resumen:  ?

Nombre Solicitante:  ?

País Solicitante: **Seleccionar** ?

Inventor:  ?

Tipo de Patente: **Seleccionar** ?

Cip:  ?

Nro. de Registro:  ?

Fecha Presentación:  Hasta:  ?

**Limpiar** **Buscar**

- Envíenos sus observaciones y comentarios del buscador al correo [info@inapi.cl](mailto:info@inapi.cl)
- Ingrese los criterios por los que desea consultar.

\* AND = Operador por defecto (ingresar terminos separados por espacio)  
OR = Ingresar terminos separados por "OR" para incrementar posibilidades de búsqueda

FIGURA 2.2. Buscador de Patentes de Dominio Público.

En base a las experiencias previamente revisadas, las bases de datos de patentes del Instituto Nacional de Propiedad Industrial no han incorporado aún mejoras en sus tecnologías de búsqueda, limitándose al uso de consultas a bases de datos relacionales y el uso de operadores booleanos para estructurar ecuaciones de búsqueda.

En términos concretos, las tecnologías semánticas y de Linked Open Data (LOD) son una gran oportunidad para visibilizar las patentes como fuentes de información relacionable con otras fuentes de datos como empresas, investigadores, proyectos de I+D+i, entre otros. De esta manera, se hace necesario incursionar en experiencias internacionales sobre la incorporación de tecnologías de Linked Data en el mundo de la propiedad industrial para analizar los elementos replicables y reusables para incorporar en el contexto nacional de las patentes.

## **2.4. Objetivos**

### **2.4.1. Objetivo General**

Establecer un procedimiento de optimización en la búsqueda y recuperación de patentes del Instituto Nacional de Propiedad Industrial basado en la integración de tecnologías semánticas que permita enlazar datos, facilitando el acceso a información de valor para las instituciones involucradas en el ecosistema nacional de I+D+i.

### **2.4.2. Objetivos Específicos**

- Identificar los campos descriptivos determinantes en el proceso de recuperación de información dentro de la base de datos de patentes del Instituto Nacional de Propiedad Industrial.
- Analizar los elementos técnicos y procedimentales que permitan efectuar una traducción de una base de datos relacional de patentes a una estructura semántica en RDF.
- Determinar una estrategia procedimental para la incorporación de tecnologías semánticas o de linked data en la base de datos de patentes del Instituto Nacional de Propiedad Industrial.
- Describir las potenciales aplicaciones que ofrecen las tecnologías de linked data al campo de la propiedad industrial en Chile.

## **2.5. Metodología**

De acuerdo a los objetivos previamente establecidos, la metodología determinada es la siguiente:

### **2.5.1. Análisis de la base de datos de patentes de INAPI**

El primer paso de la metodología supone un análisis profundo de la base de datos de patentes de INAPI, desde su modelo de datos, hasta la constitución de campos y tablas de

cada entidad con sus correspondientes atributos. En esta fase, se analizarán registros correspondientes a documentos de patentes y se definirán los principales campos descriptivos que serán determinantes en el modelamiento de datos.

### **2.5.2. Definición de Ontología(s) reusables para el modelo de datos**

El siguiente paso está enfocado en el estudio de la ontología existente que pueda ser reusada o adaptada para el contexto de la propiedad industrial nacional, particularmente con los datos de INAPI. De esta manera, se pretende analizar la ontología para determinar su relevancia potencial adopción, de acuerdo al objetivo de estudio.

### **2.5.3. Modelamiento de datos (R2RML)**

La tercera etapa, utilizando la base de datos relacional y la(s) ontología(s), consiste en un modelamiento de datos, mapping o traducción de una base de datos relacional a RDF mediante R2RML. En esta fase se busca testear el procedimiento de modelado de datos con información de patentes para publicar el resultado final en RDF.

### **2.5.4. Definición y publicación de Tripletas en RDF**

La cuarta etapa consiste en la publicación en formato RDF de los registros o datos de documentos de patentes procesados en el mapping y el análisis de su estructura sintáctica para la comprensión de las tripletas en un ambiente semántico o de linked data.

### **2.5.5. Descripción de usos y aplicaciones de datos en RDF**

Finalmente, la quinta etapa se destina a describir las aplicaciones que un conjunto de datos publicados en RDF puede ofrecer en el ámbito de la propiedad industrial en Chile, específicamente en ámbito de las bases de datos de patentes, en un contexto de innovación abierta y colaborativa.

### **2.5.6. Resultados esperados**

De acuerdo a los objetivos planteados, los resultados esperados para esta investigación se definen en los siguientes puntos:

1. Estructurar un proceso de optimización de búsqueda y recuperación de información tecnológica mediante la incorporación de tecnologías semánticas al buscador de patentes del Instituto Nacional de Propiedad Industrial, junto con establecer los principales beneficios que dicha integración conllevaría para los usuarios del servicio.
2. Definición de una estrategia procedimental de traducción de una bases de datos relacional a un esquema de datos RDF, basado en una ontología. Dicha estrategia sería potencialmente replicable en cualquier base de datos relacional.
3. Descripción de usos y aplicaciones de tecnologías semánticas o de linked data potencialmente usables por una oficina de propiedad industrial para visibilizar sus datos. En el caso de INAPI, se propone la publicación de un servicio de datos enlazados en RDF (<http://datos.inapi.cl>).

### **3. ESTADO DEL ARTE**

#### **3.1. Las patentes como fuente de información**

Las patentes son una gran fuente de información que engloba todos los ámbitos de los desarrollos tecnológicos y facilita el hallazgo de tecnología procedente de cualquier parte del mundo (OMPI, 2013).

Como se expuso previamente, el sistema de patentes y propiedad industrial tiene como objetivo principal fomentar la innovación y el desarrollo económico a través de dos elementos fundamentales: protección de la creatividad, recompensando las inversiones utilizadas en los desarrollos y publicando la información técnica asociada a las nuevas invenciones.

##### **1. Protección de la creatividad**

Una patente es el medio más recurrente que existe para proteger los derechos de los inventores y se concibe como un derecho exclusivo que cede el Estado para la protección de algún invento. Lo anterior permite la utilización, explotación comercial e impedimento de uso de la invención, y se puede demandar a quien explote el invento sin autorización previa del titular de la patente.

Las patentes otorgan dichos derechos por un período de tiempo limitado – en Chile de 20 años – y aplica solamente al país o al grupo de países donde se efectúa la solicitud de la patente. En el resto del mundo, la patente no cuenta con dichos derechos.

En el caso de Chile, los privilegios de invención y los derechos de protección de propiedad industrial están determinados por la ley 19.039 y sus modificaciones junto al reglamento. La solicitud de patente se efectúa en el Instituto Nacional de Propiedad Industrial (INAPI).

De esta manera, al brindar un derecho exclusivo al solicitante, la patente resulta un incentivo a modo de ofrecer un reconocimiento por su actividad inventiva y retribuir de forma económica su invención. Lo anterior resulta un estímulo a la

innovación, que busca transformar al sistema de patentes en un mecanismo de mejora de la calidad de vida humana.

## **2. Publicación de información técnica**

Así como la patente concede derechos exclusivos sobre la invención, el titular también tiene la obligación de publicar y divulgar toda la información técnica de la invención patentada, para que otros puedan beneficiarse del conocimiento generado y contribuir al desarrollo sistemático de algún ámbito tecnológico (OMPI, 2013).

La publicación de la información técnica de una invención es requisito fundamental para la concesión de una patente en cualquier oficina de propiedad industrial. De esta manera, el sistema de patentes pretende hacer accesible esta información publicando de manera abierta los documentos, para que cualquier persona las pueda consultar. Lo anterior fomenta la reutilización del conocimiento pre-existente para promover la innovación y los nuevos desarrollos tecnológicos. Asimismo, se busca equilibrar intereses de inventores e intereses de la comunidad que pueda hacer uso de este conocimiento.

En virtud de lo anterior, es posible identificar ciertos elementos en la propiedad industrial que contribuyen a la constitución de un dominio público o patrimonio común en el ámbito tecnológico y su innegable rol en la promoción de la creación de nuevas invenciones intelectuales. Según (Schmitz Vaccaro, 2009)), una patente puede pasar a dominio público por las siguientes razones:

- Materias impatentables como teorías científicas, elementos de la naturaleza o métodos matemáticos.
- Solicitudes de patentes que hayan sido rechazadas o abandonadas en su tramitación.
- Patentes vigentes en otros países y no solicitadas en el país en cuestión (territorialidad).
- Patentes cuya vigencia haya expirado.

- Patentes declaradas nulas por sentencia judicial.
- Patentes caducadas por falta de pago periódico, de acuerdo al requerimiento de cada oficina.

Asimismo, en relación a la información de patentes, es fundamental comprender que se publican en bases de datos a escala mundial, por lo que se puede obtener conocimiento de cualquier parte del mundo, siendo una fuente de información muy relevante no sólo útiles para grandes empresas, sino también para emprendedores, inventores, investigadores e incluso ciudadanos interesados en algún desarrollo que puedan replicar.

Por otra parte, las patentes ofrecen información de alto valor y aplicable a distintos contextos y puede ser extraída desde los documentos de patentes (OMPI, 2013):

- **Información técnica:** cada patente posee una descripción técnica y dibujos descriptivos de una invención. Lo anterior permite extraer información altamente novedosa sobre el desarrollo de una tecnología para poder replicarla o mejorarla.
- **Información jurídica:** de igual manera, las patentes ofrecen información acerca del alcance de la protección de una invención, es decir, qué elementos son los protegidos, además de su condición jurídica y su vigencia en determinados países.
- **Información comercial:** es posible obtener información comercial analizando a los solicitantes, inventores o países de origen de la solicitud de patentes para comprender los mercados donde las empresas están desarrollando y protegiendo tecnologías.
- **Información para políticas públicas:** asimismo, es indispensable considerar a las patentes como una fuente de información que permite conocer el estado del arte de algún área industrial o, si se tienen las metodologías de análisis correctas, incluso es posible hacer prospectiva tecnológica, lo que se transforma en información de alto valor para diseñar políticas públicas, como una estrategia industrial a nivel nacional, regional o internacional.

A raíz de lo expuesto, la información contenida en patentes puede ser utilizada para diversos fines como evitar la duplicación de esfuerzos en investigación y desarrollo, evitar infracciones legales, estimar valores de patentes, explotar tecnologías de dominio público, fomentar la innovación y analizar tendencias tecnológicas.

Por otra parte, las patentes se caracterizan por ser documentos que exponen su información de manera normalizada y estructurada. Toda patente debe publicarse en un formato y con ciertos datos pre-establecidos, en base a la norma ST.9 de la OMPI.

10		OFICINA ESPAÑOLA DE PATENTES Y MARCAS  ESPAÑA	
11	Número de publicación: <b>2 467 702</b>		
11	Int. Cl.: <b>A61K 39/00</b> (2006.01) <b>C07K 7/23</b> (2006.01) <b>C07K 7/08</b> (2006.01)		
12	TRADUCCIÓN DE PATENTE EUROPEA		T3
26	Fecha de presentación y número de la solicitud europea: <b>14.04.2010</b> E 10721286 (2)		
27	Fecha y número de publicación de la concesión europea: <b>02.04.2014</b> EP 2431052		
54	Título: <b>Proteína para inmunocstración de mamíferos</b>		
20	Prioridad: <b>15.04.2009 CL 9002009</b>		
40	Fecha de publicación y mención en BOPI de la traducción de la patente: <b>12.06.2014</b>		
73	Titular/es: <b>UNIVERSIDAD DE CHILE (100.0%)</b> <b>Avenida Libertador Bernardo O'Higgins No. 1058</b> <b>Santiago 8330111, CL</b>		
72	Inventor/es: <b>SÁENZ ITURRIAGA, LEONARDO ENRIQUE</b>		
73	Agente/Representante: <b>CARPINTERO LÓPEZ, Mario</b>		

FIGURA 3.1. Documento de Patente.

En primer lugar, las patentes presentan sus datos bibliográficos, incorporando información de valor para quien busca y lee patentes. De este modo, un documento de patente presenta los siguientes datos (OMPI, 2016):

- **Título:** frase que debe dar cuenta de la temática que aborda la patente. En algunas bases de datos se puede encontrar la traducción del título de la patente además de la versión original.
- **Inventor:** el nombre de la o las personas que han hecho del desarrollo intelectual y material de la invención. El nombre del inventor puede coincidir con el del solicitante en algunas ocasiones.
- **Solicitante:** también conocido como titular de la patente, corresponde a la persona o la organización que presenta una solicitud de patente y tendrá la protección sobre la invención en cuestión. Puede haber más de un solicitante en una solicitud y, en algunos casos, puede coincidir con el inventor.
- **Representante:** el nombre del agente o abogado que actúa en representación del solicitante o el inventor.
- **Fecha de presentación o solicitud:** corresponde a la fecha en que es presentada la solicitud de patente en una oficina de propiedad industrial determinada. Es preciso recordar que una solicitud de patente puede ser presentada en más de una oficina de propiedad industrial, según el territorio en el cual se quiera proteger la invención. De igual manera, con un código compuesto, se elabora el número de solicitud y es asignado por la oficina pertinente.
- **Fecha de publicación:** es la fecha en que la patente ha sido publicada por primera vez por una oficina de propiedad industrial determinada y su contenido pasa a ser parte del Estado del Arte o de la Técnica.
- **Número de prioridad:** corresponde a la primera presentación de solicitud de patente de invención en una oficina de propiedad industrial determinada. A partir de este momento, el solicitante tiene un plazo de un año para presentar solicitudes en otras oficinas. Asimismo, todas las patentes presentadas en distintas oficinas (países) correspondientes a la misma invención, pasan a constituir una familia de patentes. Por otra parte, el primer país al que se le presenta la solicitud pasa a ser el país de prioridad de dicha patente.

Por ejemplo, si una patente es publicada con prioridad internacional mediante el Tratado de Cooperación en materia de Patentes (PCT)<sup>1</sup>, este documento toma la condición de "patente prioritaria". Por otro lado, si posteriormente es presentada en Estados Unidos, en China y en la Oficina Europea de Patentes, resultarían cuatro documentos distintos, escritos con distintos idiomas y con ligeros cambios de presentación (de acuerdo a la legislación de cada país), con un número de publicación distinto, pero referidos a la misma invención. Todos estos documentos corresponden a una familia de patentes.

- **Situación jurídica:** define la situación legal de la patente, en relación a su vigencia. Da cuenta si la patente ha sido concedida o no y, en segunda instancia, si la protección de derechos sigue vigente o ha pasado a dominio público.
- **Clasificación Internacional de Patentes (CIP)<sup>2</sup>:** la CIP o Clasificación Internacional de Patentes es un sistema jerárquico de símbolos que sirve para clasificar las patentes. La CIP divide las áreas tecnológicas en ocho macro-categorías, con aproximadamente 70.000 subdivisiones, a las cuales les corresponde un código alfanumérico. Asimismo, la CIP es publicada, revisada y actualizada de manera sistemática por la Organización Mundial de Propiedad Intelectual (OMPI).  
A cada patente se le asigna uno o más códigos de clasificación que ayudan a describir su área temática. De igual manera, los códigos CIP son de gran ayuda para efectuar búsquedas de un área temática en específico.

---

<sup>1</sup>El Tratado de Cooperación en materia de Patentes (PCT) ofrece protección internacional a los solicitantes para sus invenciones y asiste a las Oficinas en las decisiones sobre el otorgamiento de patentes, y también facilita el acceso a la extensa información técnica de las patentes. Al presentar una solicitud internacional de patente según el PCT, los solicitantes tienen la posibilidad de proteger su invención a nivel mundial en un gran número de países.

<sup>2</sup>Es posible acceder a la Clasificación Internacional de Patentes en el sitio web de la OMPI <http://web2.wipo.int/classifications/ipc/ipcpub>

## Datos bibliográficos: CL31112002 (A) — 2002-11-11

★ En mi lista de patentes	Anterior ◀	8 / 8	▶ Siguiente	➤ EP Registro	📄 Informe de datos de error	🖨 Imprimir
<b>VACUNA DE ADN QUE COMPRENDE UN FRAGMENTO DE ADN QUE CODIFICA PARA LA PROTEINA DE ESTRES TERMICO 10(HSP10) DE PISCIRICKETTSIA SALMONIS, INSERTADA EN UN VECTOR ADECUADO,PROCEDIMIENTO DE PREPARACION, DNA QUE CODIFICA A HSP10 Y LOS USOS DE LA VACUNA.</b>						
Página favorito	<a href="#">CL31112002 (A) - VACUNA DE ADN QUE COMPRENDE UN FRAGMENTO DE ADN QUE CODIFICA PARA LA PROTEINA DE ESTRES TERMICO 10(HSP10) DE PISCIRICKETTSIA SALMONIS, INSERTADA EN UN VECTOR ADECUADO,PROCEDIMIENTO DE PREPARACION, DNA QUE CODIFICA A HSP10 Y LOS USOS DE LA VACUNA.</a>					
Inventor(es):	BURZIO, ERIZ, LUIS,; ROSEMBLATT, SILBER, MARIO,; VALENZUELA, VALDES, PABLO, ±					
Solicitante(s):	FUNDACION CHILE [CL]; FUNDACION CIENCIA PARA LA VIDA [CL] ±					
Clasificación:	- internacional: <b>A61K39/02; A61K48/00; C07K14/29; C12N15/31; G01N33/53;</b> - cooperativa:					
Número de solicitud:	CL20010003111 20011220					
Número(s) de prioridad:	CL20010003111 20011220					

FIGURA 3.2. Registro bibliográfico de un documento de Patente.

- **Descripción:** es una explicación clara y concisa de la invención, su relación con tecnologías existentes y cómo esta invención ayuda a la resolución de problemas no abordados por invenciones pre-existentes. Asimismo, la descripción puede dar cuenta de funcionalidades específicas de la nueva tecnología.
- **Reivindicaciones:** corresponde a la definición legal del objeto para el que se solicita u otorga protección. Cada afirmación dentro de las reivindicaciones define a la invención y determina sus características técnicas únicas que le dan calidad de innovadora y, por lo tanto, merecen protección. Las reivindicaciones deben ser claras y concisas y deben sustentarse con la descripción. De igual manera, la descripción y los dibujos ayudan a interpretar las reivindicaciones.
- **Citación y referencias:** algunas patentes incluyen también referencias y citas a información tecnológica relacionada que ha sido descubierta por el inventor, el solicitante o el examinador de la solicitud de patente; estas referencias pueden incluir tanto documentos de patente como documentación de otra índole.

Las características previamente expuestas definen a los documentos de patente como una fuente de información relevante para distintos contextos. Asimismo, su característica de documento estructurado hace factible su descripción en bases de datos documentales que fomentan el acceso a la lectura de estos documentos.

A raíz de lo anterior, se han diseñado múltiples tecnologías de búsqueda, extracción y análisis de documentos de patentes, incluyendo las tecnologías de la web semántica y linked data como una herramienta que potencia la recuperación de información tecnológica.

### **3.2. Aproximaciones a la web semántica y linked data**

Antes de analizar las aplicaciones de la web semántica en el contexto de la propiedad industrial o la gestión de I+D+i, resulta fundamental establecer una definición acerca de los principios que conlleva y las tecnologías implicadas.

En primer lugar, la web semántica no es un concepto nuevo, sino que fue la idea original de Tim Berners-Lee (Berners-Lee, Dimitroyannis, Mallinckrodt, y McKay, 1994), el inventor de la web, quien planteó en el año 1994 la necesidad de agregar el concepto semántica a la web. La aplicación de esta idea queda definida por Berners-Lee y Miller (Lee y Miller, 2002), quienes describen a la web semántica como:

Una extensión de la actual web en la que a la información disponible se le otorga un significado bien definido que permita a los computadores y las personas trabajar en cooperación. Está basada en la idea de proporcionar en la web datos bien definidos y enlazados permitiendo que aplicaciones heterogéneas localicen, integren, razonen y reutilicen la información presente en la web.

En base a esa definición, es posible identificar ciertos elementos relevantes. En primer lugar, la web semántica es una extensión de la web actual, es decir, coexiste con todas las tecnologías de la web. En segundo lugar, resalta la necesidad de definir los datos para entregarle significado (semántica). Finalmente, declara la opción de que las máquinas sean capaces de manipular y trabajar con los datos (Sánchez, 2011).

Cuando se refiere a la web actual, tradicional o web sintáctica, se habla de una serie de documentos enlazados que pueden ser comprensibles para los humanos que entienden el lenguaje natural. A pesar de esto, esos documentos resultan incomprensibles para las máquinas y éstas se limitan a visualizar los contenidos mediante HTML y CSS o realizar búsquedas mediante coincidencia de palabras clave en esos documentos. Sin embargo, la web semántica plantea la idea de no enlazar documentos, sino enlazar datos adecuadamente descritos y estandarizados, que representan conceptos concretos y abstractos del mundo real, para que sea comprensible para las máquinas (Gayo, 2012). Lo anterior permite que las máquinas reutilicen e integren la información para utilizarla de manera automatizada.

Según Pastor (Sánchez, 2011), las bases de la web semántica se componen en los siguientes postulados:

- Uso de modelos de metadatos para describir recursos de información.
- Uso de vocabularios para representar dichos metadatos
- Desarrollo y uso de esquemas u ontologías que describen las relaciones entre los recursos y las propiedades descritas.
- Identificación, localización, interconexión y reutilización de fuentes de datos que permiten su integración automatizada.
- Inferencia de nueva información a partir de las relaciones establecidas entre los datos.

De los postulados anteriores se desprende que hay una serie de elementos tecnológicos involucrados que dan vida a la web semántica como los metadatos, vocabularios, ontologías, elementos identificadores y relaciones entre los datos. Dichos elementos constituyen una arquitectura de la web semántica (Gerber, Barnard, y van der Merwe, 2007).

Para entender la arquitectura de la web semántica es preciso tener en consideración las características de la web original donde los recursos de información se localizan mediante URL, se utiliza el protocolo HTTP para establecer comunicación entre los clientes y servidores y los contenidos se muestran mediante HTML.

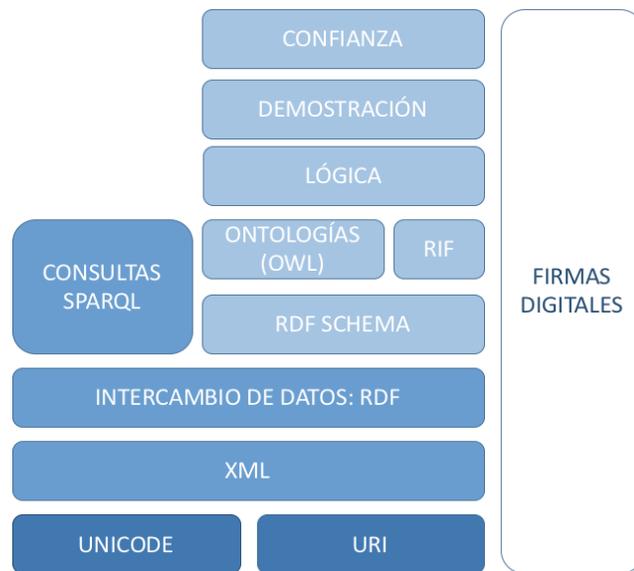


FIGURA 3.3. Arquitectura de la web semántica.  
(Berners-Lee et al., 1994)

La imagen (Figure 3.3) gráfica la arquitectura de la web semántica denominada por Berners-Lee (Berners-Lee et al., 1994) como “La tarta de la web semántica”. Dicha estructura presenta múltiples elementos, de los cuales se destacan los siguientes:

- Unicode: Se emplea como el estándar de codificación de datos y caracteres internacionales, lo que permite la utilización en cualquier idioma y alfabeto.
- URI: La web semántica fortalece la identificación y localización de recursos de información mediante los Identificadores de Recursos Uniformes o URI (Uniform Resource Identifier). Se definen como cadenas de caracteres que posibilita la identificación de recursos de información abstractos o localizables en la web.
- XML: eXtensive Markup Language (XML), es un lenguaje que viene a complementar a HTML en la representación y descripción de información mediante una sintáxis jerárquica que puede ser procesada de manera sencilla por las aplicaciones.

- **RDF:** Resource Description Framework (RDF) es un modelo de datos que permite identificar y describir recursos, sus propiedades y relaciones con otros recursos mediante tripletas (Sujeto – predicado – objeto).
- **Ontología:** El propósito de una ontología es describir objetos y sus relaciones, definiendo clases, subclases, propiedades y subpropiedades. Normalmente las ontologías funcionan con OWL (Web Ontology Language) que complementa a RDF expresando las relaciones semánticas.
- **SPARQL:** Es uno de los lenguajes de consultas para la web semántica que juega un rol similar al de SQL en el contexto de las bases de datos relacionales.
- **Reglas RIF:** Las RIF (Rule Interchange Format) son reglas que permiten la interoperabilidad entre los sistemas.

Existen otros elementos, lenguajes y esquemas que permiten hacer funcionar a la web semántica. A continuación, se describen los elementos más relevantes para efectuar el mapping de la base de datos de INAPI: URI, Ontologías y RDF.

### **3.2.1. Uniform Resource Identifier (URI)**

Un requisito fundamental para que la web de datos funcione es que sus recursos deben estar identificados de forma global y el identificador global de recursos es la URI. Como se ha expuesto previamente, una URI es una cadena de caracteres que proporciona una identificación comprensible de un recurso de información.

De acuerdo a Gayo (Gayo, 2012), es necesario hacer una distinción entre URI, recurso y representación:

- **URI** es el identificador del recurso. Por ejemplo, un código que identifique una institución.
- **Recurso** es la entidad que se está identificando mediante la URI.
- **Representación:** es el recurso identificado plasmado en una representación. Un recurso puede tener diversas representaciones.

La Figure 3.4 muestra cómo una URI (Cadena de caracteres) puede identificar un recurso (Romeo y Julieta) que es representado en algún formato (Libro).

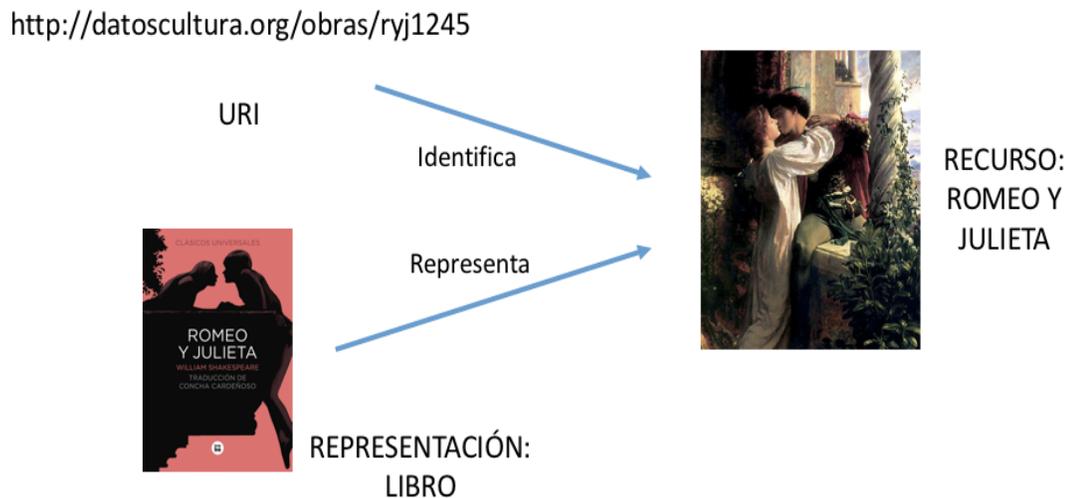


FIGURA 3.4. URI, recurso y representación.

Las posibilidades de identificación mediante el uso de una URI son amplias. En primera instancia, es fundamental reconocer que una URI puede representar entidades digitales como físicas, entidades del mundo real y conceptos abstractos. Por lo tanto, una URI puede identificar recursos de información (imágenes, documentos, PDF, etc.) o recursos no documentales como una persona, un conjunto de personas, una casa o un sentimiento.

Por otra parte, una URI puede utilizarse para describir relaciones entre recursos. Por ejemplo, la propiedad de que una entidad sea creadora o perteneciente de otra.

Las URIs son fundamentales en el contexto de la web semántica porque se busca transformar la web sintáctica a una web de datos donde la unidad mínima de información sea un dato. En este contexto, cada dato se identifica con una URI y es posible establecer una relación con otros datos, que a su vez estarán identificados con otras URIs. Es más, incluso las relaciones pueden estar identificadas con una URI.

### 3.2.2. Ontologías

En el campo de las ciencias de la computación, una ontología se define como los “objetos, conceptos y otras entidades que existen en un área de interés (o dominio) y las relaciones entre ambos” (Sánchez-Jiménez y Gil-Urdiciain, 2007). Asimismo, (Shadbolt, Berners-Lee, y Hall, 2006) se refiere a una ontología como un documento o un archivo que define formalmente las relaciones entre muchos términos, presentando normalmente una taxonomía y un conjunto de reglas de inferencia. Por un lado, la taxonomía define las clases, subclases y las relaciones entre ellas. Por otra parte, las reglas de inferencia permiten a las máquinas entender cómo manipular estos términos y sus relaciones. Por esta razón, las ontologías juegan un papel crucial en el contexto de la web semántica, permitiendo el procesamiento, compartimiento y reutilización de datos por parte de aplicaciones web (Decker et al., 2000).

En el campo de la web semántica, se propuso a OWL como uno de los lenguajes de representación de ontologías. OWL (Web Ontology Language) es un lenguaje desarrollado por la W3C que tiene por objetivo publicar y compartir en la web la información formalizada de ontologías que definen términos y relaciones en un dominio determinado. OWL utiliza las relaciones explícitas que proporciona RDF para permitir la distribución de las ontologías en los sistemas, ya que OWL permite que las ontologías se refieran a términos de otras ontologías. Por esta propiedad, OWL está diseñado especialmente para la web semántica (Shadbolt et al., 2006).

De acuerdo a (Sánchez, 2011), las ontologías son fundamentales en el ámbito de la web semántica ya que son las que precisan las semánticas bien definidas para poder estructurar y ejecutar los procesos de inferencia. Aunque el RDF presenta elementos semánticos básicos, los que ofrece OWL son mucho más potentes y avanzados porque posee mayor capacidad para interpretar el contenido por parte de máquinas que en relación a otros formatos.

En términos concretos, OWL permite definir los siguientes elementos:

- **Clases:** Las clases son conjuntos de elementos o de individuos. Por ejemplo, conjunto de documentos, de inventores, de tecnologías, países, etc.
- **Individuos:** Se definen como los elementos concretos pertenecientes a las clases. Por ejemplo, un determinado país o una determinada persona. OWL posee una propiedad particular que no asume nombres únicos, es decir, un mismo individuo podría tener dos nombres diferentes. Esta propiedad calza con la arquitectura de la web semántica desde la perspectiva en que dos URIs distintas pueden estar hablando de un mismo individuo.
- **Propiedades:** Para OWL, existen dos tipos de propiedades. Por un lado, están las propiedades de objetos (object properties) que relacionan dos individuos entre sí y, por otro lado, están las propiedades de tipo de datos (Data properties) que relacionan un individuo con un tipo de dato en particular.
- **Atributos de propiedades:** En OWL, las propiedades pueden tener atributos que definen si esa propiedad es reflexiva, simétrica, transitiva, funcional o funcional inversa.
- **Restricciones:** OWL también permite declarar restricciones que deben cumplirse dentro de la ontología. Por ejemplo, prohibir algún tipo de relación en la ontología.
- **Axiomas:** Finalmente, OWL define los axiomas que son declaraciones lógicas que deben cumplir los elementos de la ontología.

En los últimos años, OWL se ha transformado en el estándar aceptado para la creación y mantención de ontologías a nivel mundial.

### 3.2.3. RDF

El desarrollo de técnicas de descripción de recursos de información mediante el uso de metadatos es fundamental para el contexto de la web semántica (Labra, 2011). Bajo esa premisa, la W3C desarrolló RDF (Resource Description Framework), un modelo de descripción estructurada de recursos de información en la web que resulta también adecuado para representar datos (Decker et al., 2000) que busca representar de manera minimalista

el conocimiento o los recursos de información en la web (Shadbolt et al., 2006). RDF viene a dar respuesta a algunos problemas que ofrece la estructura jerárquica más limitada de XML, ya que a partir de un modelo común este esquema ofrece una descripción de recursos determinada por una semántica básica, lo que supone una ventaja significativa sobre XML.

El modelo básico de descripción en RDF tiene una estructura similar al objeto – atributo valor reconocido en modelos previos (Decker et al., 2000), pero opera con una lógica más compleja presentando la siguiente declaración:

SUJETO – PREDICADO – OBJETO

Esta declaración define todos los elementos fundamentales para la descripción de un recurso: el elemento a describir, la propiedad o relación del recurso y el valor asignado a esa propiedad o el recurso con el cual se establece relación. A esta estructura se le denomina tripleta, triple o sentencia y es la base de RDF. Una tripleta está compuesta por un sujeto, un predicado y un objeto, los cuales se caracterizan por lo siguiente:

- **Sujeto:** es el recurso a describir y que es identificado por una URI. El sujeto de una tripleta RDF puede ser cualquier tipo de recurso, como una página web, un sitio web, una imagen, una persona, un libro o incluso una entidad abstracta. Por una parte la URI identifica al recurso y la URL ayuda a localizarlo en una dirección web.
- **Predicado:** es la propiedad, característica, aspecto, atributo o relación usada para describir al sujeto. Cada propiedad está definida por un significado específico, sus posibles valores y los tipos de recursos sobre los que se puede efectuar la descripción. En algunos casos, el predicado también puede ser una URI. Por ejemplo, la URI <http://purl.org/dc/terms/creator> identifica la propiedad de que un sujeto sea creador de otra entidad (objeto).
- **Objeto:** puede ser el valor asignado a la propiedad o el recurso con el cual establece relación el sujeto. En base a eso, el objeto puede resultar un literal, es decir un valor concreto (cadena de caracteres) o una URI que identifique a otro recurso.

Lo anterior se ve reflejado en los siguientes ejemplos:

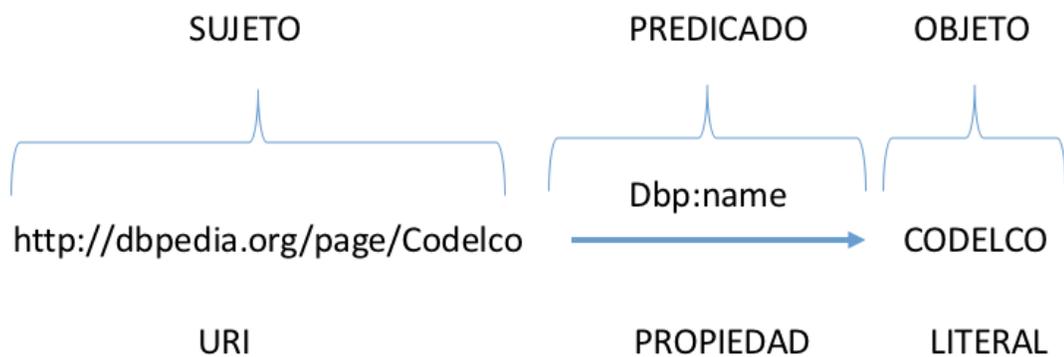


FIGURA 3.5. Tripletta en RDF (URI-Propiedad-Literal).

La Figure 3.5 identifica al sujeto, el cual es una URI de DBPedia que corresponde a su página referida a la empresa chilena Codelco. El predicado se refiere a la propiedad nombre definida en DBPedia y el objeto es “Codelco” entendido como un literal, es decir una cadena caracteres. Dicha estructura descriptiva se resume en la declaración `http://dbpedia.org/page/Codelco` lleva por nombre Codelco.

Por otra parte, la estructura previa también podría relacionar una URI con otra. Por ejemplo, la Figure 3.6 muestra como la URI de Codelco se relaciona con la URI de “Minería” bajo la relación “Industry”.

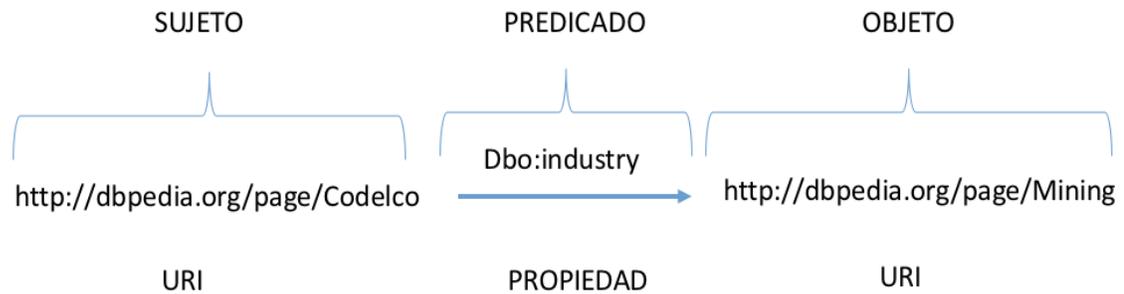


FIGURA 3.6. Tripletas en RDF (URI-Propiedad-URI).

Estas tripletas pueden ser vistas como una relación entre nodos, donde el nodo de origen es el sujeto, el enlace es el predicado y el nodo receptor es el objeto. Si se declaran muchas tripletas se obtiene la estructura de un grafo donde los nodos son los sujetos y objetos y los enlaces son los predicados. La imagen que se muestra a continuación (Figura 3.7) exhibe de forma gráfica cómo se visualizaría el grafo si conectamos las tripletas antes mencionadas. En dicho ejemplo podemos inferir que Codelco se dedica a la industria de la minería y produce cobre.

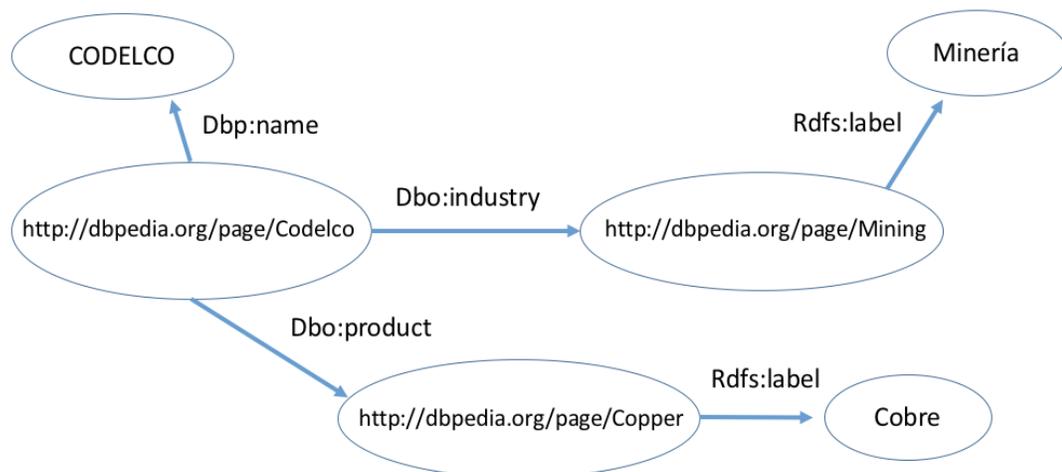


FIGURA 3.7. Grafo en RDF.

De esta manera, RDF se transforma en un esquema de descripción y construcción de grafos. El modelo de construcción de grafos en RDF es composicional, ya que es posible conectar un grafo en un servidor con elementos de otro grafo ubicado en otro servidor, mezclándose y obteniendo un grafo más grande. Por esta razón, se le ha denominado a la web semántica como el “Gran Grafo Global” (Gayo, 2012).

Para efectos de la unión de grafos, como puede verse en los ejemplos (Figure 3.7), el rol de la URI es fundamental porque permite identificar a los nodos implicados en la construcción del grafo para posteriormente conectarlos. Una URI permite la integración de datos entre dos fuentes totalmente heterogéneas y resulta un elemento clave en el modelo de construcción de grafos en RDF.



FIGURA 3.8. Representación textual de tripleta en RDF.

Para la representación textual de RDF se utilizan distintos formatos como RDF/XML, una representación basada en XML, N-Triples, N3 o Turtle. La proliferación de formatos ha perjudicado en cierta medida la adopción de RDF porque lo relevante no es el formato de la representación, sino el modelo del grafo y su relación con la ontología, lo que facilita la integración y el enlace de datos.

### 3.3. Integración de tecnologías de la web semántica y Linked Data (LD)

Todas las tecnologías anteriormente expuestas son el mecanismo para lograr un objetivo: enlazar e integrar datos en la web, lo que también se denomina como un servicio de datos enlazados o Linked Data (LD). Según (Bizer, Heath, y Berners-Lee, 2009), el Linked Data trata de utilizar la Web para crear enlaces entre datos desde distintas fuentes. Desde el punto de vista técnico, los datos enlazados se refieren a la publicación de datos en la

Web de tal manera que son legibles por una máquina, ya que tienen definida una semántica explícita, y pueden vincularse o ser vinculados con data sets externos.

Mientras la unidad primaria de la Web de hipertexto es el HTML y estos documentos son conectados por hiperenlaces, la Web de datos enlazados se basa en la publicación de datos en RDF. Los datos enlazados usan a RDF para hacer declaraciones en tripletas que relacionan elementos digitales y del mundo real en la Web.

Por otra parte, Linked Data presenta ciertas reglas para publicar datos en la web de manera de que estos datos puedan ser parte de un espacio global de datos único. Estas reglas se les conoce como los Principios del Linked Data (Bizer et al., 2009),(Gayo, 2012) y se resumen de la siguiente manera:

- Uso de URIs para nombrar entidades.
- Uso de URIs de HTTP para poder localizar dichos recursos o entidades.
- Uso de estándares (RDF, SPARQL) para entregar información útil cuando se accede a la URI.
- Incluir enlaces a otras URIs, para permitir el descubrimiento de otras entidades relacionadas al dominio que se está tratando.

De esta manera, los datos enlazados se basan en dos tecnologías fundamentales para la Web: las URIs y el Protocolo de Transferencia de Hipertexto (HTTP), debido a que ayudan a identificar y localizar, mediante URL, las entidades existentes. Adicionalmente, las URIs y el HTTP son complementados por RDF (Bizer, Heath, Idehen, y Berners-Lee, 2008), quien provee un modelo basado en grafos con el cual se estructuran y enlazan datos que describen a las entidades.

En resumen, Linked Data utiliza tecnologías como RDF y OWL para crear vocabularios que pueden describir entidades en el mundo y cómo éstas se relacionan entre sí. Dichos vocabularios son colecciones de clases y propiedades. Estos vocabularios describen recursos definidos por URIs y localizados mediante HTTP, fijando una arquitectura de la Web

de datos, donde las entidades están conectadas por enlaces RDF, creando un grafo global de datos que permite integrar, enlazar y descubrir nuevas fuentes de datos.

La interconexión entre distintas fuentes de datos no es nueva y los datos enlazados no son la única iniciativa. En ese sentido, autores como (Haslhofer y Schandl, 2010) hacen la distinción entre soluciones basadas en tecnologías como OAI-PMH, que supone la creación de repositorios de recursos de información centralizados que son agregados o cosechados de repositorios externos. Básicamente, OAI-PMH funciona con una arquitectura cliente-servidor en la que un servicio opera como recolector de metadatos y los solicita a un proveedor obteniendo una respuesta en XML, para posteriormente utilizar esos datos en su servicio.

Por otro lado, Linked Data propone hacer accesible mediante un formato estándar (RDF) aquellos datos que se pueden enlazar desde fuentes externas (Figure 3.9).

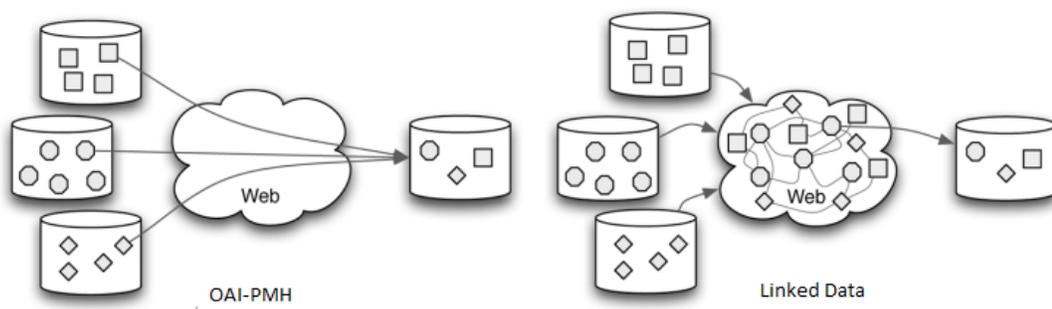


FIGURA 3.9. Diferencias entre OAI-PMH y Linked Data.

(Haslhofer y Schandl, 2010)

Sin embargo, las distintas tecnologías de integración no son excluyentes unas con otras, lo que hace que el interés por incorporar tecnologías semánticas a distintos servicios web sea creciente. (Sánchez, 2011) afirma que las tecnologías de datos enlazados son una tendencia que va ganando espacio en muchos ámbitos, incluyendo el interés por desarrollar servicios de Administración Pública y Gobierno Electrónico. De acuerdo al autor, hasta esa fecha se contabilizaban unas treinta y un mil millones de tripletas RDF disponibles en distintos servicios disponibilizados en la web.

De esta manera, la integración de tecnologías semánticas a fuentes de datos ya existentes se ha transformado en un campo activo de investigación y aplicación para diversas disciplinas (Hazber, Li, Gu, y Xu, 2016). Si a lo anterior le sumamos el antecedente de que la mayoría de los datos que están publicados en internet se almacenan en bases de datos relacionales, es posible afirmar que la transformación o adaptación de datos en modelos relacionales a RDF es uno de los principales desafíos para el desarrollo de la Web Semántica y uno de los más activos campos de investigación en la materia (Hert et al., 2011).

De esta manera, muchos investigadores han establecido metodologías y herramientas que proponen cómo convertir datos de una base de datos relacional a un modelo basado en una ontología y descrito en RDF. Incluso la W3C ha creado un grupo de investigación denominado RDB2RDF Incubator Group que busca establecer los estándares para traducir bases de datos relacionales a RDF.

De acuerdo a (Hazber et al., 2016), existen dos mecanismos para efectuar la integración de tecnologías semánticas desde una base de datos relacional: la primera, denominada transformación, es utilizada para extraer una ontología desde una base de datos relacional y la segunda, denominada mapping, efectúa un cruce entre una ontología existente y la base de datos Figure 3.10.



FIGURA 3.10. Transformación vs Mapping.

La transformación consiste de forma genérica en extraer una ontología de una base de datos aplicando nociones de ingeniería inversa. Esta técnica está compuesta por dos procesos: primero se analiza la información para extraer un esquema conceptual y transformar este esquema a su equivalente semántico en una ontología y, en segundo lugar, los registros de la base de datos son migrados al esquema de la ontología extraída.

Por otra parte, el mapping ha sido altamente estudiado y se han desarrollado estándares y herramientas para efectuar esta tarea. Consiste fundamentalmente en enlazar conceptualmente las tablas de una base de datos con las clases de una ontología para realizar una traducción de los datos a RDF.

El segundo caso ha sido objeto de múltiples estudios y se han desarrollado muchas herramientas para efectuar dicho proceso. (Hert et al., 2011) establece una comparación entre varios lenguajes y aplicativos para efectuar el mapping de una base de datos relacional a RDF, entre las cuales podemos destacar:

- **Direct Mapping:** efectúa un mapping directo desde tablas relaciones a las clases de un vocabulario en RDF y los atributos de las mismas tablas a la propiedades del vocabulario.
- **D2RQ:** es un lenguaje declarativo para describir relaciones de manera directa entre un esquema de base de datos relaciones y vocabularios en RDF y ontologías en OWL.
- **eD2R:** es una extensión de D2R y su mapping está basado en consultas SQL que extraen los registros de la base de datos relacional y aplica la transformación a los valores extraídos.
- **Virtuoso RDF Views:** Virtuoso permite exponer datos relacionales en un entorno de web semántica. En el nivel más básico, Virtuoso transforma el resultado de una consulta SQL SELECT en tripletas en RDF.
- **Triplify:** es un modelo ligero para publicar datos vinculados a partir de RDB. Se basa en hacer mapping de las solicitudes de http-URI en consultas RDB y traducir las relaciones resultantes en declaraciones RDF.

- **R2RML:** es un lenguaje para expresar asignaciones personalizadas de bases de datos relacionales a conjuntos de datos RDF. Dichas asignaciones proporcionan la capacidad de ver los datos relacionales existentes en un modelo de datos RDF, expresados en una estructura y vocabulario de destino a elección del autor de mapping.

En resumen, con el creciente interés por incorporar tecnología semántica o de datos enlazados a diversos servicios web, han proliferado las posibilidades para cualquier institución que quiera abrir y enlazar sus datos. Lo anterior no debe quedar ajeno al mundo de la propiedad industrial, debido a que con las tecnologías existentes es posible potenciar los servicios de búsqueda y recuperación de información tecnológica para potenciar los ecosistemas de innovación y desarrollo.

#### **3.4. Integración de tecnologías de la web semántica en propiedad industrial**

Si el interés por incorporar tecnología semántica o de linked data a diversos servicios web ha sido creciente, el caso de la propiedad industrial y el mundo de la patentes no ha quedado excluido, a pesar de sus características particulares.

En primer lugar, el crecimiento exponencial de la producción de patentes a nivel mundial, la complejidad y extensión de los textos, el exceso de terminología legal y técnica, entre otros factores han hecho que los mecanismos de búsqueda, recuperación y análisis de información tecnológica sean crecientes. Sin embargo, actualmente las bases de datos de patentes almacenan la información en bases de datos relacionales, y las exhiben en CVS o XML, dificultando los posibles análisis semánticos (Bermudez-Edo, Noguera, Garrido, y Hurtado, 2013). Por lo tanto, resulta fundamental contar con tecnologías que permitan el acceso, el intercambio y la interoperabilidad entre los datos y recursos de información relacionados a la propiedad industrial debido a que potencian el uso de esta información técnica por parte de empresas y organismos relacionados con la investigación, el desarrollo y la innovación, consolidando este ecosistema. En este contexto surge el rol de las tecnologías semánticas en el mundo de las patentes (Diamantini et al., 2013).

Si bien, en muchos casos algunas oficinas de propiedad industrial y otras empresas relacionadas con el rubro declara ofrecer servicios de búsqueda semántica en sus bases de datos tecnológicas (OMPI, 2012), estas funcionalidades están referidas a búsqueda de sinónimos, variaciones de palabras, términos relacionados, pero no apuntan a tecnologías de linked data basadas en ontologías y RDF.

Según (García Moreno, 2015), a diferencia de los servicios referidos, un servicio basado en linked data resulta prometedor a la hora de soportar la gestión de la I+D+i. Particularmente, en el entorno de la propiedad industrial, contribuye a establecer mejores clasificaciones basadas en una ontología, enriquecer los datos de contenido semántico y promueve su (re)utilización para entregar ventajas competitivas a las empresas o instituciones que requieran su uso. Sin embargo, de acuerdo al mismo autor, el potencial de estas tecnologías se ha tratado principalmente desde el plano de la investigación, pero de manera muy escasa en el plano de la integración a servicios reales.

De acuerdo a lo anterior, si bien las principales iniciativas que integran tecnologías de la web semántica van del lado de las publicaciones científicas (Angrosh, Cranefield, y Stanger, 2014), es preciso destacar ciertas investigaciones y/o aplicaciones que se enmarcan en el contexto de la propiedad industrial y las patentes.

Dentro de las iniciativas más concretas de integración de tecnologías semánticas al mundo de las patentes, destaca GoPatents, un prototipo orientado a la búsqueda de patentes apoyada por la extracción de términos, la categorización de patentes y el uso de ontologías de diversos dominios para establecer relaciones (Eisinger et al., 2014).

Por otra parte, destaca el desarrollo de NEST (“Newly Emerging Science and Technology”), que utiliza técnicas de minería de datos y minería de texto tanto de publicaciones científicas como patentes para identificar información clave sobre actores y tecnologías. Posteriormente, aplicaba una función semántica basada en RDF y una ontología para ejecutar una aplicación basada en TRIZ con la idea de apoyar la resolución compleja de problemas de inventiva (Zhang et al., 2014).

Siguiendo la misma línea, destacan iniciativas de extracción de información semántica a partir de las solicitudes de patentes, con el objetivo de identificar el grado de innovación o novedad de la solicitud. Lo anterior lo desarrolla utilizando extracciones de estructuras frasales en base a una ontología del dominio de la propiedad industrial y presentándolo en un formato legible para máquinas (de Carvalho, França, y Lima, 2014).

Bajo la línea de esta investigación, destaca el trabajo de (Bermudez-Edo et al., 2013), cuyo objetivo es fortalecer la información disponible en repositorios o bases de datos de patentes a través de tecnologías semánticas, ofreciendo nuevas relaciones, propiedades e inferencias. La metodología planteada consiste en procesar consultas en las bases de datos y convertir los archivos XML a un formato OWL.

Por otro lado, destaca el desarrollo de ciertas ontologías aplicadas al dominio de la I+D+i como GI2MO (“Generic Idea and Innovation Management Ontology”) basada en RDF y OWL, que incorpora terminología de distintas fuentes y tiene como objetivo vincularse con otros vocabularios estándar como FOAF (“Friend of a Friend”) o DCTerms, o la ontología IDEA, basada en OWL, que también reutiliza otros vocabularios de uso común para fortalecer la gestión de ideas de innovación (García Moreno, 2015).

Particularmente en el caso de las patentes, destaca el desarrollo del grupo de investigación “Agile Knowledge Engineering and Semantic Web” (AKSW), quienes crearon una ontología para el contexto de la oficina de patentes de los Estados Unidos (USPTO), la cual denominaron “US Patents”. Dicha ontología incluye distintas clases y propiedades relacionadas con los documentos de patentes de la USPTO y es utilizada para generar tripletas en RDF que puedan nutrir de contenido semántico a dicha documentación.

En base a lo expuesto, la necesidad por optimizar las bases de datos de patentes es notoria, y las tecnologías semánticas suponen una oportunidad con mucho potencial, a pesar de las características heterogéneas y complejas de que pueden presentar estas fuentes de información. Sin embargo, su aplicación no ha sido una práctica masiva y eso da cabida a que sea un campo de investigación con mucho potencial.

## 4. METODOLOGÍA - INTEGRACIÓN DE TECNOLOGÍAS SEMÁNTICAS

Como se ha expuesto previamente, la integración semántica en sistemas de información se ha transformado en un nicho interesante de investigación para múltiples disciplinas y han sido muchos los autores que han generado diversas metodologías y herramientas para efectuar dicho proceso.

Para el presente caso de estudio, se busca establecer una estrategia metodológica de traducción de una base de datos relacional a un modelo semántico basado en el mapping (Hazber et al., 2016), utilizando estándares de la W3C como R2RML y una ontología para el contexto de la propiedad industrial desarrollada por el grupo de investigación “Agile Knowledge Engineering and Semantic Web” (AKSW). Dicha estrategia consta de cinco etapas, las cuales buscan dar respuesta a los objetivos planteados previamente:



FIGURA 4.1. Diseño metodológico de Integración Semántica.

### 4.1. Análisis de bases de datos de INAPI

El primer paso para la integración de tecnologías semánticas en la base de datos de patentes de INAPI consiste en efectuar un análisis del modelo de datos de INAPI. Lo anterior contribuye a identificar los elementos y procedimientos relevantes que permiten estructurar el mapping o la traducción de una base de datos relacional a RDF con todos los estándares y procedimientos necesarios.

Dicho análisis se divide en dos fases:

#### 4.1.1. Análisis del modelo de datos de INAPI

En primera instancia, es preciso identificar en el modelo de datos de INAPI cuáles son los elementos relevantes y susceptibles a ser vinculados en un contexto de web semántica.

En base a lo anterior, el modelo de datos que compone la base de datos de patentes de INAPI permite identificar y describir la estructura de los datos que componen dicha base y la forma en que estos se relacionan entre sí (Figure 4.2).

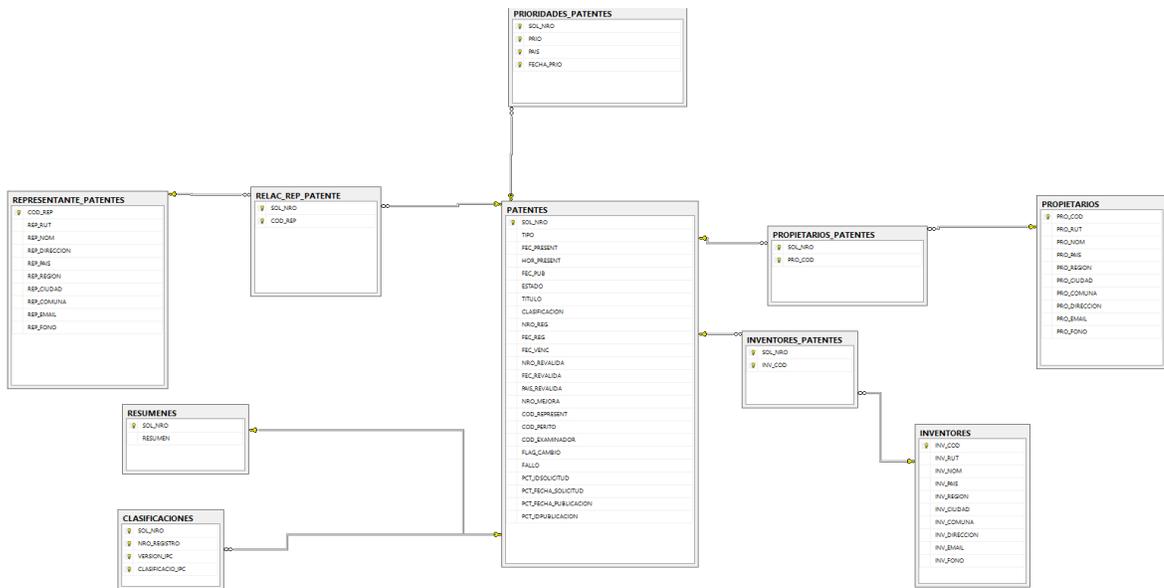


FIGURA 4.2. Modelo de datos de la base de datos de INAPI.

El modelo anterior permite identificar al menos cuatro entidades fundamentales en el ecosistema de la propiedad industrial: Documento de patente, inventor, propietario/solicitante y representante.

- **Documento de Patente:** en primer lugar, el modelo define a la entidad “Patente” como su elemento principal con una serie de campos descriptivos que resultan relevantes para identificar y caracterizar un documento de patente. Entre dichos campos destaca el número de solicitud, un código numérico que permite identificar las patentes en el modelo de datos expuesto. Cada patente tiene un número de

solicitud y el modelo de datos relaciona a las otras entidades con la patente mediante su número de solicitud. Por lo tanto, se define como el identificador de cada patente.

Por otro lado, el modelo de datos presenta una serie de campos descriptivos de la patente que entregan información de valor como el título, fechas (publicación, solicitud, vencimiento), estado, tipo de documento, país, entre otras.

El modelo de datos desarrollado por INAPI (Figure 4.2) presenta el resumen, la prioridad y la clasificación internacional (CIP) como tablas separadas a la patente. En primera instancia, el número de prioridad sirve para identificar la patente original si esta tiene una familia de patentes, entonces podría ser considerado un campo más dentro de la tabla “patentes”. En segunda instancia, el resumen es un elemento del documento de patente, por lo tanto, no se justifica que esté separado de la entidad “patentes”. Finalmente, la clasificación internacional (CIP) también podría ser considerado un campo descriptivo de una patente, pero de igual manera entrega información estructurada y jerarquizada sobre el tipo de tecnología que describe la patente, lo que podría justificar su condición de entidad separada.

- **Inventores:** el modelo define a los “Inventores” con toda la información descriptiva de cada inventor de una patente. Entre los campos relevantes que exhibe el modelo para identificar a un inventor es posible identificar su código de identificación – único para cada individuo -, su RUT, nombre, país e información de contacto. Es preciso considerar que puede haber más de un inventor en cada documento de patente.
- **Propietarios:** de manera muy similar con el caso de los inventores, el modelo define la tabla “Propietarios” con toda la información descriptiva de cada solicitante de una patente. Entre los campos relevantes que exhibe el modelo para identificar a un solicitante es posible identificar su código de identificación – único para cada propietario -, su RUT, nombre, país e información de contacto. Cabe señalar que un inventor puede ser a su vez propietario de una patente, por lo que podría coincidir el dato.

- **Representantes:** finalmente, el último actor relevante identificado en el modelo es la tabla “Representante” con toda la información descriptiva de cada representante de una patente. Los campos relevantes son similares a los otros actores, presentando un código de identificación – también único para cada representante –, su RUT, nombre, país e información de contacto.

#### 4.1.2. Levantamiento de base de datos en MySQL

Una vez expuesto y analizado el modelo de datos de INAPI, es necesario tener constituida la base de datos relacional para hacer la traducción a un modelo de datos en RDF. Para efectos de la presente investigación y el modelamiento a efectuar, se levantó una base de datos relacional en MySQL (Figure 4.3) respetando toda la estructura de la base de datos de patentes original de INAPI.

Tabla	Acción	Filas
<input type="checkbox"/> clasificaciones	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> inventores	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> inventores_patentes	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> patentes	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> prioridades_patentes	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> propietarios	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> propietarios_patentes	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> relac_rep_patente	Examinar Estructura Buscar Insertar Vaciar Eliminar	
<input type="checkbox"/> representante_patentes	Examinar Estructura Buscar Insertar Vaciar Eliminar	

FIGURA 4.3. Réplica de base de datos en MySQL.

La base de datos relacional resultante se compone de 10 tablas entre las que es posible identificar:

- **Clasificaciones:** esta tabla exhibe datos descriptivos de la CIP o Clasificación Internacional de Patentes, permitiendo relacionar un documento de patente con uno o más códigos de clasificación.

- **Inventores:** la tabla inventores muestra los datos de los inventores, como por ejemplo, su código de identificación, RUT, nombre, y otros datos personales, de acuerdo a lo establecido en el modelo inicial.
- **Inventores-patentes:** esta tabla establece relación entre los inventores y un documento de patente mediante el número de solicitud de la patente y el código único de identificación.
- **Patentes:** la tabla patentes ofrece información descriptiva de los documentos de patentes con campos como el número de solicitud, título, estado, entre otros elementos.
- **Prioridades-patentes** la tabla de prioridades exhibe los datos relacionados con número y fecha de prioridad de la patente y se relacionan con cada documento mediante el número de solicitud.
- **Propietarios:** de manera similar a la tabla de inventores, esta tabla muestra la información de los propietarios o solicitantes integrando información como código de identificación, RUT, nombre, entre otros.
- **Propietarios-patentes:** esta tabla establece relación entre los propietarios y los documentos de patente mediante el código de identificación del propietario y el número de solicitud de la patente.
- **Representante-patentes:** la tabla de representante define la información descriptiva de los representantes legales de la patente usando como datos un código de identificación, RUT, nombre, entre otros.
- **Relac-rep-patentes:** esta tabla establece relación entre los representantes y los documentos de patentes a través del código de identificación de los primeros y el número de solicitud de los segundos.

En relación a la estructura de datos original de la base de datos de INAPI anteriormente expuesta, se resuelve no crear la tabla “Resúmenes” debido a que corresponde a un campo bibliográfico del documento de patente. Por lo tanto, el resumen se incorpora como campo descriptivo de la tabla “Patentes”.

## 4.2. Definición de Ontologías

Para iniciar la segunda fase del modelo de integración de tecnologías semánticas (LOD) a la base de datos de patentes de INAPI (datos.inapi.cl), es necesario definir la ontología que ayudará a establecer las relaciones semánticas a los datos obtenidos de la base de datos MySQL.

En primera instancia, es posible desarrollar una ontología propia o utilizar alguna pre-existente para adaptarla al contexto en el cual será usada. Para el caso de la propiedad industrial, el grupo de investigación Agile Knowledge Engineering and Semantic Web (AKSW)<sup>1</sup> desarrolló un proyecto basado en la información de la Oficina de Patentes y Marcas de los Estados Unidos (USPTO), quienes almacenan y presentan sus documentos de patentes, con el objetivo de fortalecer la indexación y el proceso de búsqueda transformando sus datos de XML a RDF para ajustarlos a los principios de los datos enlazados. Para cumplir los objetivos de este proyecto, AKSW desarrolló una ontología denominada “US Patents Ontology”, basada en el modelo de datos de la USPTO para permitir la transformación de los datos a RDF.

La naturaleza estandarizada y normalizada de los documentos de patentes permiten que la “US Patents Ontology” de AKSW tenga muchos elementos aplicables al contexto de propiedad industrial en Chile. Por ejemplo, elementos relevantes como el título, el número de publicación, inventores, solicitantes, representantes, clasificaciones internacionales, entre otros, son replicables en el modelo de datos de patentes de INAPI.

De esta manera, dicha ontología toma un rol protagónico en la traducción de datos de una base de datos relacional a RDF, sin embargo, no es la única ontología que potencialmente puede ser utilizada para el contexto de propiedad intelectual. De acuerdo a lo analizado en el modelo de datos de patentes, los inventores, representantes y solicitantes toman un rol muy relevante en el entorno de la propiedad intelectual y pueden ser representados por FOAF (Friend of a friend), una ontología destinada a describir personas, sus

---

<sup>1</sup>Agile Knowledge Engineering and Semantic Web (AKSW) es un grupo de investigación dedicado a desarrollar métodos, herramientas y aplicaciones en el contexto de la Ingeniería del Conocimiento y la Web Semántica.

características, relaciones e intereses. Es fundamental considerar que, en la medida que se integren este tipo de ontologías, será más factible enlazar datos con otras fuentes que también las utilizan. De la misma manera, asumiendo la relevancia de aspectos como la territorialidad en materia de propiedad industrial, es posible utilizar otras ontologías como Geonames, para describir información geoespacial de manera semántica. En la medida que el modelo tienda a ser más complejo y aspire a enlazar más fuentes de datos, puede integrar más ontologías.

Como elemento delimitador del estudio, el modelo de datos se enfocará en abrir y relacionar datos sobre las patentes propiamente tal, los inventores y los solicitantes. Lo anterior supone el uso de algunos elementos de FOAF para describir personas, pero la ontología central será la de Patentes de AKSW.

Para efectuar de manera eficiente el proceso de modelamiento y traducción de la base de datos de patentes a un modelo RDF, es necesario analizar y comprender la estructura de la ontología a utilizar en el procedimiento. Lo anterior puede ser abordado utilizando algún software de gestión de ontologías como Protégé.

Protégé es un software libre y de código abierto desarrollado por la Universidad de Stanford que permite crear, mantener, analizar y editar ontologías. Según (Musen, 2015), Protégé se ha transformado en el software para construir y mantener ontologías más usado del mundo y, hasta la fecha de su estudio, registraba más de 250.000 personas que han usado el programa.

De acuerdo a lo expuesto por Musen, Protégé resulta un software idóneo para la edición, mantención o el análisis descriptivo de una ontología. Sus características de multiplataforma, su condición de libre y código abierto y la masificación en su uso son componentes que permiten su elección como software para estudiar la ontología de AKSW. Desde el plano práctico, Protégé entrega datos de la ontología como sus relaciones, axiomas, entre otros, que posteriormente resultan útiles para efectuar el modelamiento de forma eficiente.

Actualmente, Protégé presenta diversas versiones. Por un lado, está la versión de escritorio que presenta diversas características que permiten la construcción y mantención

de ontologías en OWL y, por otro lado, un sistema basado en web denominado WebProtégé al cual se accede mediante un navegador y resulta más sencillo de usar para tareas de ingeniería en ontologías.

Para analizar la ontología de patentes desarrollada por AKSW, se utilizará Protégé en su versión 5.1.0 de escritorio.

En primer lugar, es preciso descargar la ontología del sitio web de AKSW<sup>2</sup> en formato OWL e importarla en el software Protégé. Una vez incorporada en el sistema, es posible identificar datos relevantes de la ontología (Figure 4.4), por ejemplo, la cantidad de clases y propiedades que presenta la ontología. Para este caso se identifican 19 clases, 17 propiedades de objetos (object property) y 20 propiedades de tipos de datos (data property).

<b>Class count</b>	<b>19</b>
<b>Object property count</b>	<b>17</b>
<b>Data property count</b>	<b>20</b>

FIGURA 4.4. Descripción de elementos de la ontología de AKSW (Class, object property y data property).

La ontología presenta 18 clases correspondientes al entorno de la propiedad industrial y una clase que opera como propiedad de anotación de axiomas. Es posible graficar las clases de la ontología de AKSW mediante la herramienta de visualización OntoGraf de Protégé (Figure 4.5) y se pueden enlistar y describir de la siguiente manera:

- **Agent:** corresponde a la persona o entidad que toma el lugar del solicitante en la tramitación de la patente.
- **Applicant:** representa a la persona o entidad que solicita la patente.
- **Assignee:** representa a la persona o entidad titular de la patente, es decir, que posee los derechos de propiedad.
- **Citing:** corresponde a la información de citación de una patente a otra.
- **City:** esta clase representa a la ciudad.

<sup>2</sup>Ontología disponible en el sitio web de Agile Knowledge Engineering and Semantic Web (AKSW) <http://aksw.org/Projects/USPatents.html>.

- **ClassificationInternational:** corresponde a la clasificación internacional de la Organización Mundial de Propiedad Intelectual (OMPI).
- **ClassificationNational:** corresponde a la clasificación nacional de las patentes de Estados Unidos.
- **Country:** representa información del país incluyendo el código internacional del país.
- **Examiner - Examiner-Assistant:** esta clase corresponde a la persona encargada de comprender las especificaciones técnicas de la invención y verificar la patentabilidad de la solicitud. Presenta una relación de equivalencia con la clase “Examiner-Assistant”.
- **Examiner-Assistant - Examiner:** corresponde a la persona encargada de verificar el cumplimiento de los requisitos a una solicitud de patente. Esta persona tiene un rol secundario en relación al examinador. Presenta una relación de equivalencia con la clase “Examiner”.
- **Patent:** esta clase representa la entidad patente (documento de patente), es la clase principal de la ontología.
- **Inventor:** esta clase corresponde a la entidad inventor de una patente, referida al creador de la invención.
- **Nationality:** esta clase corresponde a la nacionalidad de una persona u de una organización.
- **Kind:** esta clase indica la tipología documental a la que corresponde una patente de invención.
- **Orgname:** esta clase representa la organización de los datos descrita en una patente.
- **Residency:** esta clase representa el lugar de residencia de una persona o de una entidad.
- **Role:** esta clase representa al rol asignado a una patente por parte de un examinador determinado.

- **State:** esta clase representa a un Estado en el contexto norteamericano y su código internacional.

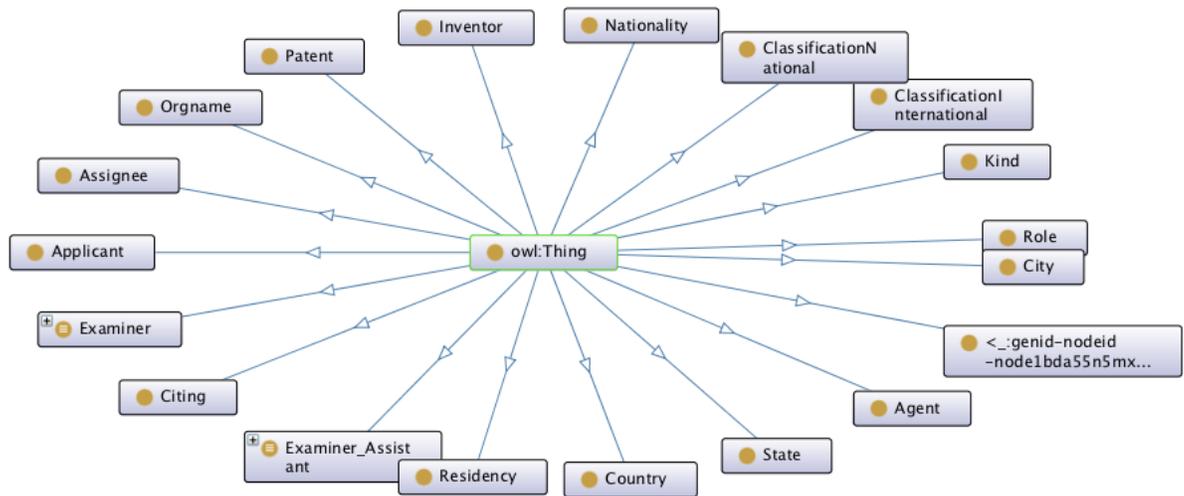


FIGURA 4.5. Visualización gráfica de la ontología de AKSW.

Las principales clases de la ontología resultan apropiadas para representar la información del entorno de la propiedad intelectual en Chile, ya que se adecuan a los principales elementos normalizados y estandarizados internacionalmente.

Por otra parte, una ontología presenta una serie de propiedades tanto de objetos (object property) como de tipo de datos (data property). Es importante resaltar que las object property relacionan individuos concretos pertenecientes a una clase y las data property relacionan a los individuos de una clase con un tipo de datos (Por ejemplo, un número entero, un valor literal, etc.).

En el caso de la “US Patent Ontology”, las propiedades de objeto establecen las distintas relaciones entre las entidades de la ontología. Por ejemplo, la relación de la entidad país con otras entidades como inventor, solicitante, patente o examinador (Figure 4.6). Asimismo, las propiedades presentan atributos que determinan si es reflexiva, simétrica, funcional o funcional inversa.

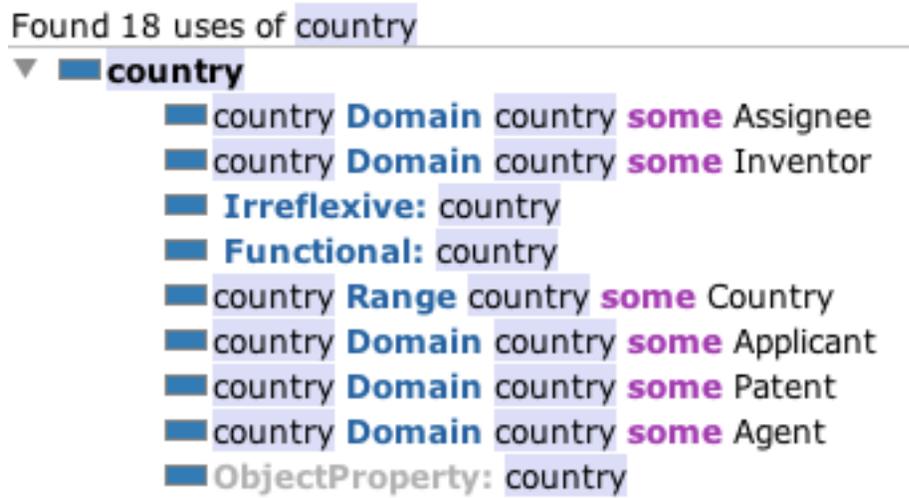


FIGURA 4.6. Propiedades de objeto (Object property) de la clase Country.

Por otra parte, las data property o propiedades de datos definen el valor de un tipo de dato que se corresponderá con un individuo de una clase. Por ejemplo, el elemento “año” de una patente corresponde a un número entero. Asimismo, en las propiedades de datos se puede hacer cruce con elementos de otras ontologías como FOAF, que permite identificar personas como inventores, examinadores o solicitantes mediante sus propiedades de datos.

El identificar las clases y sus propiedades resulta fundamental para establecer los parámetros sustanciales que determinarán la traducción o el mapping de la base de datos relacional a un modelo RDF, a través de la estructura de la ontología.

### 4.3. Mapping / Modelamiento de datos

Una vez identificados los principales elementos de la ontología, es preciso hacer la traducción de la base de datos relacional a un esquema o modelo de datos en RDF. Existen muchos procedimientos - más o menos complejos - que permiten publicar datos en RDF directamente desde una base de datos relacional. Por ejemplo, existen herramientas como D2R que permiten convertir fácilmente la estructura de una base de datos en RDF, aunque sin un fuerte componente de descripción de contenidos semánticos. Sin embargo, también es posible efectuar la traducción utilizando como estructura base una ontología dada, lo

que entrega un alto potencial semántico. Lo anterior permite generar tripletas en RDF vinculadas a un ontología, que puede ser creada o puede ser reutilizada una ya existente para describir ciertas fuentes de datos de un determinado dominio de conocimiento (Knoblock et al., 2012).

La conversión a RDF es un paso indispensable para publicar fuentes de datos enlazados y los mecanismos antes expuestos permiten hacer la conversión con la semántica subyacente de manera explícita. Asimismo, existen otros sistemas como el R2R o el R2RML que definen lenguajes para traducciones específicas entre distintas fuentes de datos. Para este caso, se utilizará una herramienta denominada “Karma” que trabaja la integración de datos y también opera como ETL.

Karma es un software de código abierto que permite a los usuarios integrar de manera sencilla una gran variedad de fuentes de datos como bases de datos, XML, hojas de cálculo, CVS, entre otras, utilizando como elemento fundamental una ontología definida por el usuario.

La versatilidad de Karma al momento de hacer integración de datos resulta fundamental para transformar o publicar datos en RDF en base a una ontología utilizando R2RML. Karma es un software que permite hacer este proceso de manera semi-automatizada (Knoblock et al., 2012). Adicionalmente, Karma presenta una serie de características que ayudan a justificar su utilización para la integración de datos desde otras fuentes:

- **Facilidad de uso:** Karma lleva todo el proceso de traducción a una interfaz intuitiva para el usuario.
- **Diversidad de fuentes:** A diferencia de otros sistemas, Karma no solo permite hacer mapping desde fuentes tabulares mediante una ontología, sino también otro tipo de fuentes como JSON o XML.
- **Uso de ontologías:** Permite integrar fácilmente una o más ontologías al proceso, junto con almacenar los modelos semánticos.

- **Proceso semi-automatizado:** Permite al usuario la realización semi-automatizada del proceso, permitiéndole intervenir los elementos que el usuario quiere ajustar en el modelado.
- **Facilidad de publicación:** Permite la publicación a RDF (y otros formatos) de manera rápida e intuitiva.
- **Multiplataforma y software libre:** Karma funciona en distintos sistemas operativos, además de ser un software libre y de código abierto.

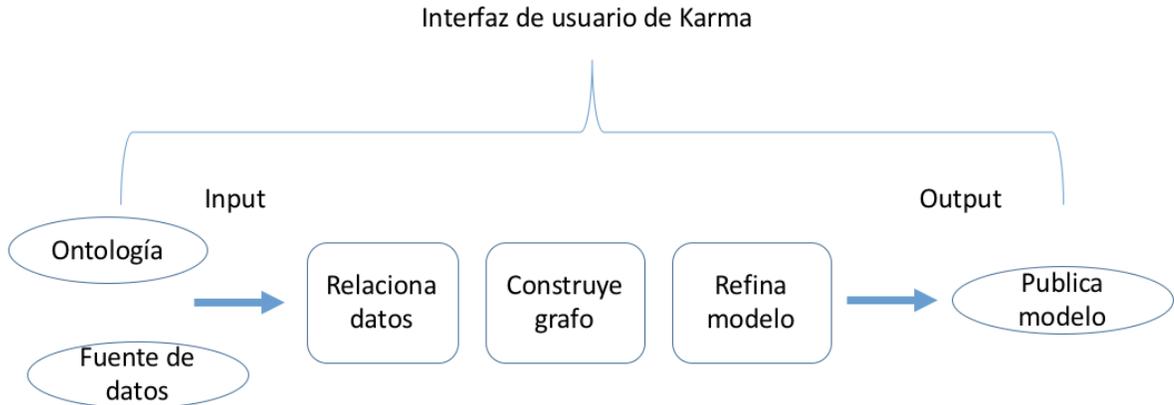


FIGURA 4.7. Proceso de Karma para modelar fuentes de datos.

Como lo demuestra la imagen (Figure 4.7), en primera instancia es preciso integrar al software la fuente de datos y una ontología a elección del usuario que permite el modelado a través de una interfaz gráfica que automatiza gran parte del proceso. Posteriormente, Karma realiza el mapeo de datos de acuerdo a las clases de la ontología y propone un modelo para vincular estas clases. Luego, Karma construye un grafo que exhibe las posibilidades de mapeo entre la fuente y la ontología para finalmente refinar el modelo. Todo lo anterior puede ser ajustado y refinado en la interfaz de usuario. Finalmente, una vez que el modelo está completo, Karma puede publicar los datos integrados como RDF u otro formato a elección del usuario.

De acuerdo a (Knoblock et al., 2012), algunas veces las fuentes de datos necesitan una operación de limpieza y transformación de los datos como parte de la traducción y esas operaciones son fácilmente aplicables en la interfaz de usuario de Karma.

La traducción de la base de datos de patentes a un modelo en RDF será descrita en 6 pasos fundamentales: 1) Input o integración de ontología y fuente de datos, 2) Definición de URIs 3) Vinculación de clases de la ontología con la fuente de datos, 4) Construcción del grafo, 5) Refinamiento del modelo y 6) Publicación del modelo. La generación de datos en RDF será objeto del siguiente apartado.

#### **4.3.1. Input o integración de ontología y fuente de datos.**

Como primer elemento para efectuar el mapeo en Karma es necesario hacer el ingreso de los datos y de la ontología al software.

En primera instancia, se debe incorporar la ontología al software en un formato adecuado como OWL. De esta manera, Karma puede identificar las clases, las propiedades de objeto y las propiedades de datos de la ontología integrada, las cuales serán fundamentales para el siguiente paso del mapeo de datos. Para este caso de estudio, es preciso importar la ontología desarrollada por AKSW, “US Patent Ontology”.

Por otra parte, es preciso hacer el ingreso de la fuente de datos, funcionalidad para la que Karma presenta una gran versatilidad. Por un lado permite el ingreso de archivos, servicios web, uso de SQL y bases de datos.

La base de datos de patentes está diseñada en MySQL, por lo tanto, la opción de importe de base de datos es la adecuada para el contexto. En dicha opción es preciso explicitar los siguientes datos: Tipo de base de datos (Oracle, MySQL, Postgis, Sybase), Hostname, puerto, usuario, contraseña y nombre de la base de datos (Figure 4.8).

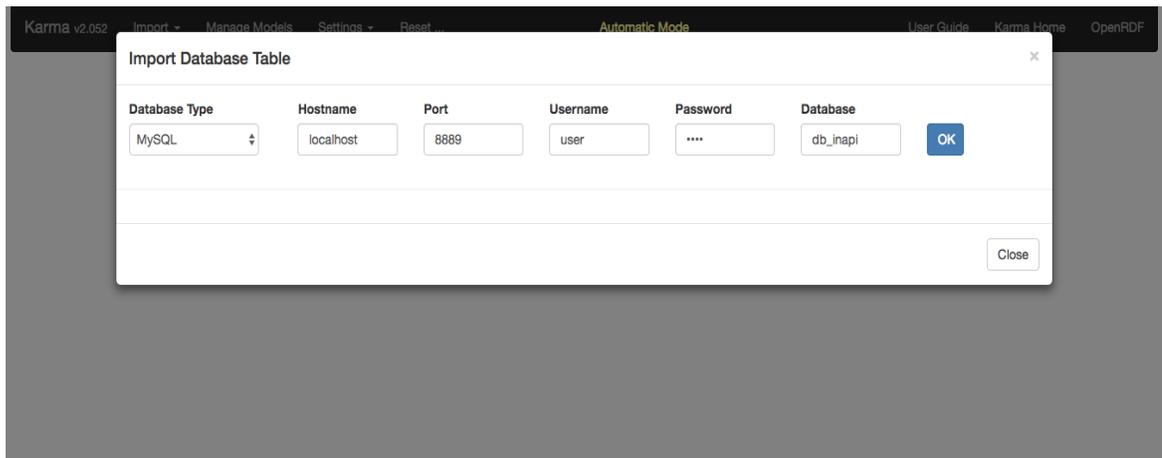


FIGURA 4.8. Importación de base de datos a Karma.

Una vez importada la base de datos, el sistema permite elegir las tablas que se quieren importar para iniciar el modelamiento. Para efecto de este estudio, se trabajará con los elementos de la tabla “patente”, incorporando los descriptores de “inventor”, “propietario” y “resumen”. Dichas columnas de la tabla son importadas desde MySQL a Karma para iniciar las relaciones entre los datos y las clases de la ontología desarrollada por AKSW.

#### 4.3.2. Definición de URIs

Para este proceso es fundamental reconocer a una URI como un identificador de algún recurso. Como lo plantea (Gayo, 2012), un recurso puede ser un elemento del mundo digital como un documento, imagen o vídeo o un elemento del mundo físico como una persona, un país o incluso un concepto abstracto.

Para el caso del modelo de datos propuesto, se trabajará con una URI principal como identificadora de recursos en el ámbito de las Patentes y dos URIs anexas, que permitan identificar inventores y solicitantes para potencialmente relacionar los datos con otras fuentes. Estas URIs serán generadas basándose en los identificadores o IDs que presenta la estructura de la base de datos MySQL original:

- **URI de Documento de Patente:** Identifica un documento de patente como recurso de información. En la base de datos del modelo relacional, el atributo clave que

identifica cada patente es el número de solicitud. De esta manera, se propone el uso de este elemento como identificador de cada documento de patente, quedando estructurada de la siguiente manera:



FIGURA 4.9. Uri de documento de patente.

- **Uri del inventor:** Identifica un inventor como recurso. En la base de datos original existe un código único de identificación de un inventor (ID de inventor) que se propone como elemento identificador, dando lugar a una URI como describe el siguiente ejemplo:

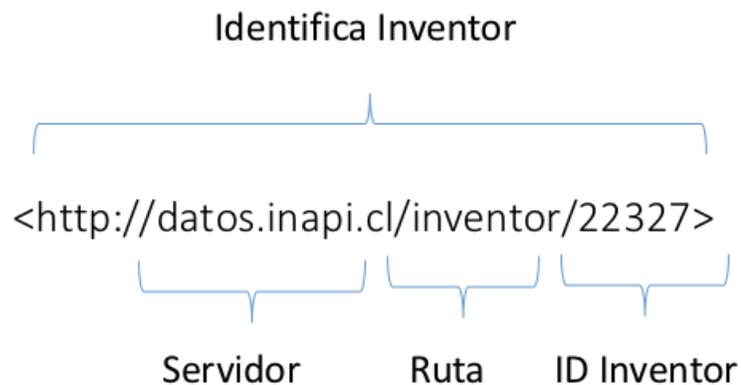


FIGURA 4.10. Uri del inventor.

- **Uri del propietario:** Identifica a un propietario como recurso. En la base de datos original existe un código único de identificación de un propietario (ID de propietario) que se propone como elemento identificador, dando lugar a una URI similar a la del ejemplo:

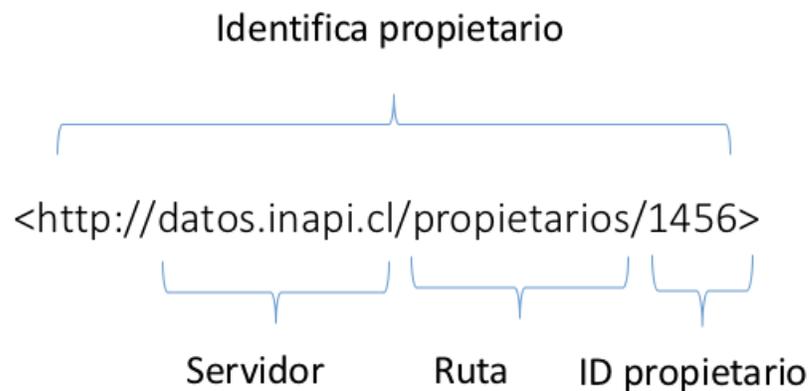


FIGURA 4.11. Uri del propietario.

Los elementos antes descritos, además de establecerse como identificadores únicos de los recursos, permiten enlazar datos con otras fuentes que también utilizan identificadores similares. Por ejemplo, si tenemos definida la URI de un propietario o solicitante como Codelco, podemos enlazar ese dato con otra fuente que entregue información más detallada sobre esa empresa, por ejemplo, un vínculo con DBPedia. De este modo, los datos del modelo expuesto originalmente, se hacen más complejos porque están abiertos y estructurados, lo que permite que se conecten con otras fuentes de datos para enriquecer el sentido de dicho dato. La imagen (Figure 4.12) expresa el rol o la relevancia de la URI en la vinculación de datos con fuentes como DBPedia o cualquier otra.

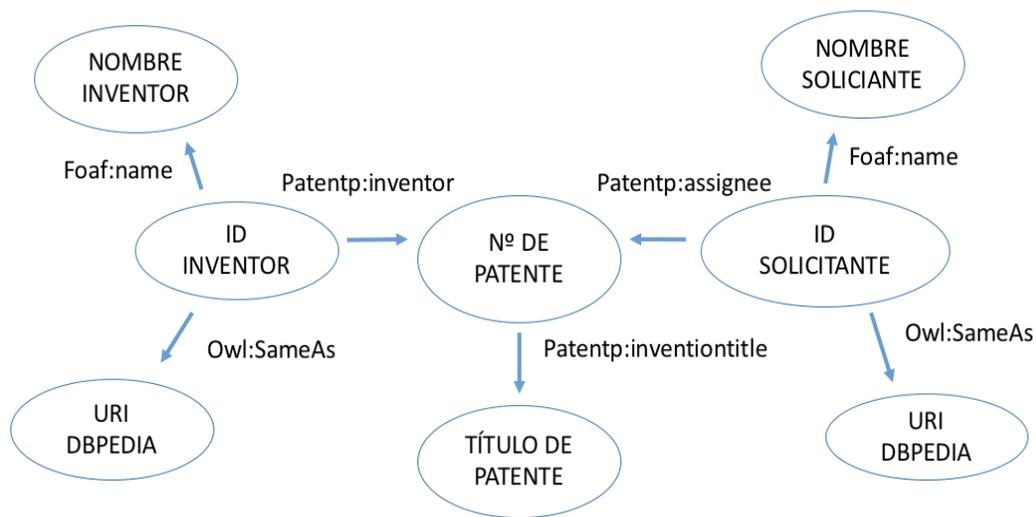


FIGURA 4.12. Relación de URIs con otras fuentes.

Para ejemplificar el esquema, podemos pensar en una patente donde su solicitante es Codelco. Identificando al solicitante y su relación con la patente, podemos relacionar la información de dicha invención con la información que tiene DBPedia sobre Codelco (Figure 4.13).

**Patente**

N Solicitud: 199501028      Fecha Solicitud: 14/12/1995  
 Fecha de Publicación: 29/08/1996      Estado: Caducada  
 N Registro Patente: 39672      Fecha Registro: 27/07/1998  
 Tipo de Solicitud: Patente de Invención  
 Prioridad:  
 Clasificaciones IPC: C22B1/02  
 C22B34/34  
 C22B4/00

**Título o Materia de la Solicitud**  
 PROCEDIMIENTO PARA LA PRODUCCION DE TRIOXIDO DE MOLIBDENO POR TOSTACION DE MOLIBDENITA MEDIANTE MICROONDAS

**Resumen**  
 PROCEDIMIENTO PARA LA OBTENCION DE TRIOXIDO DE MOLIBDENO CUYO CAMPO DE APLICACION ES LA METALURGIA DEL MOLIBDENO Y QUE RESUELVE EL PROBLEMA DE LA TECNICA DELA OBTENCION DE TRIOXIDO DE MOLIBDENO DE GRADO METALURGICO POR TUESTA OXIDANTE DIRECTA DE MOLIBDENITA CON O SIN PURIFICACION PREVIA. LA INVENCIÓN CONSISTE EN SOMETER LOS CONCENTRADOS DE MOLIBDENITA, CON O SIN PURIFICACION PREVIA, DE TAMAÑO 100% INFERIOR A 0,1 CM, COLOCADOS COMO UNA CAPA DE 2 A 10 MM DE ESPESOR, A LA ACCION DE MICROONDAS QUE TIENEN UNA FRECUENCIA COMPRENDIDA ENTRE 1.500 MHZ A 3.500 MHZ, DURANTE UN TIEMPO COMPRENDIDO ENTRE 2 A 60 MINUTOS, CON UNA POTENCIA ESPECIFICA DE 1 A 10 WATT/CM2 DANDO COMO RESULTADO UN TRIOXIDO DE PUREZA SUPERIOR AL 99%

**Solicitante**  
 Nombre/Razón Social: CORPORACION NACIONAL DEL COBRE DE CHILE CODELCO-CHILE  
 País: CHILE

---

**About: Codelco**  
 An Entity of Type: Empresa pública, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

La Corporación Nacional del Cobre de Chile (Codelco) es una empresa estatal chilena dedicada a la explotación minera cuprífera, rubro en el que es la mayor compañía del planeta. Codelco opera ocho centros de trabajo, ubicados entre la Región de Antofagasta y la Región del Libertador General Bernardo O'Higgins; la Casa Matriz se encuentra en Santiago. Empresa 100% estatal, posee el mayor nivel de reservas y recursos de cobre conocidos en el planeta. Se estima que, con los actuales niveles de producción, los yacimientos que hoy explota proyectan más de 65 años de vida útil.

Property	Value
owl:abstract	<ul style="list-style-type: none"> <li>La Corporación Nacional del Cobre de Chile (Codelco) es una empresa estatal chilena dedicada a la explotación minera cuprífera, rubro en el que es la mayor compañía del planeta. Codelco opera ocho centros de trabajo, ubicados entre la Región de Antofagasta y la Región del Libertador General Bernardo O'Higgins; la Casa Matriz se encuentra en Santiago. Codelco es el productor de cobre más grande del mundo y la empresa que contribuye más a la economía chilena, con ventas por US\$ 13.827 millones en 2014. Su producción anual equivale al 11% de la producción mundial de cobre de mina y un 34% de lo producido en Chile, con 1.796.000 toneladas de cobre fino en 2011 (incluida su participación en el yacimiento El Araj). Es también una de las primeras compañías en producción de molibdeno, con 23.098 toneladas métricas finas ese mismo año. Empresa 100% estatal, posee el mayor nivel de reservas y recursos de cobre conocidos en el planeta. Se estima que, con los actuales niveles de producción, los yacimientos que hoy explota proyectan más de 65 años de vida útil. Cabe para el desarrollo de Chile, desde 1971 Codelco ha aportado más de 98.400 millones de dólares al Estado. De estos excedentes, 48.067 millones de dólares se entregaron en el período 2004-2010; y US\$</li> </ul>

FIGURA 4.13. Potencial relación entre datos de una patente con DBPedia (Ejemplo de Codelco).

### 4.3.3. Vinculación de clases de la ontología con la fuente de datos

Una vez incorporada la fuente de datos y la ontología a Karma, es necesario establecer relaciones entre las distintas columnas de la base de datos y las clases de la ontología. Lo anterior contribuye a nutrir de semántica a la fuente de datos.

Para efectuar este proceso, es necesario establecer el tipo semántico de cada elemento de la fuente de datos, considerando la clase, las propiedades de la clase y el tipo de dato al que corresponde.

En primera instancia, es necesario establecer un vínculo de cada columna con alguna clase a la que correspondan sus datos, como se describe en los siguientes ejemplos:

- La columna “sol-nro” de la base de datos corresponde a un número de solicitud de una patente y puede ser relacionada con la clase “Patent” de la ontología.
- La columna “inventor” de la base de datos corresponde al nombre del inventor de una o más patentes y puede ser relacionada con la clase “Inventor”.
- La columna “clasificación” de la base de datos corresponde al número de Clasificación Internacional de Patentes (CIP) y puede ser relacionada con la clase “ClassificationInternational”.

Si bien, un documento de patente se caracteriza por ser altamente estructurado y normalizado, es preciso considerar que en la tabla de la base de datos de INAPI existen columnas más relacionadas con la gestión propia de la institución y no con el estándar internacional de descripción de una patente, como “hora de presentación”, “fecha de vencimiento”, “Nro de revalidación”, “fecha de revalidación”, “fallo” o “código de perito”. Dichos campos de la tabla pudiesen ser incorporados de manera forzada a la ontología, sin embargo, para efecto de este estudio, y con el objetivo de obtener como resultado un RDF pulcro, serán incorporados solo los elementos genéricos de descripción de una patente a nivel internacional, y serán relacionados de la siguiente manera:

- N° de Patente (Número de solicitud) → Clase Patent.
- Fecha de solicitud → Clase Patent.

- Fecha de presentación → Clase Patent
- Título de la patente → Clase Patent.
- Inventor → Clase Inventor
- IdInventor → Clase Inventor
- Propietario o solicitante → Clase Assignee
- IdPropietario → Clase Assignee
- Representante legal → Agent
- IdRepresentante → Agent
- N° de clasificación internacional (CIP) → Clase InternationalClassification
- Resumen → Clase Patent
- País → Clase Country

Por otra parte, elementos de otras tablas de la base de datos como la “prioridad” (Número o fecha de prioridad) podrían ser incorporados a la traducción si los datos de INAPI se estructuran de manera plana en una sola tabla para migrar a un modelo en RDF, sin embargo, en este trabajo no serán considerados para no modificar en demasía la estructura original del modelo de la base de datos.

Una vez identificadas las relaciones de la fuente de datos con la ontología, es preciso determinar las propiedades de la clase y los tipos de datos que resultarán los literales. Ambos elementos están explícitos en la ontología y pueden ser previamente captados desde la información que ofrece el software Protégé para reasignarlos de manera correcta en Karma.

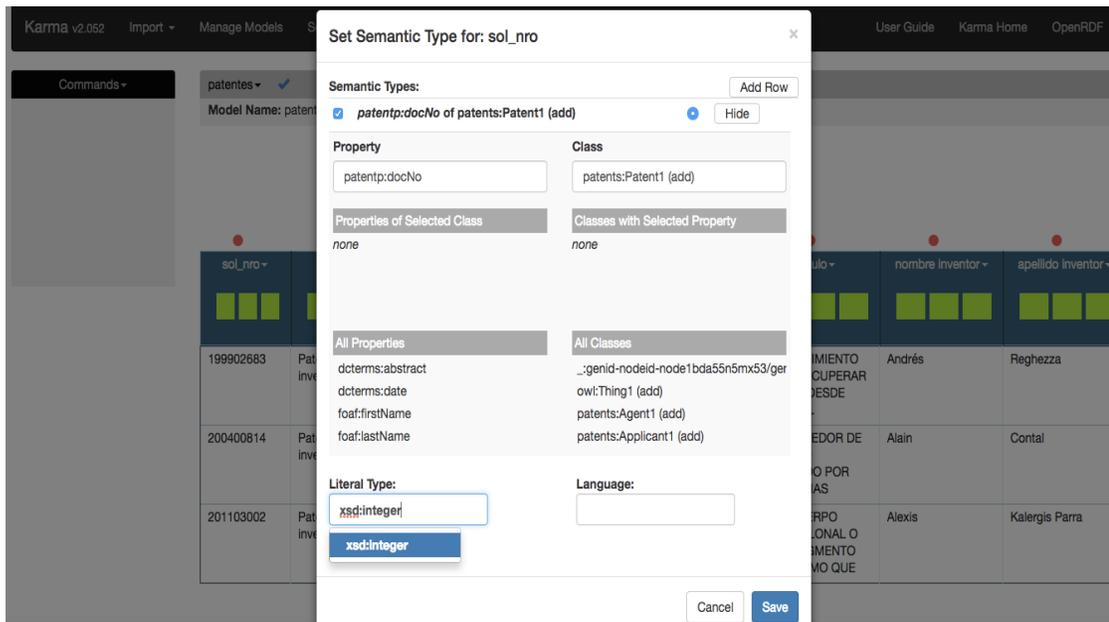


FIGURA 4.14. Interfaz para establecer clase, propiedad y tipo de literal.

Por una parte, Karma ofrece la posibilidad de hacer el proceso de manera automatizada, aunque con el riesgo de que no establezca de manera adecuada las relaciones y, por otro lado, como muestra la imagen (Figure 4.14), el software permite hacer el vínculo y la selección de propiedades y tipos de literales de forma manual.

Al establecer las relaciones de la fuente de datos con las clases antes mencionadas, los valores de la propiedad y el tipo de datos – extraídos desde la ontología – quedan reflejados en la siguiente tabla:

El cuadro anteriormente expuesto, no solo nos indica con qué clase se vincula cada elemento de la fuente de datos, sino también refleja la propiedad de esta clase – plasmada en la ontología – y el tipo de dato que será el literal.

Una vez finalizado el proceso de establecer relaciones entre la fuente de datos y la ontología, es preciso comenzar a estructurar el grafo que dará sentido a las relaciones semánticas dentro del modelo de datos.

Fuente de datos	Clase	Propiedad	Tipo de Literal
Sol_nro	Patent	Patentp:docNo	xsd:integer
Fec_present	Patent	Dcterms:date	xsd:date
Fec_pub	Patent	Dcterms:date	xsd:date
Titulo	Patent	Patentp:inventionTitle	xsd:string
nombre_inventor	Inventor	Foaf:firstName	xsd:string
apellido_inventor	Inventor	Foaf:lastName	xsd:string
pais_inv	Country	Patentp:countryCode	xsd:string
Propietario	Assignee	Patentp:Orgname	xsd:string
Pais_prop	Country	Patentp:countryCode	xsd:string
representante	Agent	Patentp:Orgname	xsd:string
Pais_rep	Country	Patentp:countryCode	xsd:string
Clasificacion	InternationalClassification	Patentp:MainClassCode	xsd:string
Resumen	Patent	Dcterms:abstract	xsd:string

TABLA 4.1. Tabla de clases, propiedades y tipos de literales

#### 4.3.4. Construcción del grafo

El siguiente proceso se le denomina construcción del grafo y busca definir todos los espacios del mapeo entre la fuente y la ontología. Se le denomina grafo porque es una construcción de una red donde, a un nivel macro, los nodos representan a las clases en la ontología y los vínculos representan las propiedades que relacionan dichas clases (Knoblock et al., 2012).

Según el autor, en Karma, el proceso de construcción del grafo se divide en tres pasos secuenciales:

- **Iniciación del grafo:** En este paso, por cada tipo semántico asignado a una columna de la fuente de datos, se genera un nodo con una etiqueta única en el grafo. Por ejemplo, al vincular la columna propietario con la clase “assignee”, se genera de manera automática un nodo que corresponde a esa clase y será parte del grafo. Así con todas las clases a las que se le asigna un tipo semántico.
- **Cierre de los nodos:** Una vez creados los nodos y definidos los tipos semánticos, se establecen enlaces (propiedades) entre la fuente de datos y las clases. El procedimiento para efectuar este cierre de nodos es que, por cada clase que tiene

un camino hacia un tipo semántico se debe agregar un nodo en el grafo para que ningún tipo semántico quede aislado.

- **Agregado o adición de enlaces:** El paso final para la construcción del grafo es agregar los enlaces que expresan las distintas relaciones entre los nodos. Se pueden conectar dos nodos en el grafo si hay algún tipo de dato, propiedad de objeto o relación que conecte expresamente esos nodos en la ontología. En la interfaz de Karma, un nodo puede recibir un enlace desde otro nodo o emanar un enlace hacia otro nodo.

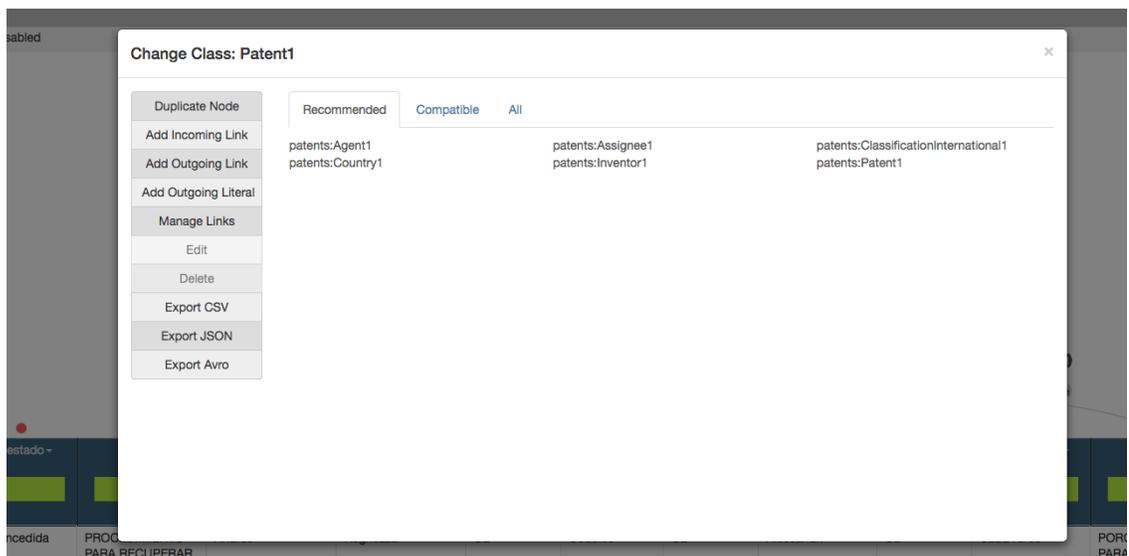


FIGURA 4.15. Interfaz de Karma para elaborar el grafo.

En el modelo de base de datos de patentes expuesto, las relaciones se tornan fundamentalmente centralizadas por la clase principal de la ontología que es “Patent” a través de un “incoming links”, es decir, enlaces que relacionan otras clases con la principal. La imagen (Figure 4.15), muestra la manera en que Karma permite incorporar enlaces entre las distintas clases de la ontología de AKSW. En resumen, se puede duplicar el nodo, agregar un link entrante, un link saliente, agregar un literal saliente o administrar los enlaces.

Una vez determinados los nodos y los distintos enlaces que tengan las clases (object property, data property o relaciones), Karma permite visualizar de manera gráfica la red de

clases con sus respectivos vínculos. La siguiente imagen (Figure 4.16) muestra la construcción de un grafo en Karma utilizando como base la ontología de AKSW, considerando la clase Patent como nodo principal y estableciendo vínculos con otras clases como inventor, solicitante, representante o clasificación internacional.

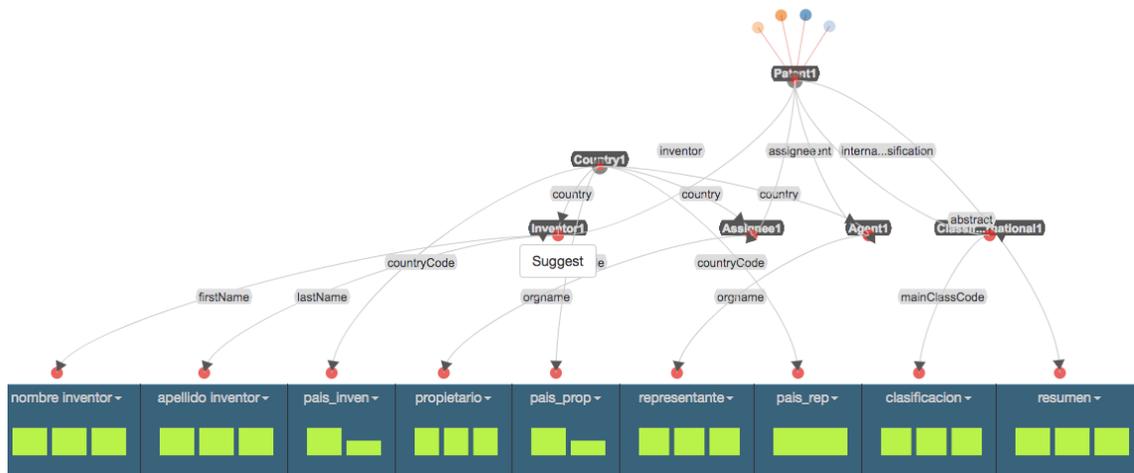


FIGURA 4.16. Construcción del grafo en Karma.

#### 4.3.5. Refinamiento del modelo

El paso final antes de generar el modelo R2RML formal y publicar contenidos en RDF u otro formato (JSON, TTL) es refinar el modelo y Karma permite trabajar esta materia a distintos niveles.

En primer lugar, el software permite trabajar al nivel del grafo construido en base a la ontología, permitiendo efectuar modificaciones en las clases (nodos) y en las relaciones del grafo, es decir, es posible reemplazar un nodo o una relación junto con integrar o quitar nodos. Por otra parte, las relaciones pueden ser cambiadas por propiedades de datos o de objetos distintas y, adicionalmente, se puede modificar el tipo de literal.

Por otro lado, Karma permite trabajar a nivel de la fuente de datos, en este caso una base de datos relacional sobre patentes, ofreciendo capacidades para desarrollar limpieza, normalización y transformación de datos antes de modelarlos. Al visualizar las distintas

columnas de la fuente de datos en la interfaz, se pueden ajustar columnas, datos, filas o celdas. Por ejemplo, si el modelo define una columna “Clasificación” y las patentes tienen dos clasificaciones internacionales y están separadas por un valor delimitador (coma, guión, espacio, etc), Karma permite dividir las celdas y crear una nueva columna para una segunda clasificación.

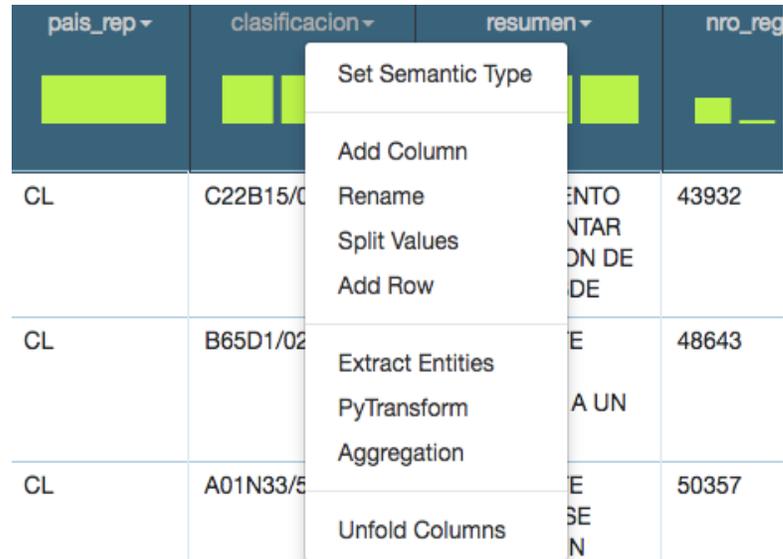


FIGURA 4.17. Herramientas de refinamiento de Karma.

Una vez ajustados tanto los elementos de la fuente de datos original (base de datos relacional) así como los elementos estructurados en el grafo en base a la ontología, es posible generar el modelo R2RML para hacer la traducción.

#### 4.3.6. Publicación del modelo

Finalmente, es posible publicar, exportar y almacenar el modelo de mapeo basado en R2RML, el estándar recomendado por la W3C para determinar las equivalencias entre elementos de bases de datos relacionales con un modelo RDF, para expresarlas finalmente en este formato, lo que permite hacer consultas en SPARQL o diseñar una interfaz de datos enlazados.

En primera instancia, el modelo muestra declaraciones sobre los elementos que serán explícitos en el mapeo de la base de datos relacional, como por ejemplo, las ontologías

activas en el proceso, los esquemas y lenguajes involucrados. Posteriormente, el modelo exhibe el input de los datos y finalmente el modelado de los tipos semánticos, de acuerdo a lo expuesto en Karma.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix purl: <http://purl.org/dc/terms/> .
@prefix patents: <http://us.patents.aksw.org/ontology/> .
@prefix patentp: <http://us.patents.aksw.org/property/> .
@prefix gndo: <http://d-nb.info/standards/elementset/gnd#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dbpedia-owl: <http://dbpedia.org/ontology/> .
@prefix patent: <http://us.patents.aksw.org/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix km-dev: <http://isi.edu/integration/karma/dev#> .

_:modelbhe7jn37x1 a km-dev:R2RMLMapping ;
  km-dev:sourceName "patentes" ;
  km-dev:modelPublicationTime "1496200109160"^^xsd:long ;
  km-dev:modelVersion "1.7" ;
  km-dev:hasInputColumns "[{\columnName\": \"titulo\"}],
  [{\columnName\": \"clasificacion\"}], [{\columnName\": \"fec_pub\"}],
  [{\columnName\": \"resumen\"}], [{\columnName\": \"pais_inven\"}],
  [{\columnName\": \"propietario\"}], [{\columnName\": \"nombre_inventor\"}],
  [{\columnName\": \"apellido_inventor\"}], [{\columnName\": \"representante\"}],
  [{\columnName\": \"fec_present\"}], [{\columnName\": \"pais_rep\"}],
  [{\columnName\": \"sol_nro\"}], [{\columnName\": \"pais_prop\"}]]" ;
  km-dev:hasOutputColumns "[{\columnName\": \"titulo\"}],
  [{\columnName\": \"clasificacion\"}], [{\columnName\": \"fec_pub\"}],
  [{\columnName\": \"resumen\"}], [{\columnName\": \"pais_inven\"}],
  [{\columnName\": \"propietario\"}], [{\columnName\": \"nombre_inventor\"}],
  [{\columnName\": \"apellido_inventor\"}], [{\columnName\": \"representante\"}],
  [{\columnName\": \"fec_present\"}], [{\columnName\": \"pais_rep\"}],
  [{\columnName\": \"sol_nro\"}], [{\columnName\": \"pais_prop\"}]]" ;
  km-dev:hasModelLabel "patentes" ;
  km-dev:hasBaseURI "http://datos.inapi.cl/records/" ;
  km-dev:hasWorksheetHistory ""[

```

FIGURA 4.18. Publicación de Modelo R2RML.

En base a todo el proceso previamente descrito, la imagen (Figure 4.18) muestra el resultado de la exportación del modelo R2RML desde Karma, generado en base al modelo de datos de INAPI y, fundamentalmente, la ontología de AKSW, aunque se hace uso y se declara participación de otras ontologías o esquemas de datos (FOAF, OWL, XMLSchema, DCTerms, entre otros). A continuación, el esquema muestra el input del modelo de la base de datos de INAPI, incorporando las distintas columnas que fueron seleccionadas para el modelamiento. Finalmente, el modelo describe las relaciones de equivalencia establecidas entre la base de datos relacional y el modelo en RDF, de acuerdo a las configuraciones determinadas en la interfaz de usuario.

Cabe señalar que es posible generar más de un modelo R2RML y almacenarlos en Karma para establecer comparaciones de eficiencia en el mapping de la base de datos relacional a un modelo en RDF, con el objetivo de determinar qué modelo traduce de mejor manera las tablas a tripletas en RDF.

Una vez concretada la publicación del modelo R2RML, se pueden definir y publicar las tripletas en RDF, lo que permitirá explotar la utilización de datos vinculados mediante distintas aplicaciones potencialmente incorporables a los servicios de información de INAPI.

#### **4.4. Definición de tripletas**

Como se ha expuesto previamente, el desarrollo de RDF Schema permite definir y describir tanto clases como propiedades e individuos. Dentro de las grandes ventajas de RDF es que permite establecer inferencias básicas en base a la generación de triples o tripletas.

En concreto, una inferencia consiste en tomar una tripleta de un grafo y relacionarla con otra tripleta. Por ejemplo, si la entidad “A” se relaciona con la entidad “B” y la entidad “B” se relaciona con la entidad “C”, se puede inferir que la entidad “A” se relaciona con la entidad “C”. Dicho elemento está altamente documentado en la literatura del análisis de redes sociales y las matemáticas (Wasserman y Faust, 2013). A este principio se le conoce como transitividad y sustenta las inferencias generadas vía RDF.

Por otra parte, RDF Schema permite declarar el dominio y el rango de una propiedad, así como definir que una propiedad es subpropiedad de otra (Gayo, 2012).

Por otra parte, junto con publicar el modelo R2RML, Karma permite exportar los datos en RDF (además de otros formatos como JSON) para generar las tripletas de manera automatizada. Lo anterior es el paso final al proceso de traducción de una base de datos relacional a un modelo de tripletas en RDF y queda graficado con el ejemplo de un registro en la siguiente imagen (Figure 4.19).

N Solicitud: 199902683	Fecha Solicitud: 18/11/1999
Fecha de Publicación: 09/01/2001	Estado: Concedida
N Registro Patentes: 43932	Fecha Registro: 22/10/2008
Tipo de Solicitud: Patente de Invención	
Prioridad: US 60/109,128 19/11/1998	
US 60/109,128 19/11/1998	
Clasificaciones IPC: C22B15/00	
(5): C22B3/08	
C22B15/00	
C22B3/08	

**Título o Materia de la Solicitud**

PROCEDIMIENTO PARA RECUPERAR COBRE DESDE MINERAL CHANCADO, EMPLEANDO UNA SOLUCION LIXIVIADORA DE ACIDO SULFURICO A LA QUE SE AGREGA UNA FORMULACION QUE EXHIBE PROPIEDADES ANTI-ESPUMANTES EN SOLUCION DE ACIDO SULFURICO DE PH 1-2.

**Resumen**

PROCEDIMIENTO PARA AUMENTAR LA LIXIVIACION DE COBRE, DESDE MINERALES QUE LO CONTIENEN, CON UNA SOLUCION LIXIVIADORA DE ACIDO SULFURICO, A LA CUAL SE AGREGA UNA FORMULACION ANTIESPUMANTE QUE ES ESTABLE A UN PH DE 1-2, SIENDO ESTA FORMULACION ANTI-ESPUMANTE PREFERIDA LA QUE COMPRENDE UN ESTER DE GLICOL Y UN ETOXIADO DE ALQUILOL FENOL EN UN SOLVENTE DE ACEITE PARAFINICO QUE SE AGREGA A LA SOLUCION LIXIVIADORA DE ACIDO SULFURICO.

```

<http://datos.inapi.cl/records/199902683> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://us.patents.aksw.org/ontology/Patent> .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/inventionTitle> "Procedimiento para recuperar cobre desde mineral
chancado, empleando una solución lixiviada de ácido sulfúrico a la que se agrega una formulación que exhibe propiedades anti-espumantes en solución
de ácido sulfúrico de PH 1-2"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://us.patents.aksw.org/country-CL> <http://us.patents.aksw.org/property/countryCode> "CL"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/docNo> "199902683"^^<http://www.w3.org/2001/XMLSchema#int> .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/year> "2001"^^<http://www.w3.org/2001/XMLSchema#Year> .
<http://datos.inapi.cl/records/199902683> <http://purl.org/dc/terms/abstract> "Procedimiento para aumentar la lixiviación de cobre, desde minerales que
lo contienen, con una solución lixiviada de ácido sulfúrico, a la cual se le agrega una formulación antiespumante que es estable a un pH de 1-2, siendo
esta formulación anti-espumante preferida la que comprende un ester de glicol y un etoxilado de alquilo fenol en un solvente de aceite parafinado que se
agrega a la solución lixiviadora de ácido sulfúrico" .@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/assignee> <http://datos.inapi.cl/proprietario/432> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Assignee> .
<http://datos.inapi.cl/proprietario/432> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://datos.inapi.cl/proprietario/432> <http://us.patents.aksw.org/property/orgname> <http://datos.inapi.cl/proprietario/432> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Organization> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/2000/01/rdf-schema#label> "Codelco"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/inventor> <http://datos.inapi.cl/inventor/2344> .
<http://datos.inapi.cl/inventor/2344> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Inventor> .
<http://datos.inapi.cl/inventor/2344> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://datos.inapi.cl/inventor/2344> <http://xmins.com/foaf/0.1/FirstName> "Andrés" @es .
<http://datos.inapi.cl/inventor/2344> <http://xmins.com/foaf/0.1/LastName> "Reghezza" @es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/MainClassCode> <http://us.patents.aksw.org/ClassificationInternational>
<http://us.patents.aksw.org/ClassificationInternational> <http://www.w3.org/2000/01/rdf-schema#label> "C22B15/00" @es .

```

FIGURA 4.19. Registro de base de datos de INAPI y su traducción a un registro en RDF.

La traducción de un registro de una base de datos relacional a RDF permite establecer relaciones semánticas entre los datos. Para fundamentar lo anterior es preciso analizar un ejemplo de las tripletas creadas en Karma. La imagen (Figure 4.20) muestra un registro que exhibe un conjunto de datos estructurados en tripletas:

```

<http://datos.inapi.cl/records/199902683> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://us.patents.aksw.org/ontology/Patent> .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/inventionTitle> "Procedimiento para recuperar cobre desde mineral
chancado, empleando una solución lixiviada de ácido sulfúrico a la que se agrega una formulación que exhibe propiedades anti-espumantes en solución
de ácido sulfúrico de PH 1-2"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://us.patents.aksw.org/country-CL> <http://us.patents.aksw.org/property/countryCode> "CL"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/docNo> "199902683"^^<http://www.w3.org/2001/XMLSchema#int> .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/year> "2001"^^<http://www.w3.org/2001/XMLSchema#Year> .
<http://datos.inapi.cl/records/199902683> <http://purl.org/dc/terms/abstract> "Procedimiento para aumentar la lixiviación de cobre, desde minerales que
lo contienen, con una solución lixiviada de ácido sulfúrico, a la cual se le agrega una formulación antiespumante que es estable a un pH de 1-2, siendo
esta formulación anti-espumante preferida la que comprende un ester de glicol y un etoxilado de alquilo fenol en un solvente de aceite parafinado que se
agrega a la solución lixiviadora de ácido sulfúrico" .@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/assignee> <http://datos.inapi.cl/proprietario/432> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Assignee> .
<http://datos.inapi.cl/proprietario/432> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://datos.inapi.cl/proprietario/432> <http://us.patents.aksw.org/property/orgname> <http://datos.inapi.cl/proprietario/432> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Organization> .
<http://datos.inapi.cl/proprietario/432> <http://www.w3.org/2000/01/rdf-schema#label> "Codelco"@es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/inventor> <http://datos.inapi.cl/inventor/2344> .
<http://datos.inapi.cl/inventor/2344> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmins.com/foaf/0.1/Inventor> .
<http://datos.inapi.cl/inventor/2344> <http://us.patents.aksw.org/property/country> <http://us.patents.aksw.org/country-CL> .
<http://us.patents.aksw.org/country-CL> <http://www.w3.org/2000/01/rdf-schema#label> "Chile"@es .
<http://datos.inapi.cl/inventor/2344> <http://xmins.com/foaf/0.1/FirstName> "Andrés" @es .
<http://datos.inapi.cl/inventor/2344> <http://xmins.com/foaf/0.1/LastName> "Reghezza" @es .
<http://datos.inapi.cl/records/199902683> <http://us.patents.aksw.org/property/MainClassCode> <http://us.patents.aksw.org/ClassificationInternational>
<http://us.patents.aksw.org/ClassificationInternational> <http://www.w3.org/2000/01/rdf-schema#label> "C22B15/00" @es .

```

FIGURA 4.20. Registro en RDF de una patente.

Como se ha definido previamente, el lenguaje RDF busca describir los recursos digitales identificados e individualizados en URIs, asignándole propiedades y valores o vinculándolos con otra entidad. RDF define las tripletas, que también pueden ser definidas como un conjunto de declaraciones simples que se compone por tres elementos: Sujeto, predicado y objeto.

El siguiente ejemplo (Figure 4.21) demuestra que, en el caso de la publicación de tripletas en RDF de la base de datos de INAPI, es posible identificar los elementos fundamentales en un esquema de esta naturaleza como los antes mencionados:

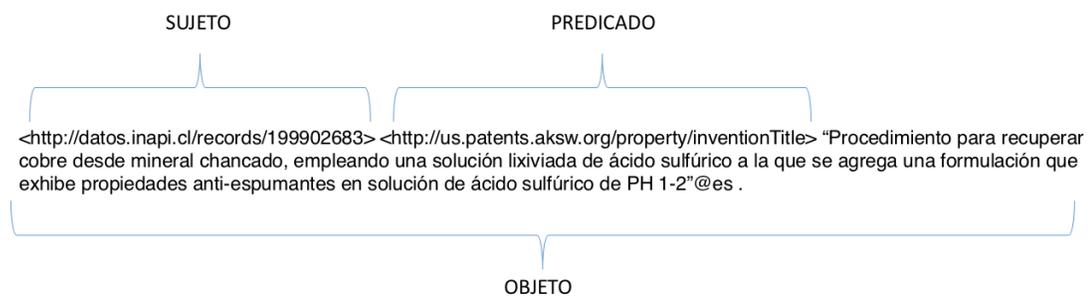


FIGURA 4.21. Tripletas de una patente y su título.

En el ejemplo expuesto en la figura (Figure 4.21), es posible asociar al sujeto con la URI que identifica el documento de patente (`<http://datos.inapi.cl/records/199902683>`), la relación o el predicado definido con una propiedad (`InventionTitle`) y el valor u objeto con un literal que en este caso es el texto del título del documento de patente. En términos concretos, la tripleta nos indica que la patente “199902683” lleva por título de la invención el valor “Procedimiento para recuperar cobre desde mineral chancado...” y, adicionalmente, la signatura “@es” nos indica que el literal está escrito en español.

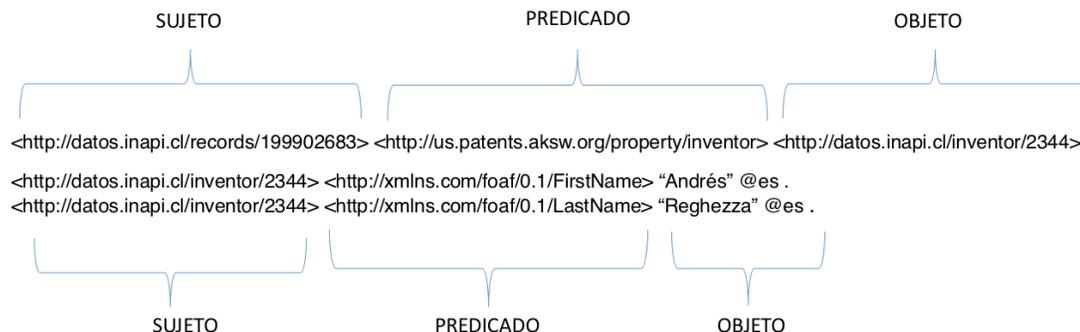


FIGURA 4.22. Tripletta de una patente y su inventor.

Si se toman otros ejemplos, como en los expuestos en las dos figuras siguientes (Figure 4.22, Figure 4.23), podemos identificar casos donde se relaciona una URI con otra URI para, posteriormente en otra tripletta, identificar los literales o los valores del dato. En el primer caso se relaciona la URI del documento de patente (datos.inapi.cl/records/199902683) a través de una propiedad (Inventor) con la segunda URI que identifica al inventor (datos.inapi.cl/inventor/2344). A continuación se especifica la misma URI del inventor que define a través de las propiedades de FOAF (“FirstName” y “LastName”) el nombre y el apellido del inventor (Andrés Reghezza). La misma lógica se aplica para el segundo caso, sólo que la URI de la patente se relaciona con la URI y la propiedad correspondiente al solicitante.

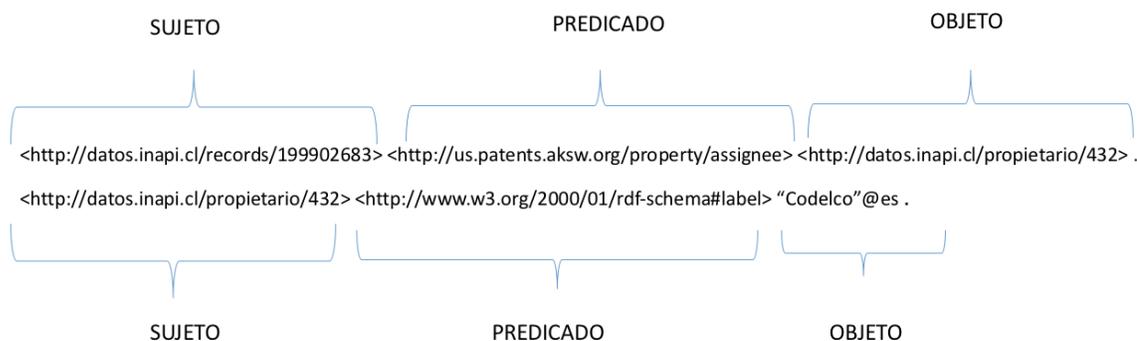


FIGURA 4.23. Tripletta de una patente y su solicitante.

Como fue expuesto previamente, si expresamos estas tripletas de RDF en un grafo, el sujeto y el objeto toman el rol de ser nodos (URIs y literales, dependiendo del caso correspondiente), mientras que el predicado (propiedades) define las relaciones entre los nodos. La siguiente figura ejemplifica un grafo basado en las relaciones en RDF expuestas previamente (Figure 4.24).

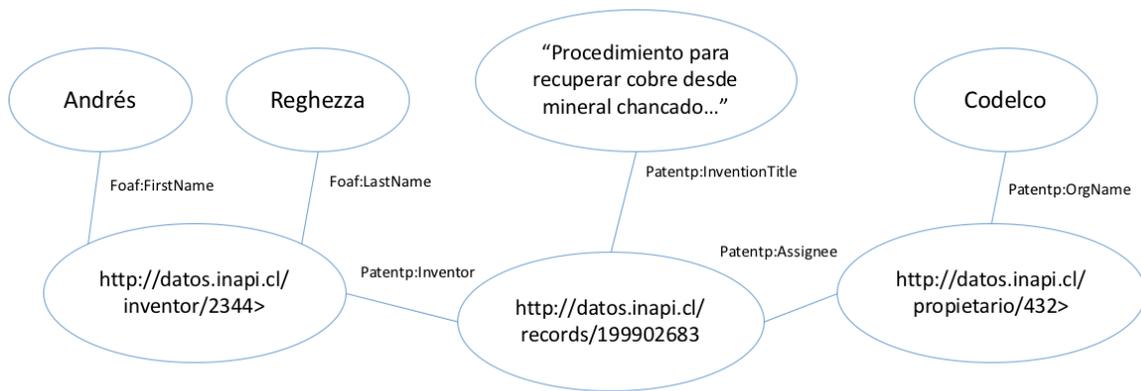


FIGURA 4.24. Grafo de relaciones Inventor – Patente - Solicitante.

De esta manera, en el ámbito de la propiedad industrial, se concreta la traducción de un modelo clásico de base de datos relacional (Entidad – atributo – valor) a un modelo en RDF semántico (Sujeto – predicado – objeto) que responde a los estándares de datos enlazados para proveer de un servicio semántico en el ámbito de las patentes.

#### 4.5. APLICACIONES

Antes de enlistar las aplicaciones que pueden tener las tripletas en RDF en el contexto de la propiedad industrial, es fundamental reconocer que las tecnologías de datos enlazados tienen mucho potencial y pueden ser fuertemente aplicadas a distintos contextos de la I+D+i, incluyendo a las patentes de invención (García Moreno, 2015). En el caso de la base de datos de INAPI, se identifican tres aplicaciones fundamentales: Optimización de búsqueda de patentes, vinculación de datos con otras fuentes y aplicación de estándares LOD.

#### **4.5.1. Optimización de búsqueda de patentes**

Debido a la complejidad de los documentos, al aumento exponencial en el número de patentes en los últimos años y el uso de un vocabulario deliberadamente complejizado, los documentos del área de la propiedad industrial son recursos de información que no han sido explotados en todo su potencial y su búsqueda resulta tremendamente difícil (Eisinger et al., 2014). Lo anterior expone a las tecnologías de datos enlazados y a la web semántica como una oportunidad de fortalecimiento de los servicios de búsqueda de información.

Desde esta perspectiva, la integración de tecnologías de datos vinculados al buscador de patentes de INAPI presenta un potencial de búsqueda más allá de la coincidencia de sintaxis que ofrece actualmente su buscador. En primer lugar, integrar, aunque sea a un nivel básico, las tecnologías de datos enlazados a cualquier servicio web trae beneficios tanto a presente como futuro debido al posicionamiento semántico que ofrecen los microdatos. Al insertar estos vocabularios estandarizados será mucho más fácil ser encontrado en buscadores externos (Google, Yahoo!, Bing) y será factible incorporar nuevas funciones de búsqueda semántica o descubrimiento de recursos de manera facetada utilizando las inferencias de RDF. En segunda instancia, la incorporación de las descripciones en RDF permiten la integración de diversas fuentes de datos de manera flexible, lo que posibilita hacer búsquedas de información más complejas, debido a que puedes tanto integrar como ser integrado a otras fuentes. Finalmente, el uso de elementos técnicos de datos enlazados contribuye a identificar de mejor forma los recursos, en este caso las patentes, mediante URIs, a localizarlos mediante protocolo HTTP y describirlos mediante las propiedades.

De acuerdo a (Gayo, 2012), “la Web Semántica se hace efectiva en el momento en que el usuario final obtiene un valor añadido” (p. 94). Desde esta perspectiva, los usuarios de la base de datos de INAPI pueden beneficiarse añadiendo aplicaciones como un buscador de inventores o solicitantes usando técnicas de integración basadas en RDF. Asimismo, INAPI podría incluir un endpoint para hacer consultas en SPARQL, crear nuevas herramientas de visualización y fomentar la reutilización de datos desde otras fuentes.

#### 4.5.2. Vinculación de datos con otras fuentes.

El crecimiento e interés a nivel global por el movimiento de datos enlazados (linked data) ha sido exponencial. De acuerdo a las cifras, “ mientras que en el año 2008 se contabilizaron 2 millones de tripletas, en el año 2010 se alcanzaron 25 mil millones de tripletas” (Gayo, 2012).

En el contexto de la propiedad industrial, las patentes son expuestas en grandes bases de datos relacionales donde se disponibilizan a texto completo en PDF. Adicionalmente, si retomamos la idea de Bermúdez-Edo et al. (2013), la información o sus metadatos son presentados en formatos estándar como CVS (Comma-Separated Value) o XML (eXtensible Markup Language). Lo anterior supone un precario potencial semántico de los datos que no permite establecer relaciones para realizar análisis u ofrecer servicios con valor añadido a sus usuarios.

En el caso de la base de datos de INAPI, integrar tecnologías semánticas traería una serie de beneficios (Gayo, 2012)

- Uso de URI para identificar entidades: Si bien en el contexto de la propiedad industrial, las nomenclaturas son complejas, por ejemplo, si consideramos las familias de patentes o lo variables que son los registros del nombre de un inventor o de un solicitante, las URI pudiesen ser una solución válida para homogeneizar estos registros. En el caso de los documentos de patentes de INAPI, resulta tremendamente recurrente el ingreso de solicitantes o inventores de distinta manera, por ejemplo, Codelco o Corporación Nacional del Cobre. Las URIs podrían contribuir a solventar ese problema de duplicidades.
- Localización de recursos en la web: En base a lo anterior, si los recursos están debidamente identificados, es posible localizarlos mediante el protocolo HTTP. Lo anterior fortalece la encontrabilidad de los recursos, sobre todo si los recursos están descritos y enriquecidos con microdatos. De esta manera, el buscador de patentes de INAPI se vería tremendamente beneficiado al momento de recuperar los documentos y exhibirlos al usuario.

- Enlaces con URIs de otras fuentes: Mediante una correcta descripción de las URIs en RDF es posible explorar otras fuentes y otros recursos para navegar en los datos, siguiendo los distintos vínculos e inferencias que ofrecen las tripletas y los grafos. De esta forma, es posible recabar datos de otras fuentes que utilicen vocabularios similares. Por ejemplo, si en la base de datos de INAPI tenemos una URI que define la entidad “Pontificia Universidad Católica” como solicitante en nuestro modelo de datos basado en la ontología de AKSW, y la relacionamos con la URI que define a “Pontificia Universidad Católica” en DBPedia, podemos obtener datos que enriquecen la descripción del recurso. La figura 40 ejemplifica cómo esta vinculación con DBPedia podría significar la obtención de datos relevantes sobre esa institución como el país, la ciudad, fecha de fundación, cantidad de estudiantes o tipo de institución, entre muchos otros.

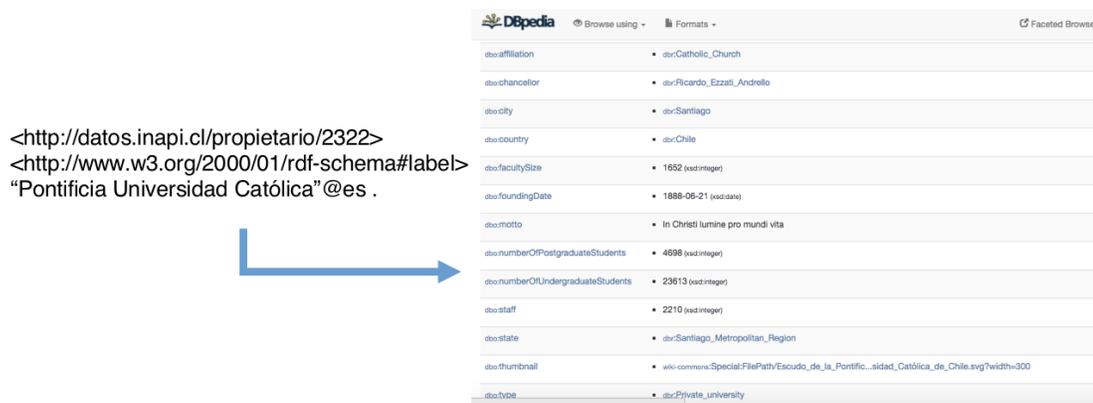


FIGURA 4.25. Datos potencialmente obtenidos de DBPedia mediante datos vinculados.

### 4.5.3. Aplicación de estándares de datos abiertos y enlazados (LOD).

La publicación de datos en un sitio web es una de alta relevancia en el contexto actual. Normalmente los desarrolladores web se enfocan en elementos como el diseño gráfico, las funcionalidades o animaciones, pero no se enfocan en lo relevante que es publicar datos abiertos y estandarizados que le den mayor valor. Bajo la premisa anterior, (Lee y Miller,

2002) propuso un modelo de publicación de datos basado en 5 estrellas, definidas de la siguiente manera:

- **1 estrella:** Implica publicar los datos en cualquier formato y bajo una licencia abierta, aunque estos no estén estructurados. Por ejemplo, publicar una tabla con datos en un PDF, donde es posible verlos pero no extraerlos de manera sencilla, ni trabajarlos fácilmente de manera automatizada.
- **2 estrellas:** Implica publicar los datos en cualquier formato y bajo una licencia abierta, aunque estos no estén estructurados. Por ejemplo, publicar una tabla con datos en un PDF, donde es posible verlos pero no extraerlos de manera sencilla, ni trabajarlos fácilmente de manera automatizada.
- **3 estrellas:** Implica publicar los datos en un formato estructurado y libre, sin limitación para ser usado por un software propietario. De esta manera, los datos pueden ser manipulados con mayor libertad, aunque no están integrados en la web.
- **4 estrellas:** Implica publicar los datos utilizando URIs para identificar los datos más importantes y los representa con RDF y/o otros formatos. De esta manera, los datos ya están integrados en la web y se pueden relacionar las URIs con otras fuentes.
- **5 estrellas:** Los datos son enlazados por otras fuentes y eso enriquece las fuentes del editor y el consumidor porque provee de contexto.

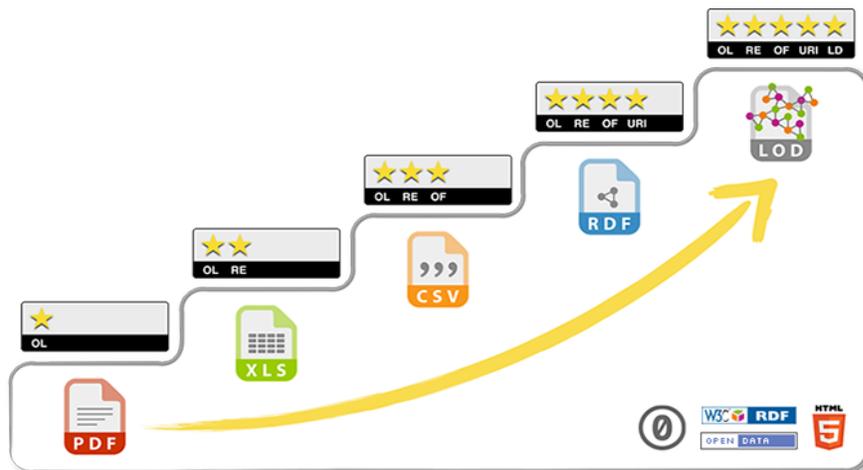


FIGURA 4.26. Modelo de 5 estrellas de datos abiertos y enlazados.

En el caso general de las bases de datos de patentes a nivel global ya se ha afirmado que la tendencia es llegar hasta publicar sus datos en un formato CSV y XML, por lo que es posible sostener que estas fuentes de información llegan hasta la tercera estrella, lo que no provee de soluciones semánticas integradas a la web. Aún más, en el caso del buscador de patentes de INAPI, los datos no se exhiben en un formato estructurado y en algunos registros solo se puede descargar un archivo zip con imágenes en PNG sobre la patente. Por lo tanto, la base de datos de INAPI incluso no alcanzaría la primera estrella en muchos de sus registros.

En términos prácticos, si la base de datos de INAPI incorpora tecnologías semánticas a su servicio, ayuda a publicar correctamente sus contenidos, a facilitar el procesamiento libre de los datos mediante máquinas, a describir adecuadamente los contenidos y enlazarlos con otros datos haciéndolos interoperables. Es relevante señalar que el modelo de estrellas puede resultar un estándar de medición para instituciones públicas aportando a la transparencia y accesibilidad a la información.

## 5. CONCLUSIONES

A raíz de la investigación bibliográfica y los resultados obtenidos en la traducción de la base de datos de INAPI a un modelo RDF basado en una ontología, se desprenden las siguientes conclusiones:

Primeramente, en el contexto actual, un sistema de I+D+i requiere más y mejores herramientas en muchos aspectos: gestión de innovación, de ideas, de información, entre otras. Particularmente, en el caso de la propiedad industrial en Chile, los servicios de información de INAPI tienen muchas oportunidades de fortalecer y mejorar la recuperación de información o la visibilidad de la documentación tecnológica protegida o de dominio público a nivel nacional.

Una de las grandes oportunidades que se presentan es la inclusión de tecnologías semánticas o de linked data, que van tomando cada vez más fuerza y relevancia. Si bien, la mayor parte de los contenidos digitales están publicados en base de datos relacionales, el interés por traducir estas fuentes de información a modelos semánticos ha sido materia creciente de investigación.

De esta manera, tecnologías como RDF, ontologías (OWL), URIs, entre otras, ofrecen ingresar al mundo del linked data, donde la interoperabilidad de los datos y la comprensión de las máquinas es su principal fortaleza. Reforzando lo anterior, el interés a nivel global por incorporar estas tecnologías es una tendencia que, aunque no ha sido fuertemente explotada en el contexto de las patentes, presenta un tremendo potencial para optimizar servicios de bases de datos de patentes como es el caso de INAPI.

Por otro lado, el crecimiento en el campo de investigación de la web semántica ha permitido que la proliferación de metodologías y herramientas sea muy grande. De este modo, es posible identificar diversas aplicaciones y procedimientos para traducir fuentes de datos existentes a un modelo de datos semántico en RDF.

Para el contexto de la base de datos de patentes de INAPI, la estrategia procedimental para la integración de tecnologías semánticas mediante la traducción de una base de datos

relacional a RDF se compone de cuatro elementos fundamentales (Figure 5.1) como son el análisis del modelo de datos, definición de ontología, modelamiento de datos y publicación de tripletas. De estas etapas del proceso se desprenden las siguientes conclusiones:

- **Análisis de modelo de datos:** para traducir una base de datos relacional a RDF es requisito fundamental conocer su modelo de datos, debido a que permite identificar las entidades y atributos clave que posteriormente pueden componerse en URIs y propiedades en un modelo en RDF.
- **Definición de ontología:** si bien es posible efectuar una traducción directa de una base de datos relacional a RDF sin el uso de una ontología (D2R), el uso de esta última resulta fundamental para potenciar las relaciones semánticas de las tripletas en RDF, por lo tanto es fundamental para un servicio de linked data sólido identificar una ontología del dominio y conocerla a fondo (Clases, propiedades, axiomas, entre otros).
- **Modelamiento de datos:** el conocimiento previo del modelo de datos y la ontología resulta clave para el modelamiento de datos, debido a que permite conectar las tablas de la base de datos relacional con las clases de la ontología. Para este proceso se sugiere utilizar el lenguaje de mapeo R2RML de la W3C, ya que permite integrar una ontología al procedimiento para fortalecer las relaciones semánticas de las tripletas resultantes.
- **Publicación de tripletas:** el output del proceso son las tripletas en RDF, las cuales deben estar correctamente estructuradas y, por lo tanto, el modelo debe ser lo suficientemente refinado para construir un grafo adecuado. Lo anterior permitirá la (re)utilización de estas tripletas para ofrecer servicios de datos enlazados.

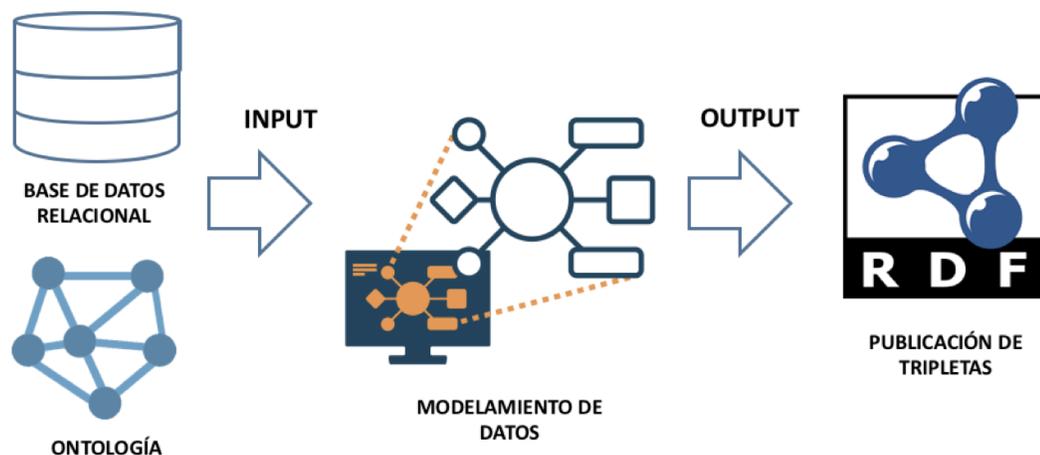


FIGURA 5.1. Modelo de traducción de base de datos a RDF.

En base a la estrategia expuesta, es posible afirmar que el uso de aplicaciones como Protégé para la creación, edición y mantención de ontologías resulta tremendamente útil para comprender la estructura de relaciones que posee uno de estos vocabularios y poder aplicarlo a mecanismos de traducción automatizada a RDF. Por otro lado, el uso de herramientas como Karma para la integración automatizada de datos - mediante la creación de un modelo basado en R2RML - facilita en gran medida la traducción de una fuente de datos (Bases de datos relacionales, CSV, entre otras) a un modelo RDF utilizando una ontología previamente estudiada con Protégé. Lo anterior permite obtener tripletas en RDF que posteriormente pueden tener múltiples aplicaciones para fortalecer la encontrabilidad y visibilidad de los datos que se quieren exhibir. Asimismo, permite enlazar datos con otras fuentes para enriquecer la fuente de datos original y proveerla de contexto.

Para el caso de la base de datos de INAPI, este trabajo propone la construcción de un portal de datos enlazados denominado datos.inapi.cl. Dicho servicio puede proveer los datos en un formato abierto y comprensible para otras fuentes de datos (RDF), junto con optimizar la búsqueda de información con relaciones más complejas basadas en una ontología. Por ejemplo, es posible establecer relaciones complejas entre un inventor, un solicitante, las

tecnologías desarrolladas (Fuente interna) y con datos que pueden ofrecer otros servicios como DBPedia (Fuente externa).

Para lo anterior, resulta fundamental conocer la estructura del modelo de datos de INAPI, identificar las principales entidades y atributos - campos clave - para relacionarlos de manera óptima con la ontología, a fin de que las URIs estén bien asignadas y las clases bien definidas para obtener las tripletas en RDF de forma correcta.

Los datos enlazados suponen una gran oportunidad para las bases de datos de patentes, a pesar de lo compleja que resulta esta documentación. Por lo tanto, esta investigación se propone como un camino entre muchos otros para obtener un mismo objetivo: fortalecer los servicios de búsqueda y recuperación de información en el ámbito de las patentes.

Las tecnologías de la web semántica se transforman en un insumo muy útil, además de necesario para el contexto de la propiedad industrial. De esta manera, se sugiere la aplicación escalable del modelo de publicación de datos de 5 estrellas para la base de datos de INAPI, definido de la siguiente manera:

- **Publicación de patentes en PDF (1 estrella):** En primer lugar, resulta fundamental optimizar la visualización de los documentos de patentes mediante el acceso a documentos PDF que faciliten su lectura.
- **Publicación estructurada de datos en CVS y XML (2 y 3 estrellas):** Para el caso de INAPI, la publicación de datos estructurados permite la explotación de estos mismos por otros actores que requieran manipular de mejor manera la información tecnológica de las patentes.
- **Publicación de datos en RDF y uso de URIs (4 estrellas):** Permite disponer los datos de las patentes identificables con URIs en un lenguaje comprensible para las máquinas, basado en una ontología.
- **Enlazar datos con otras fuentes (5 estrellas):** La publicación de datos previamente expuesta permitiría enlazar los datos con otras fuentes, desarrollando un servicio con mayor contexto semántico que potencie la comprensión de los documentos de patente y su información técnica.

De esta manera, el desarrollo de un servicio de datos enlazados para INAPI se transforma en una oportunidad para que empresas, universidades, centros de investigación y otros actores del ecosistema de innovación puedan utilizar los datos de las patentes como conocimiento para mejorar su productividad, procesos o el desarrollo de nuevas invenciones. Obtener información tecnológica comprensible para los procesos de instituciones de este tipo puede transformarse en una ventaja competitiva que, al final, reditúa en un ambiente de Investigación, Desarrollo e Innovación mucho más sólido.

Finalmente, esta investigación resulta una iniciativa incipiente a nivel nacional en relación a la incorporación de tecnologías semánticas en el ámbito de la propiedad industrial. Por lo tanto, como estrategia de continuidad a esta línea de investigación, se plantea la necesidad de desarrollar nuevas investigaciones en el área de la propiedad industrial y la web semántica desde las siguientes aristas: 1) estudio de modelos de datos en el contexto de la propiedad industrial, 2) creación o adaptación de ontologías en el dominio de las patentes a nivel nacional, 3) estudio de nuevos mecanismos de traducción manual o automática de registros de bases de datos de patentes a tecnologías de linked data y 4) nuevas y mejores aplicaciones basadas en tecnologías de la web semántica para fortalecer el acceso a la documentación científico-técnica.

## BIBLIOGRAFIA

- Adams, S. (2012). *Information sources in patents*. Walter de Gruyter.
- Angrosh, M., Cranefield, S., y Stanger, N. (2014). Contextual information retrieval in research articles: Semantic publishing tools for the research community. *Semantic Web*, 5(4), 261–293.
- Bermudez-Edo, M., Noguera, M., Garrido, J. L., y Hurtado, M. V. (2013). Semantic patent information retrieval and management with owl. En *Advances in information systems and technologies* (pp. 33–42). Springer.
- Berners-Lee, T., Dimitroyannis, D., Mallinckrodt, A. J., y McKay, S. (1994). World wide web. *Computers in Physics*, 8(3), 298–299.
- Bizer, C., Heath, T., y Berners-Lee, T. (2009). Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, 205–227.
- Bizer, C., Heath, T., Idehen, K., y Berners-Lee, T. (2008). Linked data on the web (ldow2008). En *Proceedings of the 17th international conference on world wide web* (pp. 1265–1266).
- de Carvalho, D. S., França, F. M. G., y Lima, P. M. V. (2014). Extracting semantic information from patent claims using phrasal structure annotations. En *Intelligent systems (bracis), 2014 brazilian conference on* (pp. 31–36).
- Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., ... Horrocks, I. (2000). The semantic web: The roles of xml and rdf. *IEEE Internet computing*, 4(5), 63–73.
- Diamantini, C., Potena, D., Proietti, M., Smith, F., Storti, E., y Taglino, F. (2013). A semantic framework for knowledge management in virtual innovation factories.

*International Journal of Information System Modeling and Design (IJISMD)*, 4(4), 70–92.

Eisinger, D., Mönnich, J., y Schroeder, M. (2014). Developing semantic search for the patent domain. En *Ipamin@ konvens*.

Etzkowitz, H., y Leydesdorff, L. (2000). The dynamics of innovation: from national systems and “mode 2” to a triple helix of university–industry–government relations. *Research Policy*, 29(2), 109 - 123. doi: [http://dx.doi.org/10.1016/S0048-7333\(99\)00055-4](http://dx.doi.org/10.1016/S0048-7333(99)00055-4)

García Moreno, C. (2015). *Desarrollo de un modelo para la gestión de la i+d+i soportado por tecnologías de la web semántica* (phdthesis). Universidad de Murcia.

Gayo, J. E. L. (2012). *Web semántica*. NETBIBLO, S.L.

Gerber, A. J., Barnard, A., y van der Merwe, A. J. (2007). Towards a semantic web layered architecture. En *Proceedings of the 25th conference on iasted international multi-conference: Software engineering* (pp. 353–362).

Haslhofer, B., y Schandl, B. (2010). Interweaving oai-pmh data sources with the linked data cloud. *International Journal of Metadata, Semantics and Ontologies*, 5(1), 17–31.

Hazber, M. A., Li, R., Gu, X., y Xu, G. (2016). Integration mapping rules: Transforming relational database to semantic web ontology. *Appl. Math*, 10(3), 1–21.

Hert, M., Reif, G., y Gall, H. C. (2011). A comparison of rdb-to-rdf mapping languages. En *Proceedings of the 7th international conference on semantic systems* (pp. 25–32).

Hunt, D., Nguyen, L., y Rodgers, M. (2012). *Patent searching: Tools & techniques*. John Wiley & Sons.

Knoblock, C., Szekely, P., Ambite, J., Goel, A., Gupta, S., Lerman, K., ... Mallick, P. (2012). Semi-automatically mapping structured sources into the semantic web. *The Semantic Web: Research and Applications*, 375–390.

Lee, T. B., y Miller, E. (2002). The semantic web lifts off. *ERCIM News No, 51*.

Magdy, W., y Jones, G. J. (2014). Studying machine translation technologies for large-data clir tasks: a patent prior-art search case study. *Information retrieval*, 17(5-6), 492–519.

Manglano, B. G.-A., y Zulueta, M. Á. (2008). Estudio comparativo de bases de datos de patentes en internet. En *Anales de documentación* (Vol. 10, pp. 145–162).

Musen, M. A. (2015). The protégé project: A look back and a look forward. *AI matters*, 1(4), 4–12.

OMPI. (2012). *wipo\_pub\_1434\_11.pdf*. [http://www.wipo.int/edocs/pubdocs/es/patents/434/wipo\\_pub\\_1434\\_11.pdf](http://www.wipo.int/edocs/pubdocs/es/patents/434/wipo_pub_1434_11.pdf). ((Accessed on 06/09/2017))

OMPI. (2013). *L434-2e*. [http://www.wipo.int/edocs/pubdocs/es/patents/434/wipo\\_pub\\_1434\\_02.pdf](http://www.wipo.int/edocs/pubdocs/es/patents/434/wipo_pub_1434_02.pdf). ((Accessed on 06/09/2017))

OMPI. (2016). *Principios básicos de la propiedad industrial*. [http://www.wipo.int/edocs/pubdocs/es/wipo\\_pub\\_895\\_2016.pdf](http://www.wipo.int/edocs/pubdocs/es/wipo_pub_895_2016.pdf). ((Accessed on 06/09/2017))

Paranjpe, P. P. (2012). Patent information and search. *DESIDOC Journal of Library & Information Technology*, 32(3).

Salampasis, M., y Hanbury, A. (2014). Perfedpat: An integrated federated system for patent search. *World Patent Information*, 38, 4–11.

Sánchez, J. A. P. (2011). *Tecnologías de la web semántica*. Editorial UOC.

- Sánchez-Jiménez, R., y Gil-Urdiciain, B. (2007). Lenguajes documentales y ontologías. *El profesional de la información*, 16(6), 551–560.
- Schmitz Vaccaro, C. (2009). Propiedad intelectual, dominio público y equilibrio de intereses. *Revista chilena de derecho*, 36(2), 343–367.
- Shadbolt, N., Berners-Lee, T., y Hall, W. (2006). The semantic web revisited. *IEEE intelligent systems*, 21(3), 96–101.
- Singh, V., Chakraborty, K., y Vincent, L. (2016). Patent database: Their importance in prior art documentation and patent search.
- Wasserman, S., y Faust, K. (2013). *Análisis de redes sociales. métodos y aplicaciones* (Vol. 10). CIS-Centro de Investigaciones Sociológicas.
- Zhang, Y., Zhou, X., Porter, A. L., y Gomila, J. M. V. (2014). How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: “problem & solution” pattern based semantic triz tool and case study. *Scientometrics*, 101(2), 1375–1389.

## ANEXO A. POTENCIALES DATASETS ENLAZABLES CON INAPI

Nombre del Dataset	URL	Descripción del servicio
DBPedia	<a href="http://wiki.dbpedia.org/">http://wiki.dbpedia.org/</a>	DBPedia es un proyecto colaborativo de datos abiertos enlazados de Wikipedia que propone una versión semántica. Actualmente posee más de 3,77 millones de entidades descritas en Resource Description Framework (RDF) y se puede consultar mediante SPARQL.
EU Open Data Portal	<a href="https://data.europa.eu/">https://data.europa.eu/</a>	El Portal de datos abiertos y enlazados de la Unión Europea ofrece un total de 10761 datasets disponibles para ser reutilizados. Las temáticas son diversas entre las que se puede destacar la ciencia, los negocios, medio ambiente, productividad y tecnologías, economía y finanzas.
WorldBank	<a href="http://data.worldbank.org/">http://data.worldbank.org/</a>	El Banco de datos (DataBank) es una herramienta basada en datos abiertos y enlazados que permite el análisis y visualización de datos sobre diversos temas vinculados con economía y desarrollo.
LinkedGeoData	<a href="http://linkedgeodata.org/">http://linkedgeodata.org/</a>	LinkedGeoData es un proyecto que ofrece añadir una dimensión espacial a la Web de Datos / Web Semántica. LinkedGeoData utiliza la información recopilada por el proyecto OpenStreetMap y la hace disponible como una base de conocimiento en RDF, de acuerdo con los principios de Linked Data.
Geonames	<a href="http://www.geonames.org/">http://www.geonames.org/</a>	La base de datos geográfica GeoNames cubre todos los países y contiene más de once millones de nombres de lugares que están disponibles para su enlace y descarga gratuita.
Biblioteca del Congreso	<a href="http://datos.bcn.cl/es/">http://datos.bcn.cl/es/</a>	La Biblioteca del Congreso Nacional de Chile ofrece sus datos bajo los estándares de Linked Data para ser enlazados, presentando información legislativa, ordenanzas, normas, entre otros.

TABLA A.1. Datasets potencialmente enlazables con INAPI