



UNIVERSIDAD DE DEUSTO

# ARQUITECTURA SOFTWARE DISTRIBUIDA PARA LA GENERACIÓN Y LA PROVISIÓN DE INFORMACIÓN SEMÁNTICA PARA EL TRANSPORTE MULTIMODAL

Tesis doctoral presentada por Asier Moreno Emborujó  
dentro del Programa de Doctorado en Ing. Informática y Telecomunicación

Dirigida por Dr. Asier Perallos Ruiz y  
Dr. Diego López de Ipiña González de Artaza

Vitoria, julio de 2016





UNIVERSIDAD DE DEUSTO

# ARQUITECTURA SOFTWARE DISTRIBUIDA PARA LA GENERACIÓN Y LA PROVISIÓN DE INFORMACIÓN SEMÁNTICA PARA EL TRANSPORTE MULTIMODAL

Tesis doctoral presentada por Asier Moreno Emborujó  
dentro del Programa de Doctorado en Ing. Informática y Telecomunicación

Dirigida por Dr. Asier Perallos Ruiz y  
Dr. Diego López de Ipiña González de Artaza

El doctorando

El director

El director

Vitoria, julio de 2016





*A mis padres, Domingo y Francisca.*



## **Abstract**

Rapid technological progress in the field of software applied to transport services is enabling the development of new kinds of applications focused, among other things, on improving citizen mobility. This objective is currently considered an institutional priority, supported by the technological investment of the private sector in recent years and the increasing provision of public infrastructures. This makes the Intelligent Transportation Systems (ITS) play a key role in the context of effective management of such infrastructures and services.

The existing literature has been able to establish a correlation between the added value provided by the ITS and the quality of the transport information actually available. Thus, one of the fundamental premises of this thesis is the necessity to establish mechanisms that give support to applications through proper management and publication of transit data. Several opportunities in this regard were found, based on integration, interoperability and openness capabilities, among others.

To meet these challenges two main instruments are developed: a multimodal transport ontology, promoting the harmonization of available transit data, and a distributed and interoperable architecture for the publication of semantic information. Thus, following the strategy pointed out by the EU in the White Paper on Transport, it is to improve and expand the capacity of public institutions, companies and users in order to obtain, share and provide complete transport information.

This thesis proposes a new approach for the management and exchange of multimodal transport information, and does so using a wide range of innovative technologies.

A novel software architecture is presented, with particular emphasis on the design of the data model and the enablement of mobility-related services, providing a semantic model for the representation of transport information.

All these contributions have been validated through a laboratory test suite reproducing actual transport scenarios. Such experimentation has served to demonstrate the hypothesis stated at the beginning of the work. Finally, a working solution of a semantic trip planner running on the proposed architecture is presented, as a demonstration and validation of the system.

## **Resumen**

El rápido progreso tecnológico en el ámbito de los servicios software aplicados al transporte está posibilitando el desarrollo de nuevos tipos de aplicaciones y servicios enfocados, entre otros aspectos, en la mejora de la movilidad ciudadana. Dicho objetivo es considerado actualmente una prioridad institucional, avalada por la inversión tecnológica del sector privado durante los últimos años y la creciente dotación de infraestructuras públicas. Esto hace que los Sistemas Inteligentes de Transporte (ITS) cobren un papel fundamental en el marco de la gestión y la explotación eficaz de dichas infraestructuras y servicios.

La literatura existente ha podido establecer una correlación entre el valor añadido aportado por los servicios ITS y la calidad de la información de transporte disponible. Así, una de las premisas fundamentales sobre las que se constituye esta tesis doctoral es la necesidad de establecer mecanismos que den soporte a las aplicaciones mediante la correcta gestión y publicación de los datos de tránsito. Se encontraron diversas oportunidades en este sentido, atendiendo a las capacidades de integración, interoperabilidad y apertura de la información, entre otras.

Para dar respuesta a dichos retos se desarrollan dos instrumentos fundamentales: una ontología para el transporte multimodal, favoreciendo la armonización de los datos de tránsito disponibles, y una arquitectura distribuida e interoperable para la publicación de información semántica. Así, siguiendo la estrategia apuntada por la UE en el Libro Blanco del transporte, se pretende mejorar y ampliar la capacidad de instituciones públicas, empresas y usuarios de cara a obtener, compartir y proporcionar información de transporte actualizada y completa.

Esta tesis propone un nuevo enfoque para la gestión y el intercambio de información de transporte multimodal, y lo hace mediante la utilización de un amplio abanico de tecnologías innovadoras. Ejemplo de ello es el desarrollo efectuado de una arquitectura software original, con especial énfasis en el diseño del modelo de datos y la habilitación de servicios relacionados con la movilidad, proporcionando un modelo semántico para la representación de la información de transporte.

Todas estas aportaciones han sido validadas a través de un conjunto de pruebas en laboratorio reproduciendo escenarios de actuación reales. Dicha experimentación ha servido para demostrar la hipótesis planteada al inicio del trabajo, permitiendo además contextualizar los beneficios derivados de la aplicación de la arquitectura propuesta, mediante la construcción de una solución de planificación de rutas semántica corriendo sobre la misma.

## Agradecimientos

A Asier Perallos, en primer lugar por haberme dado la oportunidad de trabajar en DeustoTech, por haber apostado por mí y por haber empujado en momentos en los que hacerlo era más difícil. También a nivel profesional por su criterio y su labor como director, estando ahí en todo momento para guiar el resultado de este trabajo hacia buen puerto.

A Diego López de Ipiña, por ser también el *culpable* de que hoy esté investigando. Por su apoyo, su confianza, su consejo y su tiempo. Y también por su manera de enfocar las cosas, profesional y personalmente.

Por supuesto a mis compañeros, a toda la tropa de seniors y juniors que forma la familia Mobility. Un gran equipo en todos los sentidos y uno de los valores más importantes en el día a día. A Nacho, Rober, Pili, Itziar, Pablo, Ander y Bruno, por hacer del trabajo algo tan agradecido y agradable. También al autodenominado *Team Science*: Hugo, Eneko, Pedro, Idoia, Laura y Luis, que ciertamente aporta conocimiento al conjunto, pero sobretudo energía, optimismo y buen ambiente. Trabajar con todos vosotros es una verdadera suerte y un placer.

No me quiero olvidar de los que han pasado por el equipo y ya no están, pero que de una u otra forma me han ayudado en el proceso, gracias a Gorka, Asier, Lucio, Txertu, Unai y Xabi. Tampoco de los que he conocido por el camino, grandes investigadores y mejores personas. Especialmente me quiero acordar de Jesús, Gervasio, Borja y Maria.

Gracias también a mi familia y amigos; en especial a Saioa, por estar ahí durante 10 años ya, que se dice pronto. Y por aguantarme también con la tesis. Gracias *compi*.

Y como no, quiero agradecer a mis padres, Domingo y Francisca, por dárme todo, entre otras cosas una educación y unos valores de los cuales no puedo estar más orgulloso. Gracias por estar ahí, por vuestro apoyo y cariño, y por haber hecho de mí la persona que soy hoy.

*¡Muchas gracias!*

Asier.



# Índice general

<b>1</b>	<b>Introducción y motivación</b>	<b>1</b>
1.1	Definición del problema . . . . .	2
1.2	Trabajo de tesis . . . . .	4
1.2.1	Hipótesis . . . . .	4
1.2.2	Objetivos . . . . .	5
1.2.2.1	Objetivos específicos . . . . .	6
1.3	Metodología de investigación . . . . .	7
1.4	Estructura de la memoria . . . . .	9
<b>2</b>	<b>Estado del arte</b>	<b>13</b>
2.1	Sistemas Inteligentes de Transporte . . . . .	13
2.1.1	Marco social, económico y medioambiental . . . . .	14
2.1.2	Intermodalidad y multimodalidad . . . . .	17
2.1.3	Arquitecturas de servicios multimodales . . . . .	18
2.2	Sistemas de planificación de rutas . . . . .	19
2.2.1	Cronología y características . . . . .	20
2.2.2	Google Transit . . . . .	21
2.2.3	Moveuskadi . . . . .	22
2.2.4	OpenTripPlanner . . . . .	24
2.3	Representación de la información de transporte . . . . .	25
2.3.1	<i>DATA EXchange</i> (DATEX II) . . . . .	26
2.3.2	<i>General Transit Feed Specification</i> (GTFS) . . . . .	28
2.3.3	<i>Web Feature Service</i> (WFS) . . . . .	30
2.3.4	Soluciones Ad-hoc . . . . .	31

## ÍNDICE GENERAL

---

2.4	Interoperabilidad y enriquecimiento de los datos . . . . .	34
2.4.1	Modelos formales y ontologías . . . . .	34
2.4.1.1	Metodologías para el diseño de ontologías . . . . .	36
2.4.2	Web Semántica . . . . .	37
2.4.3	Datos enlazados . . . . .	39
2.4.3.1	Datos enlazados abiertos . . . . .	40
2.5	Hacia la semantización de datos de transporte . . . . .	44
2.5.1	Ontologías con información geoespacial . . . . .	45
2.5.2	Modelización semántica de datos de transporte . . . . .	51
<b>3</b>	<b>Oportunidades de mejora y contribuciones</b>	<b>57</b>
3.1	Modelización de los datos . . . . .	58
3.1.1	Caracterización del modelo de datos . . . . .	59
3.1.2	Diseño del modelo ontológico . . . . .	60
3.2	Interoperabilidad y enriquecimiento . . . . .	61
3.3	Explotación y publicación de la información . . . . .	62
3.3.1	Datos reales de transporte multimodal . . . . .	63
3.3.2	Arquitectura distribuida . . . . .	63
3.4	Análisis y diseño conceptual . . . . .	64
<b>4</b>	<b>Ontología para el transporte multimodal</b>	<b>67</b>
4.1	Metodología de diseño . . . . .	68
4.2	Formalización de la ontología . . . . .	69
4.2.1	Vocabulario . . . . .	72
4.2.2	Gestión de la información geográfica . . . . .	77
4.2.2.1	Gestión de la información geoespacial . . . . .	77
4.2.2.2	Gestión de la información geopolítica . . . . .	79
4.2.3	Adaptador GTFS . . . . .	80
4.3	Enriquecimiento de la información . . . . .	82
4.3.1	Vocabularios enlazados . . . . .	83
4.3.2	Información contextual . . . . .	85

<b>5</b>	<b>Arquitectura del sistema</b>	<b>87</b>
5.1	Capacidades y características . . . . .	88
5.2	Diseño y componentes software . . . . .	91
5.2.1	<i>Linked Data Interface</i> y servidor RDF . . . . .	93
5.2.2	<i>Triple store</i> y servidor SPARQL . . . . .	94
5.3	Integración de datos de transporte distribuidos . . . . .	96
5.3.1	Motivaciones y diseño . . . . .	96
5.3.2	Federación de consultas SPARQL . . . . .	98
<b>6</b>	<b>Experimentación y resultados</b>	<b>103</b>
6.1	Estrategia de validación de la hipótesis . . . . .	104
6.2	Despliegue del sistema y escenario de pruebas . . . . .	105
6.2.1	Presentación del escenario de pruebas . . . . .	106
6.2.2	Carga de datos de transporte multimodal . . . . .	107
6.2.3	Despliegue de servidores semánticos distribuidos . . . . .	109
6.3	STP: Planificador Semántico de Transporte . . . . .	111
6.3.1	Consideraciones iniciales: objetivos y alcance . . . . .	112
6.3.2	Modificación del proyecto OTP . . . . .	114
6.3.2.1	Configuración de las fuentes de datos . . . . .	115
6.3.2.2	Gestión de contexto y vocabularios enlazados . . . . .	117
6.3.3	Consulta y acceso a la información semántica . . . . .	119
6.4	Resultados de la experimentación . . . . .	122
6.4.1	Modelización de los datos y enriquecimiento . . . . .	122
6.4.2	Explotación y publicación de la información . . . . .	123
<b>7</b>	<b>Conclusiones</b>	<b>127</b>
7.1	Visión general, motivación y objetivos . . . . .	128
7.1.1	Contribuciones principales . . . . .	129
7.1.2	Resultados de la experimentación . . . . .	131
7.2	Líneas futuras de trabajo . . . . .	132
7.3	Consideraciones finales . . . . .	134



# Índice de figuras

1.1	Metodología de investigación empleada . . . . .	8
2.1	Inversión pública en infraestructuras en la UE (en % del PIB) . . .	15
2.2	Inversiones realizadas en modos de transporte e infraestructuras .	16
2.3	Cadena de transporte multimodal . . . . .	17
2.4	Arquitectura para la intermodalidad de servicios . . . . .	18
2.5	Interfaz de usuario de Google Transit . . . . .	22
2.6	Interfaz de usuario de Moveuskadi . . . . .	23
2.7	Interfaz de usuario de OpenTripPlanner . . . . .	24
2.8	Lenguajes para la representación del conocimiento . . . . .	35
2.9	Pila completa de la Web Semántica . . . . .	38
2.10	Mapa de datos enlazados abiertos . . . . .	43
2.11	Clases básicas de la ontología OTN . . . . .	47
2.12	Estructura general de la ontología para el dominio del transporte .	48
2.13	Contexto de descubrimiento y composición de servicios . . . . .	49
2.14	Ontología puente entre WFS y las consultas SPARQL . . . . .	50
2.15	Arquitectura de la interfaz ontológica RDF . . . . .	51
2.16	Diagramas UML de la ontología . . . . .	54
2.17	Propiedades definidas para modelar una ruta de transporte . . . . .	55
2.18	Arquitectura del portal semántico planteado por A. Gunay . . . . .	56
3.1	Evolución de la información de transporte . . . . .	61
3.2	Esquema conceptual de la solución propuesta . . . . .	66
4.1	Enfoque incremental por fases para el desarrollo de ontologías . .	69

## ÍNDICE DE FIGURAS

---

4.2	Fichero OWL de MTO cargado en Protégé . . . . .	71
4.3	Entidades y relaciones principales de la ontología MTO . . . . .	73
4.4	Enriquecimiento de la información de transporte . . . . .	82
4.5	Resultados de una consulta SPARQL: relación de restaurantes . . .	84
4.6	Evolución de los datos de transporte: información contextual . . .	86
5.1	Arquitectura para la provisión de información semántica . . . . .	91
5.2	Despliegue de Pubby sobre los datos semánticos de transporte . .	93
5.3	Parliament: arquitectura software y componentes . . . . .	95
5.4	Arquitecturas para la provisión de información de transporte . . .	97
6.1	Arquitecturas para la provisión de información semántica . . . . .	110
6.2	Interfaz web de STP ( <i>Semantic Trip Planner</i> ) . . . . .	113
6.3	Interfaz web para la búsqueda facetada y de POIs . . . . .	118
6.4	Agregación de información externa en relación a la ruta generada .	121
6.5	Clasificación y filtrado de información complementaria . . . . .	122

# Índice de tablas

1.1	Actividades relevantes y objetivo específico al que contribuyen . .	7
2.1	Formatos para la representación de información de transporte . . .	32
2.2	Las cinco estrellas de los datos enlazados . . . . .	41
3.1	Formatos para la representación de información de transporte (b) .	59
3.2	Análisis de limitaciones y contribuciones planteadas . . . . .	65
6.1	Datos de tránsito. Distribución por provincias y modos . . . . .	106
6.2	Datos de tránsito. Distribución de la carga de proceso . . . . .	107
6.3	Proceso de semantización de la información de transporte . . . . .	108
6.4	Comparativa de arquitecturas: modelos centralizado y distribuido .	110
6.5	Comparativa entre OTP y el planificador semántico desarrollado .	119
6.6	Actividades conducentes a la validación de la hipótesis . . . . .	125





# Índice de listados

2.1	Respuesta DATEX. Listado de horarios . . . . .	27
2.2	Fichero CSV de agencias . . . . .	28
2.3	Fichero CSV de rutas . . . . .	29
2.4	Respuesta WFS. Listado de distritos municipales . . . . .	30
4.1	Definición del fichero mto-core.owl . . . . .	70
4.2	Definición del fichero mto-top.owl . . . . .	70
4.3	Descripción lógica de la clase Agency . . . . .	73
4.4	Descripción lógica de la clase Route . . . . .	74
4.5	Descripción lógica de la clase Trip . . . . .	74
4.6	Descripción lógica de la clase Stop . . . . .	75
4.7	Descripción lógica de la clase POI . . . . .	75
4.8	Descripción lógica de la clase Schedule . . . . .	76
4.9	Descripción lógica de la clase Shape . . . . .	76
4.10	Descripción lógica de la clase ShapeLoc . . . . .	76
4.11	Definición de un elemento geolocalizado mediante GeoSPARQL .	77
4.12	Consulta SPARQL. Selección de POIs a 5km de una recta dada . .	78
4.13	Especificación de una parada ( <i>Stop</i> ). Enlace con GeoNames . . . .	79
4.14	Consulta SPARQL. Selección de rutas para una provincia dada . .	80
4.15	Especificación de un POI proveniente de LinkedGeoData . . . . .	83
5.1	Consulta SPARQL. Selección de agencias y exportación a CSV . .	89
5.2	Conjunto de agencias en formato CSV compatible con GTFS . . .	90
5.3	Consulta SPARQL. Ejemplo directiva remota <i>Service</i> . . . . .	99

## ÍNDICE DE LISTADOS

---

5.4	Descripción de la entidad <i>Dataset</i> . URL al <i>endpoint</i> SPARQL . . .	99
5.5	Consulta federada. Listado de paradas . . . . .	100
5.6	Listado de paradas ( <i>Stops</i> ). Se muestran las 10 primeras . . . . .	101
6.1	Fichero de configuración de OTP. Fuentes de datos GTFS . . . . .	115
6.2	Fichero de configuración de STP. Fuente <i>endpoint</i> SPARQL . . .	116
6.3	Consulta SPARQL dinámica en función del contexto de búsqueda	118

*Toda tecnología lo suficientemen-  
te avanzada es indistinguible de la  
magia*

Arthur C. Clarke

# 1

## Introducción y motivación

**D**urante los últimos años se ha venido produciendo un cambio destacable en los hábitos de transporte y movilidad ciudadana. De caracterizarse por la utilización casi exclusiva de medios de transporte privados, se ha pasado a un uso cada vez más responsable (en parte debido a la sensibilización ecológica, en parte debido a la carestía del combustible) de dichos medios, adoptando paulatinamente como alternativa viable el uso del transporte público [UITP 14].

También a nivel institucional el impulso al transporte público de modo que se convierta en opción de transporte preferencial, favoreciendo el desarrollo económico y minimizando el impacto en el medio ambiente, es actualmente un aspecto prioritario [Europea 11]. Se espera que en 2025 el transporte público multimodal haya conseguido duplicar su cuota de mercado con respecto a 2009, según datos de la Asociación Internacional de Transporte Público [UITP 09] completando de este modo la transición hacia un modelo de transporte más sostenible.

Los avances realizados en cuanto a la aplicación de las Tecnologías de la Información y la Comunicación (TIC) al ámbito del transporte han sido constantes y de diversa índole. Tal es así que los Sistemas Inteligentes de Transporte (ITS), cuyo objetivo es optimizar la utilización del transporte, conforman un área de investigación prolífica con múltiples temáticas de trabajo y problemáticas a abordar.

## 1. Introducción y motivación

---

Problemáticas tan diversas como la gestión del tráfico, la logística y la trazabilidad de mercancías, los sistemas vehiculares cooperativos, los sistemas de conducción autónoma o los servicios software enfocados al transporte, entre otras, están despertando un fuerte interés. En relación a este último aspecto y en consonancia con el referido impulso institucional, una de las plataformas que más han evolucionado aportando un valor añadido al usuario son las llamadas soluciones de planificación multimodal de transporte. Dichas soluciones permiten al usuario encontrar la mejor ruta entre varios puntos geográficos en base a diferentes criterios, entre ellos el medio de transporte a utilizar. En esta área, soluciones como Google Transit<sup>1</sup> u OpenTripPlanner<sup>2</sup>, han supuesto un importante avance a la hora de facilitar la gestión y la planificación de los viajes.

En definitiva, se puede observar como el progreso tecnológico dentro del ámbito de los servicios software aplicados al transporte está permitiendo el desarrollo de nuevos tipos de aplicaciones dirigidas a mejorar tanto la gestión de la movilidad ciudadana como, subsecuentemente, la propia calidad de vida en nuestras ciudades.

Se hace necesario destacar que dichas soluciones basan su funcionamiento en la cantidad y la calidad de los datos disponibles. Por consiguiente, cada vez adquiere más importancia establecer mecanismos que den soporte a las aplicaciones mediante la correcta y efectiva gestión y publicación de la información [Davies 02]. En este sentido, paradigmas de reciente aparición, como *Linked Open Data*, que establecen la apertura de la información disponible, almacenada de forma estructurada e integrada, son una gran herramienta para la creación de servicios que se apoyen en dicho conocimiento común [Speiser 11, Norton 10].

### 1.1 Definición del problema

Siendo como se ha comentado la movilidad sostenible una prioridad explícita, y habiendo ya soluciones para favorecer la toma de decisiones por parte del usuario, se considera que dichas soluciones están basadas en modelos de gestión cerrados sobre los cuales, tanto los agentes participantes (ayuntamientos o empresas de transporte público y privado) como el usuario final, no tienen el suficiente control.

---

<sup>1</sup> Google Transit, <http://maps.google.com/transit>

<sup>2</sup> OpenTripPlanner, <http://www.opentripplanner.org>

## 1.1 Definición del problema

---

Por un lado, las soluciones existentes no son lo suficientemente interoperables de cara a integrar diversas fuentes de datos ya que no comparten un estándar ni existe un formato universal para representar la información multimodal de transporte, lo que se traduce en información inconexa, desestructurada y geográficamente aislada. De igual modo, el acceso a dicha información es también en la mayor parte de los casos un problema, ya que o bien los datos no son de dominio público o bien no están disponibles en un formato que permita una consulta efectiva, quedando restringidos al uso de una aplicación concreta, aun siendo ésta de libre acceso. Así mismo, no se tienen en cuenta factores relevantes que podrían incorporarse si se considerase tanto la información existente y relacionada como el contexto del usuario (preferencias, situación, etc.) a la hora de obtener los resultados. Además, como se ha comentado, en muchos casos se encuentran cerradas de modo que es muy costoso el acceso a su información.

Por otro lado, aunque la información generada por dichas herramientas de planificación ha evolucionado en los últimos años, la mayor parte de ellas siguen limitándose a ofrecer únicamente resultados en base a parámetros como la distancia o la duración del viaje, no siendo posible combinar este tipo de información con otros parámetros externos relacionados que, en muchos casos, también son relevantes a la hora de realizar dicha gestión. Es en este área donde el uso de información externa relacionada, proveniente de otras fuentes de datos o de la experiencia de los propios usuarios, y enlazada en el modelo de datos, puede suponer una mejora significativa en cuanto a la calidad de la información provista.

Hay que tener en cuenta que, tal y como se hace explícito en [Campbell 98], la efectividad de los sistemas de información de transporte depende en gran medida por un lado de la capacidad para integrar información desde diversas fuentes de transporte, y por otro, de la adecuación de la información al usuario concreto. Ambos son factores limitantes que, junto con las oportunidades detectadas al estudiar las soluciones de planificación de rutas existentes, ofrecen un marco de investigación al que se adscribe el presente trabajo.

## 1. Introducción y motivación

---

### 1.2 Trabajo de tesis

Esta tesis propone la aplicación de una arquitectura software novedosa con la que abordar ese cambio que debe producirse en la movilidad urbana, a la vez que se intentan resolver las limitaciones encontradas mediante la incorporación de tecnologías innovadoras, como la utilización de un middleware semántico para estructurar los datos, la gestión del contexto del usuario o el uso de información externa enlazada; tecnologías que ya han sido utilizadas con éxito en otras áreas de aplicación [Wang 02, Chen 03, Rosse 03].

Después de realizar un análisis crítico y en profundidad del problema encontrado en el estado del arte actual, se plantea la siguiente hipótesis a validar durante el desarrollo de la tesis doctoral. Además, se plantean una serie de objetivos que delimitan el alcance del trabajo, dividiéndose éstos en generales y específicos.

#### 1.2.1 Hipótesis

Una vez identificado el problema y con la intención de resolverlo, se plantea la siguiente hipótesis:

«Es posible desarrollar una arquitectura software que, partiendo de un conjunto heterogéneo de fuentes de datos de transporte e incorporando información relacionada relevante, sea capaz de generar información de transporte multimodal estructurada e integrada, habilitando de este modo la creación de servicios avanzados de consulta mediante la publicación distribuida e interoperable de dicha información»

Es decir, se va a desarrollar una arquitectura software distribuida, con especial énfasis en el diseño del modelo de datos y la habilitación de servicios para la consulta de información, con la que se pretenden integrar y estructurar los datos de transporte multimodal disponibles, obteniendo de este modo un modelo semántico para la representación de la información de transporte. Dicho modelo tendrá capacidad para extender y enlazar información relacionada relevante y estará soportado por herramientas que permitan la consulta efectiva de sus datos, atendiendo al contexto concreto de interés y habilitando la creación de servicios avanzados de información al viajero.

### 1.2.2 Objetivos

El objetivo general de esta tesis es el *diseño de una arquitectura software distribuida que permita, gracias a la estructuración de datos de transporte provenientes de fuentes heterogéneas junto con la integración de información relevante relacionada, ofrecer una plataforma interoperable para la habilitación de servicios software avanzados relacionados con la movilidad multimodal.*

La explotación de dicha arquitectura favorecerá la aparición de soluciones alternativas en cuanto a la gestión y la provisión de información de transporte, proporcionando nuevas opciones al usuario de cara a la planificación de sus desplazamientos o la gestión de su movilidad urbana. La arquitectura permitirá obtener información básica de transporte mediante el uso de parámetros tradicionales como el tiempo o la distancia. Sin embargo, a diferencia de las soluciones existentes, también proporcionará información más específica y concreta, mediante la agregación de información relacionada relevante. Para ello, además de la información disponible actualmente y que el sistema consumirá (rutas, distancias, etc.) será necesario disponer de una arquitectura que, primero, posibilite el que esa información sea enriquecida con datos externos (proporcionados en algunos casos por los mismos usuarios) y, segundo, que esos datos puedan ser después explotados, filtrados y visualizados en función del interés o contexto de búsqueda.

Del mismo modo, la arquitectura propuesta persigue solventar las limitaciones encontradas en los sistemas de gestión de datos de transporte existentes, en muchos casos convertidos en auténticos silos de información, donde el acceso y la integración de los datos se hace muy costoso. Esto dificulta en gran medida la aparición de herramientas enfocadas en mejorar la movilidad ciudadana, cuya funcionalidad exige necesariamente apoyarse en dicha información.

El resultado de este trabajo de tesis proporcionará nuevas herramientas para proveer de servicios a la sociedad de la información y el conocimiento, mediante el enriquecimiento y la publicación efectiva de los datos de transporte disponibles, aportando a su vez una plataforma abierta e interoperable desde la que construir las bases para una gestión de la movilidad eficiente y sostenible.

## 1. Introducción y motivación

---

### 1.2.2.1 Objetivos específicos

Los principales objetivos específicos, derivados del objetivo general y que a su vez definen las acciones necesarias para la confirmación de la hipótesis, podrían agruparse de la siguiente manera:

- ◇ *Definición y diseño de la arquitectura para la gestión de la información de transporte* [OE1]. Es decir, especificar todos los aspectos relacionados con la funcionalidad del sistema, tanto desde el punto de vista del middleware de datos, analizando y definiendo el formato en el que se modelará la información de transporte, como desde el punto de vista de la arquitectura del sistema, definiendo la integración con fuentes de información relacionada y estableciendo el acceso distribuido a la información disponible.
- ◇ *Desarrollo de la arquitectura del sistema y del middleware de datos* [OE2]. Se trata de implementar el software derivado del diseño de la arquitectura propuesta. Así, el desarrollo se centra principalmente en la implementación del modelo de datos y la definición de la arquitectura que permita la integración, contextualización y publicación de la información. Dichos componentes deberán satisfacer los requisitos establecidos de integración, interoperabilidad y acceso distribuido a la información.
- ◇ *Definición y configuración del entorno de pruebas* [OE3]. Se trata de definir un conjunto de escenarios que permitan evaluar el comportamiento del sistema y así validarlo funcionalmente. De este modo, se planteará un banco de pruebas donde se valore el desempeño y se comparen las características de la arquitectura propuesta con otros formatos y enfoques existentes para la representación y la publicación de información multimodal de transporte.
- ◇ *Evaluación de los resultados obtenidos* [OE4]. Una vez realizadas las pruebas funcionales del sistema se podrán evaluar los resultados obtenidos, valorando si éstos son los esperados y si la arquitectura propuesta supone o no una ventaja sobre los sistemas existentes según los parámetros establecidos. Dicho estudio permitirá probablemente hallar puntos de mejora que posibiliten optimizar el sistema o enfocar la investigación hacia otras áreas de interés, detectadas mediante el análisis de los resultados.



### 1.3 Metodología de investigación

La tabla 1.1 recoge el conjunto de actividades necesarias para la consecución del objetivo general y, por consiguiente, la validación de la hipótesis. Se representa también su relación con el objetivo específico al que contribuyen.

ACTIVIDADES	OBJ. ESP.
Análisis de las soluciones existentes para la representación de información de transporte	OE1
Definición y diseño del middleware de datos	OE1
Definición y diseño de la arquitectura para la distribución de la información de transporte	OE1
Desarrollo software del middleware de datos e integración de los componentes de la arquitectura	OE2
Carga de los datos de transporte y generación del contenido ontológico correspondiente	OE2
Despliegue de los servidores distribuidos y publicación de la información semántica	OE2
Desarrollo de un planificador multimodal semántico corriendo sobre la arquitectura propuesta	OE3
Definición y diseño de los escenarios de evaluación	OE3
Ejecución de las pruebas funcionales en laboratorio	OE4
Evaluación y análisis de los resultados obtenidos	OE4

**Tabla 1.1:** Actividades más relevantes y objetivo específico al que contribuyen

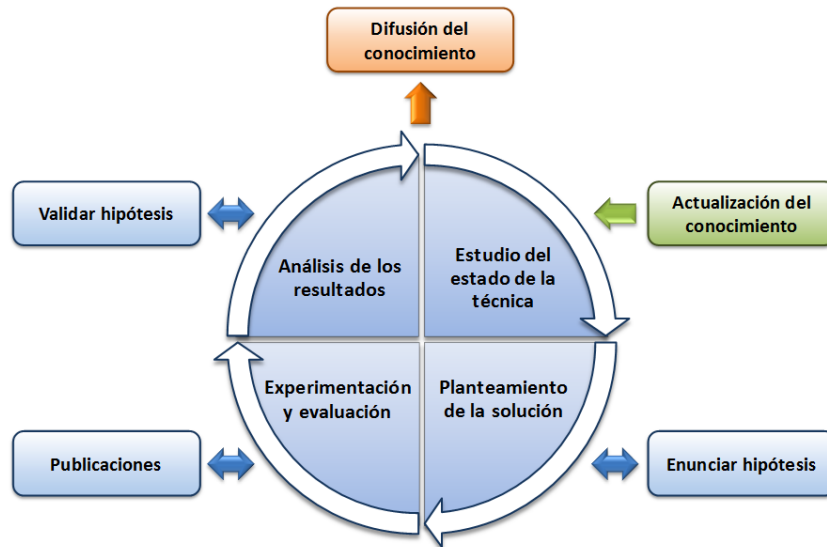
### 1.3 Metodología de investigación

El área de conocimiento a la que se adscribe la tesis, los sistemas inteligentes de transporte, es una de las que más rápido han evolucionado en cuanto a desarrollo tecnológico durante los últimos años, más aún en el ámbito de las soluciones software y la planificación de rutas. Esto hace imprescindible la utilización de una metodología de investigación que se adecúe a dicha velocidad. Por ello, es necesario un ciclo continuo de revisión del estado del arte, que contemple la aparición de posibles mejoras de forma continuada.

## 1. Introducción y motivación

---

Con estas necesidades, se plantea un proceso definido de forma iterativa y cíclica, en el que se vayan planteando soluciones cada vez más refinadas mediante la completitud de ciclos, en cada uno de los cuales se desarrollan nuevas ideas, que son puestas en práctica y comprobadas en el ciclo siguiente.



**Figura 1.1:** Metodología de investigación empleada

Esta metodología cíclica, mostrada en la figura 1.1, puede resumirse en los siguientes puntos:

- ◇ Estudio del estado de la técnica actual, mediante bibliografía relacionada y publicaciones de la comunidad científica en este ámbito, de tal manera que tengamos un conocimiento actualizado sobre el estado del arte en la materia, con el fin de detectar limitaciones, necesidades u oportunidades de mejora.
- ◇ Diseño de la solución que solvante los problemas en los que se centra la tesis. Para ello, se definirá una hipótesis y unos objetivos, basados en el análisis previo de la materia y que constituirán el enlace de unión entre los problemas y la solución.
- ◇ Experimentación y evaluación, para medir de forma cuantitativa y cualitativa el resultado obtenido mediante la aplicación de la solución diseñada, así como su comparación con otras soluciones ya existentes.

- ◇ Validación de la comunidad científica, por medio del análisis crítico de las publicaciones realizadas, de tal forma que se determine la relevancia de la investigación realizada.
- ◇ Adaptación de la solución, de acuerdo a la retroalimentación obtenida desde la comunidad científica, y a los resultados de la experimentación. Parte indispensable en el proceso iterativo de mejora continua establecido por la metodología.
- ◇ Análisis de los resultados. Una vez se han realizado las iteraciones necesarias para la corrección y mejora de la solución, se comprueban los resultados obtenidos, permitiendo la validación de la hipótesis.
- ◇ Divulgación de los conocimientos obtenidos durante la investigación, contribuyendo a la producción científica y a la generación de conocimiento, así como, en su caso, a la transferencia de dicho conocimiento a otros sectores económicos y productivos.

## 1.4 Estructura de la memoria

La presente memoria describe el trabajo de tesis realizado. Se encuentra estructurada en siete capítulos, cada uno de los cuales aborda los siguientes contenidos:

- ◇ El primer capítulo, el presente, ha proporcionado una visión global acerca del ámbito de trabajo de la tesis y de las motivaciones principales que han propiciado su realización. Se ha identificado la problemática que este trabajo pretende resolver y se han definido las características más relevantes de la solución propuesta, la hipótesis de trabajo de la tesis, los objetivos generales y específicos que guían su desarrollo, así como la metodología a seguir.
- ◇ El segundo capítulo proporciona al lector una visión general del estado del arte en las áreas de conocimiento sobre las que se apoya el trabajo de tesis. Así, se presenta una contextualización inicial de la investigación en el ámbito de los ITS para, a continuación, analizar y describir las funcionalidades y características de los sistemas de planificación de rutas más relevantes.

## 1. Introducción y motivación

---

El capítulo continúa con un análisis crítico de los distintos formatos para la representación y la publicación de información de transporte, de manera que permite entender las diferentes estrategias existentes y por qué se ha escogido la modelización semántica como vía para la estructuración de los datos. En este sentido, se presentan los conceptos clave de web semántica y *linked open data* como ejes de la arquitectura propuesta, y se describen una serie de trabajos notables en el estado del arte relacionados con su aplicación tanto en el área del transporte como en otras áreas de actuación afines a ésta.

- ◇ El tercer capítulo aborda en mayor profundidad el problema existente en la representación y la publicación de información multimodal de transporte, estudiando por qué no ha podido ser resuelto hasta el momento. Se efectúa un análisis comparativo de los distintos enfoques presentados en el estado del arte, aportando soluciones de forma teórica para la mejora de la calidad de la información provista y se presentan las oportunidades de mejora.
- ◇ El cuarto capítulo describe el desarrollo principal del trabajo de tesis: la modelización semántica de la información de transporte multimodal. Una vez expuestas las oportunidades de mejora, se presenta la solución principal de esta tesis. Inicialmente se definen tanto las características de la ontología propuesta como la metodología seguida en su construcción para, a continuación, especificar con detalle sus peculiaridades, derivadas de la necesidad de gestionar y modelar el contenido geográfico presente en el dominio del transporte.
- ◇ El quinto capítulo describe la arquitectura software que da soporte al modelo de datos desarrollado. Primero se especifican las capacidades principales de la misma, atendiendo a su naturaleza distribuida y a la adaptación al contexto concreto de interés en la consulta de información de transporte, para después profundizar en su diseño y funcionalidad, así como en las herramientas y mecanismos empleados para su despliegue.

- ◇ El sexto capítulo aborda la experimentación como medio para la validación de la hipótesis enunciada en la tesis. Se analiza la hipótesis y se justifica mediante la presentación de un banco de pruebas. El desarrollo de un planificador de rutas multimodal, corriendo sobre la arquitectura propuesta y consumiendo la información semántica proporcionada por el sistema posibilita la realización de dichas pruebas y permite por tanto comparar los distintos enfoques presentados con el trabajo realizado en la presente tesis. Se muestran, de esta manera, los resultados de la experimentación, se analizan los datos obtenidos y se comprueba la veracidad de la hipótesis.
- ◇ El séptimo capítulo recoge finalmente las conclusiones de la tesis, así como las posibilidades de mejora y las líneas futuras de trabajo.



*Un experto es una persona que ha  
cometido todos los errores posibles  
en un campo muy pequeño.*

Niels Bohr

# 2

## Estado del arte

La arquitectura a desplegar como resultado del presente trabajo se fundamenta en diversas áreas de conocimiento, tecnologías de desarrollo y paradigmas de diseño software. Se hace necesario por tanto, antes de explicitar las particularidades de la solución propuesta, realizar un análisis del estado del arte en dichas áreas. Como punto de partida y con el objetivo de situar el contexto de la investigación, se presenta el escenario principal en el que encuadrar el trabajo de tesis: los Sistemas Inteligentes de Transporte.

### 2.1 Sistemas Inteligentes de Transporte

El objetivo prioritario de una gestión adecuada del transporte es la optimización de todos los recursos existentes, al mismo tiempo que se promueven políticas modernas de transporte basadas en la innovación. Tecnológicamente es posible exponer este objetivo enfocándolo desde otro punto de vista: la tecnología ofrece medios para conseguir un intercambio inmediato de información que se transforma en un aprovechamiento más eficaz de los recursos. La tecnología hace presencia en los medios de transporte a través de la incorporación de los ITS.

## 2. Estado del arte

---

Los ITS se pueden definir como un conjunto de aplicaciones avanzadas dentro de la tecnología informática, electrónica y de comunicaciones que, desde un punto de vista social, económico y medioambiental, están destinadas a mejorar la movilidad, seguridad y productividad del transporte, optimizando la utilización de las infraestructuras existentes, aumentando la eficiencia del consumo de energía y mejorando la capacidad del sistema de transportes [Fomento 10].

Las primeras aplicaciones de sistemas de comunicaciones se desarrollaron en los años sesenta en el transporte aéreo y, desde entonces, la aplicación de las tecnologías ITS se ha venido desarrollando con más o menos intensidad para los diferentes modos de transporte. Las mejoras derivadas del uso de estas tecnologías en el ámbito del transporte son claramente evidentes. Los sistemas de ayuda a la explotación [Ozbay 99], las aplicaciones de localización de flotas [Zhao 00], las tarjetas sin contacto para el cobro, el tacógrafo digital o, en una temática más alineada con el trabajo propuesto, los sistemas avanzados de información al viajero, son ejemplos claros de los beneficios que todos los agentes involucrados en el transporte obtienen con la implantación de estas tecnologías.

### 2.1.1 Marco social, económico y medioambiental

El cambio en la política de transporte que se ha venido experimentando durante los pasados años, basada en una apuesta por la gestión de la demanda y apoyada casi exclusivamente en el lado de la oferta mediante la creciente dotación de infraestructuras, hace que los productos y servicios ITS cobren un papel fundamental también en el marco de la gestión y la promoción de la movilidad sostenible [Fomento 10].

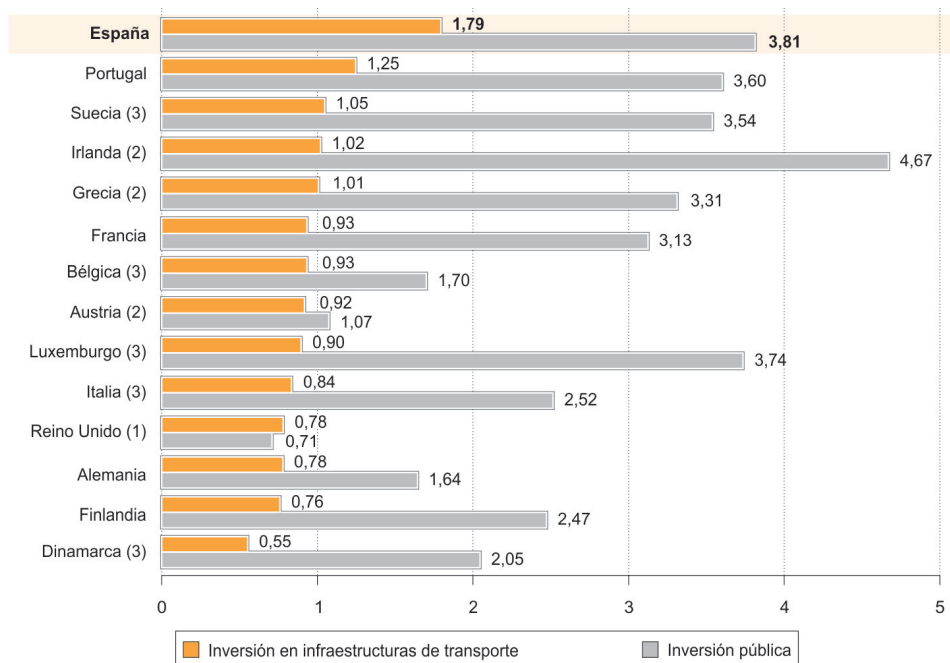
Este hecho se apoya además en una serie de factores que han permitido un despliegue relativamente rápido de estas tecnologías: agentes sociales cada vez más exigentes en materia de información en tiempo real; universalización y reducción de costes de la electrónica utilizada para captar, tratar y difundir la información; toma de conciencia y apuesta drástica por la seguridad vial, etc. Así, la inversión en infraestructuras (construyendo nuevas vías y manteniendo las existentes) forma parte de la acción de los estados de la Unión Europea (UE) para impulsar el desarrollo económico, conseguir un mejor equilibrio territorial y lograr una mayor cohesión social.



## 2.1 Sistemas Inteligentes de Transporte

Lógicamente, el esfuerzo inversor depende de la riqueza del país. España, en concreto, viene haciendo un gran esfuerzo inversor para mejorar la red de infraestructuras de transporte. Como ejemplo, según el informe [Encuentro 12] a partir de datos de la OCDE y Eurostat, en 2010 se dedicaron 410 euros por habitante a infraestructuras de transporte, mientras que en Francia o en Alemania el gasto fue de 279 y 236 euros, respectivamente. Si se analiza la inversión en relación con la riqueza, España está también a la cabeza de Europa, ya que invirtió en infraestructuras de transporte un 1,79 % del PIB en 2010, es decir, un 70 % más que en Francia, Portugal o Bélgica.

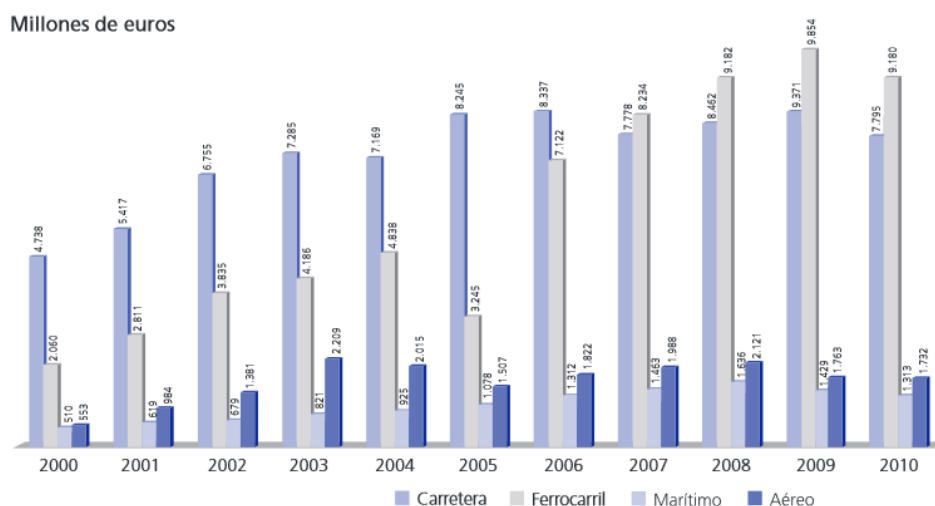
El siguiente gráfico (figura 2.1) muestra la inversión pública en infraestructuras de transporte en relación con el PIB de algunos países europeos, así como la inversión pública en todos los sectores en 2010. Como media, en Europa se invirtió un 0,8 % del PIB en infraestructuras de transporte, mientras que en España se duplicó esa cifra. La inversión pública total fue de un 3,81 % del PIB, valor también elevado en el contexto europeo. Estas inversiones en infraestructuras de transporte, realizadas en un periodo de gran inestabilidad económica, son las que se usarán durante los próximos 20 años y sobre las que tendrán que operar los ITS.



**Figura 2.1:** Inversión pública en infraestructuras en la UE (en % del PIB)

## 2. Estado del arte

El importante gasto en infraestructuras producido en la década 2000-2010 (ver figura 2.2) [Fomento 13] condiciona la necesidad actual de gestionar y mantener un número muy elevado de redes de transporte multimodal y supone una motivación específica para el desarrollo de sistemas inteligentes de información enfocados en la provisión de información de transporte integrada, que posibiliten una planificación adecuada de la movilidad urbana e interurbana, fomentando el uso de dichas infraestructuras y habilitando un uso económica y medioambientalmente sostenible de las redes de transporte disponibles.



**Figura 2.2:** Inversiones realizadas en modos de transporte e infraestructuras

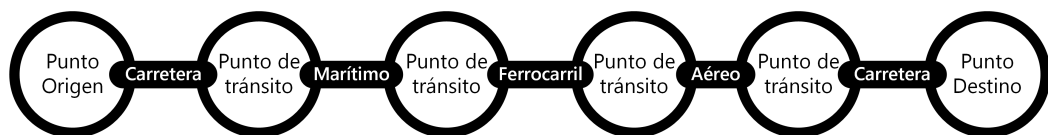
Por otro lado, en cuanto al uso real de dichas infraestructuras y redes de transporte, cabe destacar que el transporte público urbano de viajeros se aproximó a los 2.800 millones de viajeros en 2013 [Fomento 15] manteniendo la demanda de años anteriores. El transporte público, junto con los viajes a pie y en bicicleta, ha alcanzado cuotas de más del 65 % de los viajes en zonas metropolitanas como Madrid o Barcelona. En esta última, el porcentaje se eleva hasta el 85 % de los viajes realizados en el ámbito urbano. No se considera que a corto plazo se vayan a producir incrementos en la demanda que justifiquen mayores inversiones de ampliación. No obstante, sí resultarán necesarias inversiones para la mejora de los servicios existentes, área donde los ITS y en concreto las soluciones para la provisión de información integrada al viajero cobran verdadera importancia.

### 2.1.2 Intermodalidad y multimodalidad

Un modo de transporte es una tipología particular para el transporte de bienes o de personas. De forma genérica, se distinguen tres modos de transporte fundamentales: aéreo, marítimo y terrestre. Por otro lado, se conoce como cadena de transporte a la secuencia de modos de transporte aplicados para el movimiento de la carga desde su origen a su destino, incluyendo uno o más trasbordos [Fomento 14].

En este punto surge el concepto de intermodalidad: el transporte intermodal designa el movimiento de mercancías utilizando sucesivamente dos o más modos de transporte sin manipular la mercancía en los intercambios de modo. Por lo tanto, el término intermodalidad se utilizará para describir un sistema de transporte en el que dos o más modos de transporte están involucrados en el transporte de mercancías de una manera integrada, sin procesos de carga y descarga.

Por otro lado, el transporte multimodal (figura 2.3) es la articulación entre diferentes modos de transporte, a fin de realizar más rápida y eficazmente las operaciones de trasbordo de materiales y mercancías. El transporte multimodal designa el movimiento de mercancías en el que es necesario emplear más de un tipo de vehículo para transportar la mercancía desde su lugar de origen hasta su destino final, pero mediando un solo contrato de transporte. El transporte intermodal se considera así un tipo de transporte multimodal.



**Figura 2.3:** Cadena de transporte multimodal

Adicionalmente, el término multimodalidad designa la organización del transporte mediante la simultaneidad de diferentes modos para un mismo itinerario o en una zona geográfica concreta. Sin embargo, en relación con estos conceptos, además de las ineficiencias existentes en cada uno de los modos de transporte, se observa también una falta de integración entre ellos, haciendo que la carga se encuentre durante más tiempo en tránsito lo que provoca un incremento de costos y precios, y complicando los procedimientos administrativos.

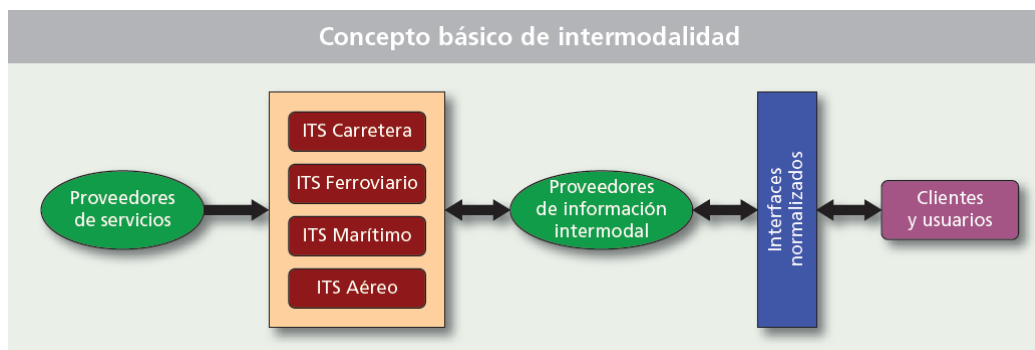
## 2. Estado del arte

Así, siendo uno de los objetivos de los ITS mejorar la movilidad de personas y mercancías, un sistema de soporte inteligente ideado para optimizar el transporte multimodal debería ser capaz de garantizar un flujo de información continuo y eficaz entre modos de transporte y servicios heterogéneos.

### 2.1.3 Arquitecturas de servicios multimodales

En la última década, de cara a la gestión de servicios ITS, cada vez más numerosos, desde la Unión Europea se está impulsando la creación de arquitecturas ITS que funcionen esencialmente como marco normalizador y de integración de los servicios de este tipo que se vayan a abordar en los próximos años. En este sentido, el Libro Blanco del Transporte [Europea 11] establece una ruta a seguir con el fin de lograr un transporte competitivo y sostenible. Centrándose en el ámbito de la multimodalidad, la UE tiene como reto optimizar la eficiencia de las cadenas logísticas multimodales. El objetivo es conseguir una red básica eficiente para el transporte interurbano y multimodal.

Cada modo de transporte, e incluso cada tipo de servicio ITS, dispone de una serie de necesidades en infraestructura que, por lo general, tienen carácter específico. Esto obliga a desarrollar un plan de despliegue para cada uno y un plan global que permita extraer la información necesaria para permitir la intermodalidad efectiva, tal y como se muestra en la figura 2.4. A su vez, la arquitectura resultante debe ser modular y fácilmente ampliable tanto en el ámbito geográfico como en número de usuarios y operadores implicados.



**Figura 2.4:** Arquitectura para la intermodalidad de servicios

Dichas arquitecturas surgen ante la necesidad de intercambio de información entre numerosos organismos o entidades que desarrollan ITS. Mejorar la integración de las redes modales conduce a mejores elecciones entre dichas opciones modales. Habrá más conexiones entre las estaciones de autobús, ferrocarril, metro, puertos o aeropuertos, transformándose en plataformas para el transporte multimodal de pasajeros, favoreciendo de este modo opciones de movilidad ciudadana más eficientes y sostenibles.

Desde el punto de vista tecnológico, se promueve la creación de las condiciones necesarias para el desarrollo y el uso de sistemas inteligentes de transporte interoperables, habilitando la creación de servicios avanzados de información al viajero. Servicios y aplicaciones que, mediante la integración y la clasificación de información relevante, permitan a los usuarios tomar decisiones informadas sobre la planificación y la gestión de sus viajes.

## 2.2 Sistemas de planificación de rutas

Para fomentar un comportamiento más sostenible y eficiente energéticamente, hay que promocionar activamente una mejor planificación de la movilidad. Para ello, tal y como se ha resaltado anteriormente, es condición indispensable que exista información ampliamente disponible sobre todos los modos de transporte, tanto de pasajeros como de mercancías, y sobre las posibilidades de su uso combinado y su impacto medioambiental. Dicha información disponible deberá estar soportada sobre una arquitectura modular, integradora e interoperable, de cara a gestionar grandes volúmenes de datos.

A este respecto, la investigación se centra en los llamados Sistemas Avanzados de Información al Viajero (ATIS), sistemas diseñados para ayudar al viajero en la planificación de sus viajes y la optimización de las rutas. Estos sistemas utilizan las TICs para recopilar, procesar y distribuir información actualizada al usuario sobre el tráfico, las condiciones de la vía, la duración del viaje, retrasos esperados, rutas alternativas o condiciones meteorológicas, dando a los viajeros la oportunidad de tomar decisiones informadas sobre cuándo viajar, qué modo de transporte utilizar o qué ruta tomar.

## 2. Estado del arte

---

### 2.2.1 Cronología y características

Múltiples estudios se han realizado en este área a lo largo del tiempo, atendiendo fundamentalmente a dos cuestiones: inicialmente el objetivo de la investigación era observar el comportamiento de los usuarios ante dichos sistemas de información, midiendo los beneficios percibidos [Al-Deek 93], la adecuación de las soluciones al usuario [Dingus 97] o su grado de aceptación [Schofer 93]. Tras dichos estudios, y en vista tanto de los resultados positivos obtenidos como del avance de las TICs en esos años, artículos posteriores centraron su atención en el diseño arquitectónico y en la información a utilizar para el correcto desarrollo de dichos sistemas [Adler 98, Szczerba 00, Caulfield 07], dando lugar a soluciones experimentales como la desplegada por Kumar en India [Kumar 05] o Gan en China [Gan 06]. La investigación realizada durante este periodo ha sido decisiva para el desarrollo de herramientas software avanzadas, como los actuales sistemas de planificación de rutas, favoreciendo también la aparición de aplicaciones comerciales en este ámbito enfocadas al uso del transporte multimodal.

Tradicionalmente, el usuario tenía la opción de planificar manualmente un viaje de largo recorrido mediante la combinación de distintos medios y agencias de transporte. Hacer este tipo de combinaciones implicaba varias horas de navegación web e incluso algunas llamadas telefónicas, necesarias para vincular las diferentes rutas, horarios y tarifas, sin garantizar que el usuario seleccionase la mejor combinación para el viaje deseado. Este es el escenario en el que un planificador multimodal puede ayudar a decidir la mejor opción para el viajero.

Un planificador de rutas multimodal es un sistema informático que puede proporcionar al viajero de forma automatizada un itinerario de transporte optimizado para la ejecución de sus desplazamientos. Dicha optimización viene dada por la provisión de información de transporte enriquecida, mediante la integración de fuentes de información con horarios, rutas y otros datos útiles y la generación de nuevo conocimiento [Delling 09, Niaraki 09]. Un solo viaje puede utilizar una secuencia de varios modos de transporte, lo que significa que el sistema debe contar con información sobre los servicios públicos de transporte (autobús, tren, avión, tranvía, metro) y sobre las redes de transporte (carreteras, senderos, bici-carriles) para el transporte privado (automóvil, a pie o en bicicleta).

A este respecto, un sistema de planificación de rutas multimodal debería ser capaz de calcular la mejor ruta (en cuanto a tiempo empleado entre un origen y un destino) basándose en todos los medios de transporte disponibles y teniendo en cuenta tiempos de conexión, trasbordos, etc. Soluciones actuales como Google Transit<sup>1</sup> o Moveuskadi<sup>2</sup> en el ámbito privativo u OpenTripPlanner<sup>3</sup> en el ámbito del software libre, han supuesto un avance muy importante en este área. Este tipo de soluciones software son un claro exponente de la aplicación de las TICs como herramienta para facilitar la toma de decisiones, en este caso en el ámbito del transporte, suponiendo una clara mejora en la gestión de la movilidad ciudadana y enfocando esta además hacia un modelo más sostenible.

A continuación se describen brevemente algunas de estas soluciones junto con sus características y funcionalidades principales.

### 2.2.2 Google Transit

Google Transit es una aplicación web para la planificación de rutas de transporte ofrecida por Google (ver figura 2.5). La herramienta permite planificar viajes combinando la información más reciente provista por empresas gestoras de transporte público con la tecnología de Google Maps, es decir, integrando cartografía actualizada. Google Transit integra información sobre paradas de transporte público, rutas, horarios y tarifas, permitiendo planificar un viaje de manera rápida y sencilla. Para las empresas de transporte público de todo el mundo es una solución rentable dirigida a viajeros tanto inexpertos como experimentados que utilizan el transporte público. La plataforma de Google ofrece además un API (Interfaz de Programación de Aplicaciones) para uso gratuito (con funcionalidad limitada) y de pago (para uso comercial), mediante la cual aplicaciones y servicios de terceros pueden acceder a la información de transporte provista por Google Transit.

En diciembre de 2005, el servicio se inició con tiempos y rutas para la ciudad de Portland, Estados Unidos. En marzo de 2007 ya incluía datos de nueve ciudades en siete estados de Estados Unidos. Actualmente, más de 800 ciudades de todo el mundo ofrecen su información mediante Google Transit.

---

<sup>1</sup> Google Transit, <https://developers.google.com/transit>

<sup>2</sup> Moveuskadi, <http://www.moveuskadi.euskadi.eus>

<sup>3</sup> OpenTripPlanner, <http://opentripplanner.org>

## 2. Estado del arte

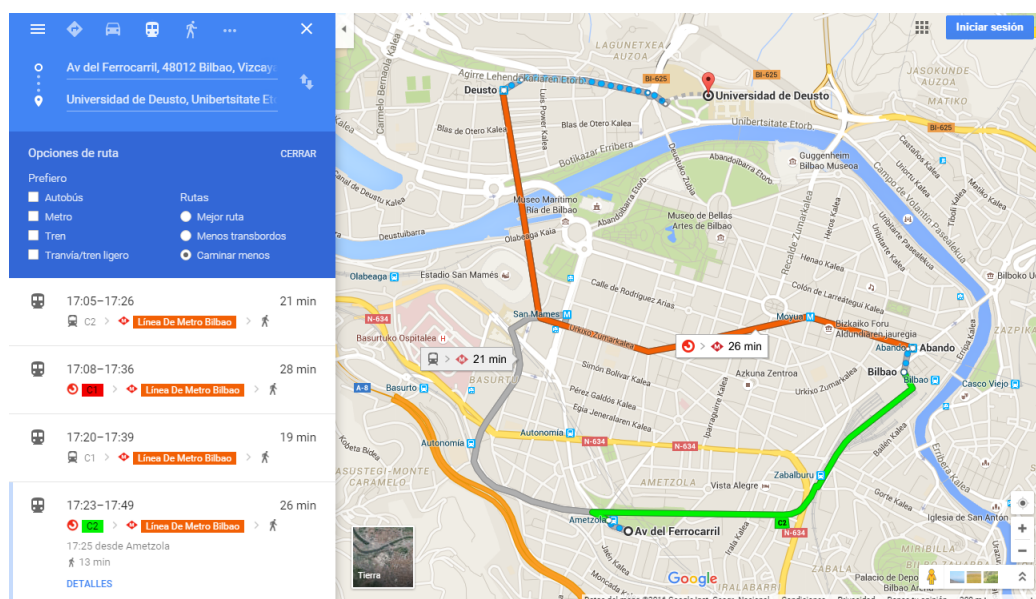


Figura 2.5: Interfaz de usuario de Google Transit

En cuanto a la información de transporte utilizada, Google Transit trabaja con la Especificación general de *feeds* de transporte público (GTFS) definiendo un formato común para los horarios de transporte público, así como para el resto de elementos que componen el dominio del transporte multimodal y la información geográfica asociada a ellos. De este modo, una empresa y/o institución gestora de transporte público puede utilizar esta especificación para publicar información de los servicios ofrecidos en una zona geográfica determinada. En el apartado 2.3.2 de la presente memoria se describen con detalle las características y la estructura de dicha especificación.

### 2.2.3 Moveuskadi

El sector del transporte es el responsable de la emisión de casi el 25 % de los gases de efecto invernadero que se emiten a la atmósfera en el País Vasco [Ihobe 12], propiciando el deterioro continuado del medio ambiente. Por ello, el Gobierno Vasco está haciendo una apuesta clara por potenciar el transporte público, no sólo mediante inversiones en nuevas infraestructuras o normativas, sino buscando la complicidad y el cambio de hábitos de la ciudadanía.



## 2.2 Sistemas de planificación de rutas

Siguiendo esa línea de actuación se desarrolló la herramienta Moveuskadi, una aplicación web accesible desde navegadores y dispositivos móviles (ver figura 2.6) que provee información de transporte público dentro de la Comunidad Autónoma Vasca y que además, mediante información referente a las emisiones de CO<sub>2</sub> para cada ruta y medio de transporte elegido, propone alternativas ecológicas a la hora de realizar los desplazamientos seleccionados. Así, el objetivo del Gobierno Vasco es promover la movilidad sostenible, mostrando a la ciudadanía cuales son las alternativas más ecológicas para viajar, tomando conciencia del impacto ambiental provocado por sus desplazamientos.

**Figura 2.6:** Interfaz de usuario de Moveuskadi

A nivel técnico, la solución propuesta desde el Gobierno Vasco delega la mayor parte de la funcionalidad en la plataforma Google Transit. Así, utiliza la solución provista por Google para la ejecución de los algoritmos de selección de rutas, como formato para la gestión de información de transporte (GTFS) y como modelo cartográfico. El proyecto cuenta con cierta relevancia dado que más de 15 entidades tanto públicas como privadas decidieron colaborar con la propuesta contando de este modo con una gran cantidad de información abierta y actualizada sobre agencias de transporte de diversa índole (metro, tranvía, autobuses urbanos e interurbanos, trenes, etc.) dentro del territorio.

## 2. Estado del arte

### 2.2.4 OpenTripPlanner

OpenTripPlanner (OTP) es una plataforma de código abierto para la planificación de viajes de carácter multimodal. Se trata por tanto de una alternativa abierta a Google Transit, solución con la cual sólo comparte el formato GTFS como mecanismo para representar la información de transporte.

Lanzado en 2009, el proyecto ha atraído a una creciente comunidad de usuarios y desarrolladores, recibiendo igualmente el apoyo de organismos públicos, empresas y agencias de transporte. Originalmente creado para TriMet, la agencia de transporte público que opera en Portland, Oregon; actualmente existen despliegues de OTP por todo el mundo siendo también OpenTripPlanner el motor de enrutamiento detrás de varias aplicaciones avanzadas de información al viajero.

Proporciona su propia cartografía, derivada del proyecto OpenStreetMap<sup>4</sup>, junto con una API REST para su uso por aplicaciones de terceros, habilitando la búsqueda de itinerarios incluyendo rutas a pie, en bicicleta, en transporte público o en coche. La figura 2.7 muestra la interfaz de usuario de la aplicación web provista.

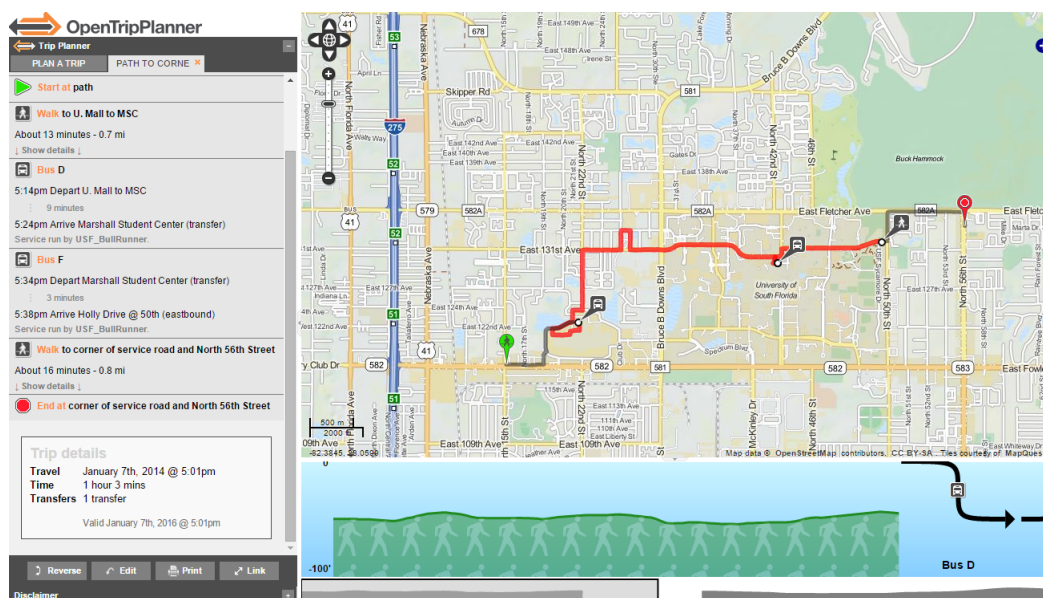


Figura 2.7: Interfaz de usuario de OpenTripPlanner

<sup>4</sup> OpenStreetMap, <http://www.openstreetmap.org>

## 2.3 Representación de la información de transporte

---

Como se ha indicado, una de las principales diferencias con soluciones como Google Transit es la apertura de su código fuente. OTP es un proyecto de software libre de código abierto bajo licencia *GNU Lesser General Public License* (LGPL) lo cual implica que puede ser modificado y distribuido libremente.

## 2.3 Representación de la información de transporte

Del análisis de los ITS y los sistemas de planificación de rutas, se puede extraer como conclusión relevante la dependencia directa de dichas soluciones en la cantidad y la calidad de los datos disponibles. Por consiguiente, cada vez adquiere más importancia la correcta y efectiva gestión y publicación de la información de transporte. Se produce, por lo tanto, un proceso complejo de recopilación, integración y transformación de datos, provenientes de diversas fuentes y en distintos formatos, con el objetivo de generar información valiosa para el usuario, y en el que entran en juego varios factores limitantes [Cali 13].

Por un lado, tal y como se ha explicitado con anterioridad, la necesidad de integrar datos desde diversos sistemas y modos de transporte implica la necesaria apertura e interoperabilidad de los datos mediante el uso de formatos y/o convenciones estándar. Por otro lado, el modelo de gestión de la información ya integrada deberá ser capaz de extender y enlazar dicha información y de proveer de herramientas que permitan la consulta de la misma, habilitando de este modo la creación de servicios en torno a ella, como pueden ser los ya citados planificadores de rutas multimodales. La efectividad de los sistemas de información de transporte, dependerá de esa capacidad para integrar información desde diversas fuentes y de la adecuación de dicha información al usuario concreto [Campbell 98].

Las compañías de transporte público y privado, como agentes involucrados en la generación y la gestión de información de tráfico, disponen de datos propios de planificación de servicios relativos a las rutas y horarios efectuados por su flota de vehículos. Actualmente existen iniciativas<sup>5,6</sup> para la publicación e intercambio de dichos datos de tránsito, lo cual está permitiendo a los desarrolladores consumirlos e integrarlos en sus aplicaciones de forma interoperable.

---

<sup>5</sup> GTFS Data Exchange, <http://www.gtfs-data-exchange.com>

<sup>6</sup> Public Feeds, <https://code.google.com/p/googletransitdatafeed>

## 2. Estado del arte

---

Se presentan a continuación las tres soluciones principales (en cuanto a su uso generalizado y el apoyo institucional recibido), impulsadas desde tres ámbitos diferentes y dirigidas al modelado y la provisión de datos de transporte: Datex II (estándar Europeo), GTFS (estándar *de facto* de Google) y WFS (iniciativa del *Open Geospatial Consortium*, OGC). Junto a ellas conviven soluciones propietarias definidas por las propias agencias de transporte. En los siguientes apartados se realiza un estudio entre los distintos formatos y arquitecturas actualmente utilizadas, incidiendo en las soluciones propuestas por cada alternativa para resolver los factores limitantes anteriormente descritos.

### 2.3.1 *DATA EXchange* (DATEX II)

*DATA EXchange* (DATEX) es un estándar europeo de intercambio de información de tráfico entre sistemas heterogéneos. Esta información incluye cualquier situación de tráfico que afecte o pueda afectar al viajero, desde un accidente que provoca una retención, al cierre de una estación de servicio en una autopista. El propósito de DATEX es facilitar el correcto intercambio de datos entre administraciones y otras entidades estableciendo modelos comunes que permitan representar la información de cualquier sistema de gestión de tráfico.

DATEX I surgió en 1996 como resultado de los esfuerzos de la UE en el dominio de los ITS. En 2003 el estándar se actualizó, ampliando el modelo de datos y favoreciendo la extensibilidad, estableciendo así la especificación DATEX II<sup>7</sup>. Dicha actualización añade, además, un modelo de consulta basado en servicios web, con la intención de facilitar la comunicación entre diferentes nodos, ampliando las posibilidades de publicación e intercambio de información [Wei-feng 08, Raines 08].

DATEX II provee de este modo un amplio catálogo de datos de tráfico estructurados en base a modelos UML y expuestos a través de servicios web WSDL. Se añaden también formatos específicos y estándar para la geolocalización de entidades, como WGS84<sup>8</sup>, permitiendo así realizar consultas y recuperar elementos geográficos. Para realizar estas operaciones utiliza peticiones HTTP SOAP, una variante de servicios web basados en el intercambio de ficheros XML.

---

<sup>7</sup> DATEX II, <http://www.datex2.eu>

<sup>8</sup> Basic Geo (WGS84 lat/long), <http://www.w3.org/2003/01/geo>

## 2.3 Representación de la información de transporte

El fragmento de código 2.1 muestra a modo de ejemplo la respuesta de un servidor DATEX II ante una petición web solicitando información acerca de los horarios de una determinada agencia de transporte. El fragmento ha sido recortado ya que el fichero original<sup>9</sup> contenía más de 6700 líneas para un total de 180 horarios.

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <d2LogicalModel xmlns:xsd="http://www.w3.org/2001/XMLSchema">
3   <exchange xmlns="http://datex2.eu/schema/1_0/1_0">
4     <supplierIdentification>
5       <nationalIdentifier>SRA</nationalIdentifier>
6     </supplierIdentification>
7   </exchange>
8   <payloadPublication xsi:type="ElaboratedDataPublication">
9     <publicationTime>2006-12-21T14:48:58.29+01:00</publicationTime>
10    <publicationCreator>
11      <nationalIdentifier>SRA</nationalIdentifier>
12    </publicationCreator>
13    <headerInformation>
14      <confidentiality>noRestriction</confidentiality>
15    </headerInformation>
16    <elaboratedData id="SE_SRA_ELABORATEDDATA_GBG107108">
17      <basicDataValue xsi:type="TravelTimeValue">
18        <time>2006-12-21T14:40:32</time>
19        <affectedLocation>
20          <locationContainedInGroup xsi:type="Linear">
21            <alertCLinear xsi:type="AlertCMethod4Linear">
22              ...
23            </alertCLinear>
24          </locationContainedInGroup>
25        </affectedLocation>
26        <travelTime>33</travelTime>
27        <freeFlowTravelTime>31</freeFlowTravelTime>
28      </basicDataValue>
29    </elaboratedData>
30    ...
31  </payloadPublication>
32 </d2LogicalModel>
```

**Código 2.1:** Respuesta DATEX. Listado de horarios

<sup>9</sup> DATEX II Example Messages, <http://datex2.eu/content/example-messages>

## 2. Estado del arte

---

Las implementaciones del estándar DATEX II se encuentran muy limitadas a las instituciones ligadas al desarrollo y financiación del proyecto. Esto es debido principalmente a la complejidad general de la arquitectura planteada, sobre todo en lo referente al modelo de datos y los mecanismos de explotación de la información, lo que ha conducido a una escasa aparición de implementaciones de terceros.

Así, la necesidad intrínseca del estándar de incorporar y clasificar toda la información de tráfico disponible, ha derivado en esquemas XML muy diversos y estructuralmente complejos. Esto, unido a la utilización de servicios web WSDL (y no tecnologías más modernas como REST) hace que el detalle de las respuestas sea en algunos casos desproporcionado (sirva como ejemplo el código 2.1), al incluir en el cuerpo del mensaje toda la información del modelo. Esto provoca que, en algunos casos, la información de transporte como tal suponga únicamente un 20 % del mensaje recibido, complicando el acceso a los datos, incrementando el tráfico de red necesario para su transmisión y favoreciendo la aparición de errores.

### 2.3.2 *General Transit Feed Specification (GTFS)*

La Especificación general de *feeds* de transporte público (GTFS) define un formato común para la gestión de los datos de transporte público y la información geográfica asociada a ellos. Esto permite a las compañías de transporte público exportar su información de tránsito haciéndola accesible. GTFS se ha situado como el estándar de facto para la representación de datos de transporte, en gran medida por su sencillez y el apoyo recibido por parte de Google.

En cuanto al formato, un *feed* GTFS se compone de una serie de archivos de texto en formato CSV (contenido separado por comas) recopilados a su vez en un fichero comprimido en formato ZIP. Cada archivo modela un aspecto particular de la información de tránsito: paradas, rutas, viajes y otros datos relacionados con la planificación. A continuación se muestran dos fragmentos (código 2.2 y 2.3) correspondientes con los archivos *agency.txt* y *routes.txt*, a modo de ejemplo de la estructura de los ficheros GTFS.

```
1 agency_id,agency_name,agency_url,agency_timezone,agency_lang  
2 TuvisaTran,Tuvisa,www.vitoria-gasteiz.org,Eu,
```

**Código 2.2:** Fichero CSV de agencias

## 2.3 Representación de la información de transporte

Como puede observarse el fichero contiene la información referente a agencias de transporte. La primera línea del fichero establece el orden y el contenido de los campos requeridos y las siguientes muestran los datos pertinentes. En este caso se muestra la información de la agencia de transportes Tuvisa-Euskotran, que opera en Vitoria-Gasteiz con dos medios de transporte: autobús y tranvía.

```
1 route_id,agency_id,route_short,route_long,route_type,color
2 L1,TuvisaTran,L1,Circular,3,FFFFFF,
3 L4,TuvisaTran,L4,Lakua-Mariturri,3,FFFFFF,
4 T1,TuvisaTran,T1,Ibaiondo,0,8CC63F,
5 T2,TuvisaTran,T2,Abetxuko,0,8CC63F,
6 G1,TuvisaTran,G1,Lakua-Abetxuko,3,F499C1,
```

**Código 2.3:** Fichero CSV de rutas

En cuanto a la publicación de los datos, es destacable reseñar que no existe un sistema estandarizado para compartir la información GTFS. Así, la publicación de los datos de transporte por parte de las agencias se realiza en muchos casos simplemente alojando dicho fichero ZIP en un servidor web o directamente enviándoselo a Google, que lo utilizará para alimentar su propio sistema: Google Transit.

Esto conlleva evidentes problemas a la hora de hacer interoperable dicha información. En primer lugar, un desarrollador que quiera contar con esos datos puede no conocer el servidor en el que está alojada la información. También es probable que dicha información no se actualice; teniendo en cuenta además que, en el caso de hacerlo, el desarrollador tampoco será consciente de ello, debiendo consultar el servidor cada vez. Por otro lado cada agencia de transporte almacenará sus datos en servidores diferentes, de modo que el desarrollador deberá gestionar todos ellos en busca de información relevante.

Se puede inferir por tanto que la inexistencia de un modelo de publicación estándar esta favoreciendo la creación de silos de información con datos desactualizados y no interoperables. La tabla 2.1 resume dichas carencias mediante la comparativa entre los principales formatos para la representación de la información multimodal y el conjunto de características consideradas deseables para la correcta compartición de la misma.

## 2. Estado del arte

---

### 2.3.3 Web Feature Service (WFS)

Los *Web Feature Service* (WFS) son otra familia de los servicios orientados a facilitar el intercambio de información geográfica a través de Internet y desarrollados según especificaciones del *Open Geospatial Consortium* (OGC) en cuanto a código abierto e interoperabilidad.

A través de WFS es posible la descarga de capas de información geográfica vectorial completa, es decir, la descarga de su geometría y de su tabla de atributos asociada. Del mismo modo, proporciona una interfaz de comunicación para realizar solicitudes de características geográficas concretas (denominadas *features*) a través de la web. WFS permite realizar consultas y recuperar elementos geográficos e información relacionada. Para realizar estas operaciones se utiliza el lenguaje GML (*Geography Markup Language*), una variante de XML específica para tratar con información geográfica, a través del cual se transmiten estas operaciones.

El fragmento de código 2.4 muestra a modo de ejemplo la respuesta de un servidor WFS ante una petición web *GetFeature* solicitando los distritos municipales (una de las *features* expuestas por el servicio).

```
1 <wfs:FeatureCollection
2   xmlns:gml="http://www.opengis.net/gml"
3   xmlns:wfs="http://www.opengis.net/wfs"
4   xmlns:ogc="http://www.opengis.net/ogc"
5   xsi:schemaLocation="http://b5m.gipuzkoa.net/ogc/wfs/
6     gipuzkoa_wfs?REQUEST=DescribeFeatureType&TYPENAME=DISTRICT">
7   <gml:featureMember ms:DISTRICT gml:id="DISTRICT.17">
8     <gml:boundedBy srsName="EPSG:25830">
9       <gml:lowerCorner>562459.63 4792350.85</gml:lowerCorner>
10      <gml:upperCorner>562459.63 4792350.85</gml:upperCorner>
11    </gml:boundedBy>
12    <ms:msGeometry srsName="EPSG:25830">
13      <gml:pos>562459.63 4792350.85</gml:pos>
14    </ms:msGeometry>
15    <ms:NAME>6 Sektorea</ms:NAME>
16    <ms:TYPE>Poligono industrial</ms:TYPE>
17  </gml:featureMember> ...
18 </wfs:FeatureCollection>
```

**Código 2.4:** Respuesta WFS. Listado de distritos municipales



## 2.3 Representación de la información de transporte

---

WFS define interfaces para describir las operaciones de manipulación de datos disponibles, incluyendo la consulta y la gestión de instancias geográficas. Las implementaciones de la norma son sin embargo escasas, debido principalmente a la complejidad del modelo de datos y el nivel de detalle de las consultas requeridas.

La consulta de los datos de transporte queda limitada al conjunto de interfaces y servicios que decida exponer el administrador del servidor. Esto conlleva problemas a la hora de hacer interoperable dicha información con la de otros servicios WFS y complica su utilización, dado que el desarrollador debe conocer de antemano las capacidades o métodos públicos así como su especificación concreta.

### 2.3.4 Soluciones Ad-hoc

Existen también otros sistemas para representar y/o proveer información de transporte, en su mayor parte definidos por las propias agencias o instituciones que se encargan de obtener los datos. Uno de los más destacados dentro de esta categoría es TransXChange<sup>10</sup> usado en Reino Unido y Australia para el intercambio de información relacionada con la planificación del servicio de autobuses.

Como se ha podido comprobar, diversos sistemas coexisten a la hora de representar la información de tránsito. DATEX II es el estándar europeo para el intercambio de datos de tráfico. Sin embargo, la gran variedad de información manejada y el complejo modelo de datos han complicado su implementación efectiva. Por otra parte, la especificación WFS del OGC, está más enfocada a la realización de consultas sobre datos geoespaciales almacenados en GIS pero establece un sistema similar de intercambio de datos a través de servicios web SOAP y mensajes GML lo que deriva también en una utilización compleja.

Las soluciones ad-hoc, en cambio, son óptimas para su dominio de actuación y sencillas de utilizar, pero no son interoperables debido a sus inherentes características propietarias. GTFS es el estándar *de facto* en cuanto al almacenamiento de datos de tránsito y también es sencillo de utilizar por lo que cuenta con un alto grado de implantación. Sin embargo, el acceso y la consulta de sus datos no es trivial, al no encontrarse estos estructurados ni formalmente enlazados.

---

<sup>10</sup> TransXChange, <http://www.dft.gov.uk/transxchange>

## 2. Estado del arte

La tabla 2.1, a modo de resumen general, presenta una comparativa entre los formatos descritos para la representación de información de transporte de acuerdo a un conjunto de características consideradas relevantes para la correcta gestión, integración y compartición de la misma.

	DATEX II	GTFS	WFS	Soluciones Ad-hoc
<b>Clasificación</b>	<i>Open Data</i>	<i>Open Data</i>	<i>Open Data</i>	Privada
<b>Estructura</b>	XML	CSV	GML (XML)	Variable
<b>Extensibilidad</b>	Sí	No	No	No
<b>Enlazabilidad</b>	No	No	No	No
<b>Consultabilidad</b>	Servicio Web	Programática	Servicio Web	API
<b>Acceso a datos</b>	Limitado	Completo	Limitado	Variable

**Tabla 2.1:** Formatos para la representación de la información de transporte

A continuación se describen con mayor detalle cada una de esas características.

- ◇ *Clasificación.* Define el tipo de restricciones legales asociadas a los datos provistos (p. ej. *Open Data*: de acceso abierto, uso libre y redistribuible; Privada: de uso exclusivo por la organización gestora de los datos; o Comercial: de uso comercial supeditado al pago de una licencia).
- ◇ *Estructura.* Se refiere al formato en el que los datos están almacenados. Es importante clasificar los datos en estructurados: formatos legibles por una máquina (CSV, XML, una taxonomía, etc.), y no estructurados (p. ej. un archivo PDF) donde no es trivial el acceso a los datos de forma automatizada.
- ◇ *Extensibilidad.* Define la capacidad intrínseca del formato para extender su información con otros datos relevantes enriqueciendo así los datos originales. Este tipo de formatos contarán con algún mecanismo que permita extender los datos mediante su relación con otros conjuntos de datos.
- ◇ *Enlazabilidad.* Expone la capacidad del formato para ser incluido en otro conjunto de datos como una referencia (por ejemplo, proveyendo una URI).

## 2.3 Representación de la información de transporte

---

Mediante este mecanismo varios conjuntos de datos pueden estar enlazados entre sí, de modo que su información pueda ser extendida.

- ◇ *Consultabilidad.* Hace referencia a la disponibilidad de la información en relación con el mecanismo de consulta de la misma. Se revisa de este modo si los datos se pueden consultar de forma semiautomatizada y se indica el mecanismo utilizado para ello (p.ej. programáticamente, mediante un API, un servicio web, etc.).
- ◇ *Acceso a datos.* Establece la completitud de la información proporcionada. Se comprueba que toda la información se encuentra disponible y pueda ser consultada de forma efectiva.

De acuerdo a estas características, DATEX II proporciona una estructura estable y con posibilidad de extenderse mediante ampliaciones del modelo de datos. También proporciona una interfaz para el acceso a datos basada en el uso de servicios web SOAP, pero su utilización es compleja debido al uso de ficheros XML sobre-descriptivos, ofuscando los datos recibidos y penalizando el rendimiento en cuanto al procesamiento necesario.

GTFS por su parte, proporciona un acceso completo a los datos, aunque no provee de ningún mecanismo avanzado para ello. Será el desarrollador quien deba localizar en primer lugar el servidor en el que se encuentra alojada la información para después acceder a ella de forma programática, recorriendo y filtrando el contenido según su interés, al no existir una interfaz estándar definida para ello.

El formato WFS sí proporciona una interfaz estandarizada, pero las consultas, al igual que en DATEX, están predefinidas por lo que el acceso a la información disponible es limitado. Así mismo su utilización es también más compleja debido al uso de ficheros GML que añaden mucha información de cabecera innecesaria penalizando el rendimiento en cuanto al ancho de banda necesario.

En cuanto a las soluciones Ad-hoc, a nivel general se pueden destacar como características fundamentales el carácter cerrado de los datos y el uso mayoritario de APIs internas propietarias de cara al manejo de la información provista.

### 2.4 Interoperabilidad y enriquecimiento de los datos

Uno de los conceptos principales sobre los que se fundamenta este trabajo doctoral es la mejora en la información de transporte que actualmente se provee al usuario. En los últimos años, tal y como se ha descrito, la tendencia tanto general en cuanto a los servicios al usuario como en el ámbito específico de la información de transporte es buscar la interoperabilidad de los datos y de este modo el enriquecimiento de la información [Doan 12, Dong 13].

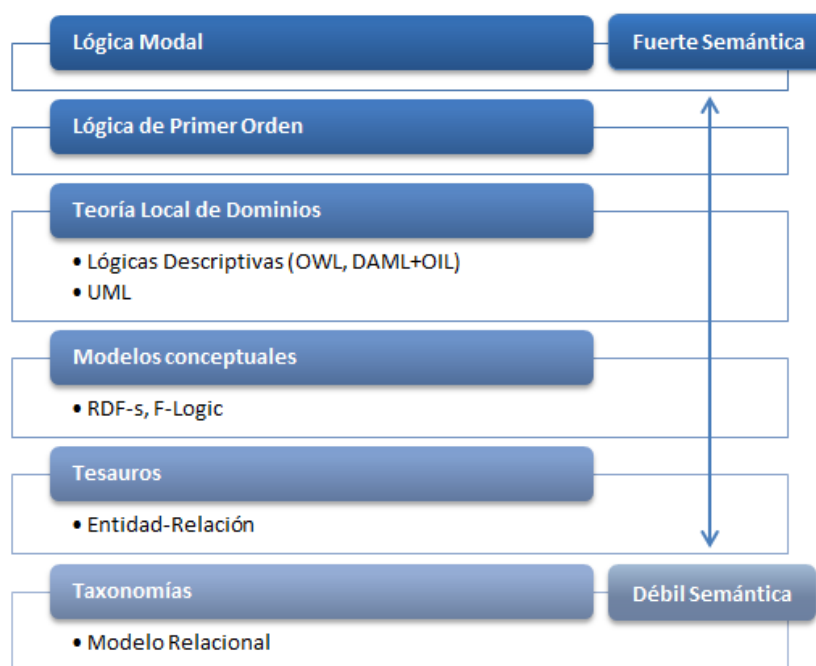
Con la idea de dar soporte a dicha tendencia integradora y de dar solución a las limitaciones encontradas en los formatos existentes en cuanto a la compartición y la publicación de nueva información se han estudiado las posibilidades ofrecidas por paradigmas de relativamente reciente aparición en el ámbito de la computación como son las ontologías, la web semántica y los datos enlazados. Los siguientes apartados describen dichos enfoques y cómo pueden ser aplicados al ámbito objeto de estudio en esta tesis, el transporte multimodal.

#### 2.4.1 Modelos formales y ontologías

El término ontología se utiliza en filosofía como una disciplina que estudia la naturaleza y organización de la realidad. En computación el término se adopta con ciertos matices, refiriéndose a la formulación de los conceptos y su relación taxonómica dentro de dominio, de forma rigurosa y exhaustiva. El objetivo es facilitar la comunicación y la compartición de la información entre diferentes sistemas y entidades, incluyendo a las personas. Se trata, dando una definición simplificada, de un sistema de clasificación particular de cierta visión del mundo [Guarino 98].

Un enfoque ontológico puede satisfacer de este modo los requisitos de compartición y publicación de información en contraste con otras alternativas que suelen fallar en la riqueza y calidad de la información representada así como caer en la ambigüedad. El trabajo realizado desde los años 70 por la comunidad de inteligencia artificial mostró evidencias de que las ontologías formales podían ser usadas como mecanismo para especificar y reutilizar conocimientos entre diferentes entidades [Gruber 95, Studer 98]. En la figura 2.8 podemos ver cómo se sitúan las ontologías dentro de los métodos y lenguajes de representación del conocimiento (adaptado de los trabajos de Ruckhaus [Ruckhaus 05]).

## 2.4 Interoperabilidad y enriquecimiento de los datos



**Figura 2.8:** Lenguajes para la representación del conocimiento

Es posible expresar ontologías en su versión más simple, a través de la descripción de conceptos únicamente; a través de modelos conceptuales, definiéndose en este caso tanto los conceptos como las propiedades y, por último, por medio de lenguajes lógicos, incluyendo restricciones además de conceptos y propiedades.

Independientemente de la codificación concreta de una ontología, es posible identificar los elementos que pueden componerla. La definición de ontología de [Gruber 93]: “*descripción formal y explícita de conceptos, considerados como un conjunto de entidades, relaciones, funciones, instancias y axiomas*” recoge los elementos principales. A continuación se describen con mayor detalle:

- ◇ *Conceptos o Clases*. Son las ideas básicas que se intentan formalizar. Los conceptos pueden ser clases de objetos, métodos, estrategias, procesos, etc.
- ◇ *Relaciones*. Representan la interacción y enlace entre los conceptos del dominio. Pueden formar la taxonomía del dominio.
- ◇ *Instancias*. Se utilizan para representar objetos determinados de un concepto.

## 2. Estado del arte

---

- ◇ *Funciones*. Son un tipo de relación donde se identifica una instancia mediante el cálculo de una función que considera varias instancias de la ontología.
- ◇ *Axiomas*. Son teoremas que se declaran sobre relaciones que deben cumplir las instancias de la ontología, permiten junto con la herencia de conceptos, inferir conocimiento que no esté indicado explícitamente en la taxonomía de conceptos. Mediante los axiomas se tendría mayor capacidad expresiva del dominio almacenado.

Es ampliamente aceptado que el uso de ontologías en inteligencia artificial aporta numerosos beneficios y funcionalidad adicionales. Sin embargo su uso más extendido es el de representar conocimiento sin mayor pretensión que ayudar a su comprensión de cara a las personas, sin que sea procesado computacionalmente. La formalización de ontologías es el primer paso para hacerla computable y, en consecuencia, tener las herramientas para aprovechar todo su potencial.

### 2.4.1.1 Metodologías para el diseño de ontologías

El diseño y generación de ontologías debe seguir un proceso que permita guiar y evaluar el modelo generado. No existen metodologías ampliamente aceptadas pero podemos encontrar estudios que ayuden en esta tarea. Un ejemplo son los criterios identificados por Gruber [Gruber 95]:

- ◇ *Claridad*. El objetivo es comunicar de manera efectiva el significado de los elementos definidos. Esta definición debe ser objetiva, independiente del contexto social y computacional.
- ◇ *Coherencia*. Debe soportar inferencias consistentes con las definiciones.
- ◇ *Extensibilidad*. Debe permitir que se añadan nuevos términos para usos más especializados sin necesidad de revisar las definiciones existentes.
- ◇ *Sesgo de codificación mínimo*. La especificación no debe estar ligada a la codificación de la ontología.
- ◇ *Mínimo compromiso*. Una ontología debe contener el menor número de suposiciones sobre el mundo modelado.

## 2.4 Interoperabilidad y enriquecimiento de los datos

---

También existen metodologías más detalladas que pretenden establecer una estrategia para el desarrollo guiado de ontologías. Ya en 1995 Uschold y King [Uschold 95] definieron una metodología proponiendo como fases: identificar el objetivo de la ontología, construirla, evaluarla y documentarla.

Una metodología más ambiciosa es la denominada *Methontology* que abarca todo el ciclo de vida [Fernández 97], equiparando la ontología a un producto software. Noy y McGuinness [Noy 01] por su parte, describen un proceso iterativo formado por varias etapas: determinar el dominio y alcance, considerar la reutilización de ontologías existentes, enumerar los términos importantes, definir clases y su jerarquía, definir propiedades, definir restricciones y crear instancias.

En 2009, Suarez-Figueroa [Suárez 09] presenta un nuevo enfoque con la metodología NeOn, un conjunto de nueve escenarios para la construcción de ontologías y redes de ontologías, haciendo hincapié en la colaboración y el dinamismo, la reutilización de los recursos ontológicos y no ontológicos, la reingeniería y la fusión de los datos.

### 2.4.2 Web Semántica

La Web Semántica es una visión extendida de la Web, dotando de mayor significado a toda su infraestructura, incluyendo sus contenidos. En la Web tenemos acceso a millones de recursos, lo que ha proporcionado su éxito pero también ha originado problemas de sobrecarga de información y de heterogeneidad de las fuentes. La Web Semántica ayuda a resolver estos problemas, reduciendo los esfuerzos cognitivos de los usuarios para delegarlos en agentes que puedan razonar, combinar y hacer deducciones para ofrecer la información adecuada a las personas.

La Web Semántica, como infraestructura basada en metadatos, aporta un camino para razonar en la Web, extendiendo así sus capacidades. Para obtener esa adecuada definición de los datos, se utilizan esencialmente los lenguajes RDF, SPARQL, y OWL, mecanismos que ayudan a convertir la Web en una infraestructura global en la que es posible compartir y reutilizar datos y documentos entre diferentes tipos de usuarios [W3C 15b]. A continuación se describen brevemente:

- ◇ *Resource Description Framework (RDF)*. Marco para generar información descriptiva simple sobre los recursos que se encuentran en la Web.

## 2. Estado del arte

- ◇ *SPARQL*. Lenguaje de consulta sobre RDF, que permite hacer búsquedas sobre los recursos de la Web Semántica utilizando distintas fuentes de datos.
- ◇ *Ontology Web Language (OWL)*. Lenguaje para desarrollar ontologías o vocabularios específicos en los que asociar esos recursos.

Las ontologías se encargan de definir los términos utilizados para describir y representar un área de conocimiento, son utilizadas por los usuarios, las bases de datos y las aplicaciones que necesitan compartir información [Gomez 02].

Además de estos tres, la web semántica se sustenta en otros lenguajes y protocolos. La pila completa, presentada por Tim Berners-Lee [Berners-Lee 01] y adaptada por [Fernández 06] se presenta en la figura 2.9.

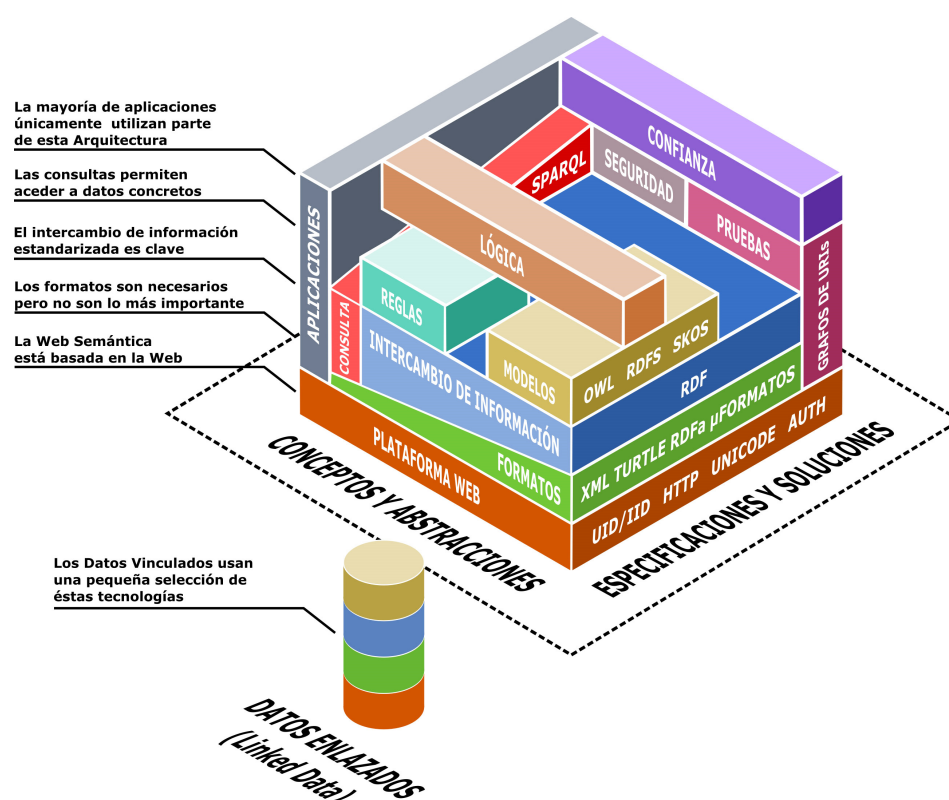


Figura 2.9: Pila completa de la Web Semántica



## 2.4 Interoperabilidad y enriquecimiento de los datos

---

Aunque RDF es conocido como un lenguaje para ontologías debido a que permite expresar clases y propiedades, así como definir estructuras jerárquicas, se trata de un lenguaje poco conveniente para describir elementos en detalle. Las principales carencias de RDF son la imposibilidad de definir participación y cardinalidad de conjuntos y la ausencia de propiedades como la transitividad, simetría, etc. Por esta razón es más conveniente modelar a través de OWL u otros lenguajes basados en lógicas descriptivas, sirviendo RDF como apoyo para definir elementos sencillos o vocabularios de un subdominio específico. En general, el modelado a través de lenguajes como OWL y RDF aporta las siguientes ventajas:

- ◇ RDF y OWL son lenguajes ricos en expresividad y permiten representar la variedad de términos involucrados en el ámbito del transporte: agente, ruta, parada, viaje, horario, etc.
- ◇ Debido a la representación explícita y formal de las semánticas asociadas a estos lenguajes, es posible implementar motores de inferencia lógicos.
- ◇ OWL y RDF han sido diseñados como estándares y tienen el apoyo de colectivos conocidos y reconocidos, como el *World Wide Web Consortium* (W3C).
- ◇ Al utilizar lenguajes de la Web Semántica, tenemos disponibles múltiples herramientas que apoyen, en mayor o menor medida, el proceso de generación y utilización de ontologías.

### 2.4.3 Datos enlazados

Una vez establecidos los formatos (RDF y OWL) sobre los que modelar vocabularios concretos, generando de este modo ontologías para dominios específicos y habiendo definido la web semántica como infraestructura para darles soporte, el siguiente paso es interrelacionar la información contenida en dichos dominios o, dicho de otra forma, enlazar los datos [Noy 04].

En informática, los datos enlazados (capitalizados como *Linked Data*, LD) describen un método para la publicación de información estructurada, de modo que pueda ser interconectada y así resultar más útil. Se basa en tecnologías Web estándar, como HTTP, RDF y URIs, pero en vez de utilizarlas para servir páginas

## 2. Estado del arte

---

web a las personas, las extiende para compartir información de una manera que puede ser procesada automáticamente por ordenador. Esto permite interconectar y consultar datos de diversas fuentes [Berners-Lee 09b].

Tim Berners-Lee, director de la W3C, acuñó el término en una nota de diseño que trataba de cuestiones relativas al proyecto de la Web Semántica donde definió cuatro principios que caracterizan los datos enlazados:

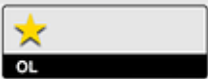




1. Utilizar URIs para identificar los recursos.
2. Aprovechar las URIs bajo HTTP para que tanto las personas como los ordenadores puedan localizar y consultar (“desreferenciar”) estos recursos.
3. Proporcionar información útil acerca del recurso cuando la URI haya sido desreferenciada, haciendo uso de estándares como RDF o SPARQL.
4. Incluir enlaces a otros recursos relacionados (mediante sus URIs) de forma que se potencie el descubrimiento de información en la Web.

En 2010, con el objetivo de estandarizar y medir la calidad de los datasets publicados hasta ese momento, que ya empezaban a ser considerables en número, se introdujo una métrica conocida como “las 5 estrellas de los datos enlazados”. Básicamente, se trata de un conjunto de características incrementales que deben cumplir los datos publicados para ser considerados como datos enlazados según criterio de la comunidad. Cuantas más funcionalidades cumpla el conjunto de datos, mejor será la calidad de este. La tabla 2.2, adaptada de la información en [Hausenblas 15] muestra dicha clasificación de forma gráfica.

### 2.4.3.1 Datos enlazados abiertos

Durante los últimos años, el paradigma de los datos enlazados ha encontrado un área de aplicación a explotar derivada de la publicación de información bajo licencias libres. El movimiento *Open Data*, que tiene como objetivo liberar grandes cantidades de conjuntos de datos mediante licencias no privativas, habitualmente provenientes de autoridades gubernamentales locales [Ding 12], adoptó la tecnología *Linked Data* y aplicó sus mejores prácticas para conseguir la publicación de una multitud de conjuntos de datos diferentes y relacionados entre sí.

## 2.4 Interoperabilidad y enriquecimiento de los datos

Estrellas	Descripción	Acrónimo	Ejemplo
	Datos disponibles en la web en cualquier formato	On-Line	PDF
	Datos estructurados y legibles por una máquina	Readable	XLS
	Datos disponibles en un formato no propietario	Open Format	CSV
	Datos referenciables según estándar de la W3C	URI	RDF
	Datos con enlaces hacia otros datasets Linked Data	Linked Data	RDF

**Tabla 2.2:** Las cinco estrellas de los datos enlazados

Los datos enlazados abiertos (LOD) son una especialización de los datos enlazados en donde la información se publica bajo licencias libres como *Creative Commons* (CC). En el ámbito de la Unión Europea existen multitud de proyectos que involucran datos enlazados y que han obtenido financiación bajo el Séptimo Programa Marco de Investigación y Desarrollo Tecnológico<sup>11</sup>. A continuación se citan algunos de los proyectos más relevantes:

- ◇ *Linked open data around the clock (LATC)* [LATC 11] cuyo objetivo es facilitar a usuarios y organizaciones el proceso de publicación y consulta de datos enlazados, incrementando de este modo el número, la calidad y la precisión de los enlaces entre los conjuntos de datos.
- ◇ *PlanetData* [PlanetData 10] pretende definir una comunidad interdisciplinar de investigadores europeos que ayuden a las organizaciones a publicar sus datos en la web de forma útil. El fin último es administrar de forma efectiva y eficiente una enorme cantidad de datos, favoreciendo la transición hacia una sociedad europea del conocimiento.

<sup>11</sup> 7th Framework Programme for Research and Technological Development (FP7), <http://ec.europa.eu/research/fp7>

## 2. Estado del arte

---

- ◇ *Data and Platform as a Service (DaPaaS)* [DaPaaS 13] se centra en proporcionar un entorno donde los desarrolladores puedan publicar y alojar tanto conjuntos de datos como aplicaciones con un uso intensivo de datos. El objetivo del proyecto es dirigirse hacia una plataforma de servicios que reduzca la complejidad tanto en la publicación como en el consumo de los datos.
- ◇ *Linked Open Data 2 (LOD2)* [LOD2 10, Auer 12] es un proyecto europeo a gran escala cuyo reto es convertir a la web de datos en una realidad viable.

Para ello se propone mejorar la coherencia y la calidad de los datos publicados, reducir la diferencia en rendimiento entre las bases de datos relacionales y el manejo de datos RDF, mejorar la confianza en la web de datos y reducir de forma general la barrera de entrada tecnológica a los usuarios.

Como se puede observar, el impulso al uso de los datos enlazados es uno de los principales objetivos de la Unión Europea, siendo el Portal de datos abiertos<sup>12</sup> uno de los activos más importantes en este sentido.

El Portal de datos abiertos es un punto único de acceso a gran variedad de datos de las instituciones y otros organismos de la Unión Europea. Miles de conjuntos de datos se encuentran disponibles para que cualquiera los pueda reutilizar y enlazar libremente con fines comerciales o no comerciales.

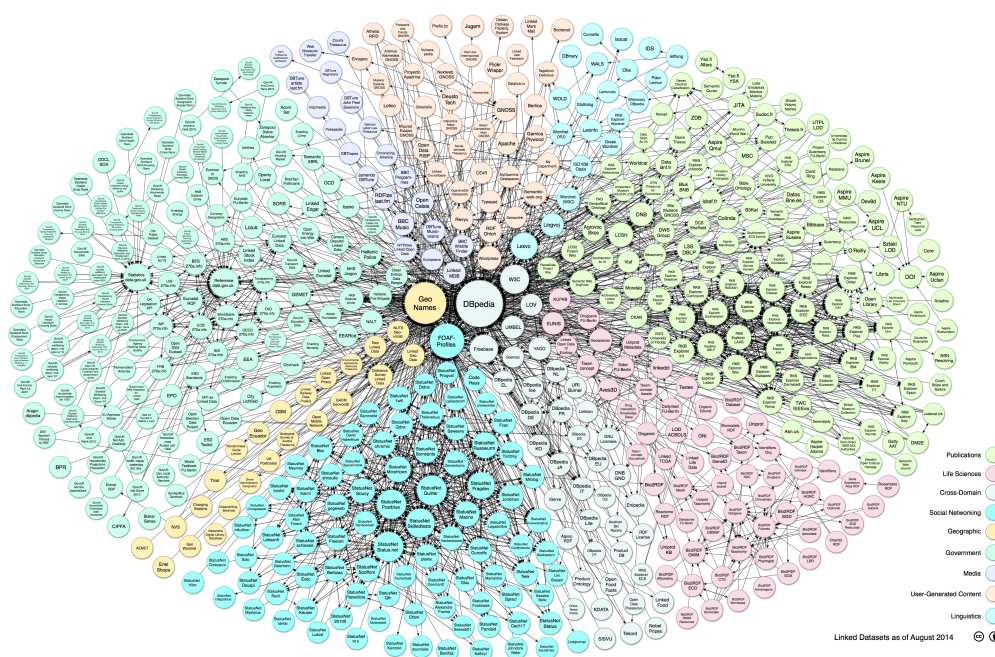
Desde la W3C el proyecto *Linking Open Data* [W3C 15a] desarrollado por el grupo encargado de divulgar y explicar la Web Semántica (*Semantic Web Education and Outreach*) tiene como objetivo extender la web con una base de datos común mediante la publicación de bases de datos en RDF y mediante el establecimiento de enlaces RDF entre datos de diferentes fuentes.

En octubre de 2007 se contabilizaron conjuntos de datos de más de dos mil millones de tripletas, relacionados entre sí por medio de más de dos millones de enlaces RDF [Fensel 11]. En septiembre de 2011, estos datos ya habían crecido hasta los treinta y un mil millones de tripletas RDF y los más de quinientos millones de enlaces. La figura 2.10 muestra el estado del mapa de datos enlazados abiertos a fecha de Agosto de 2014.

---

<sup>12</sup> Portal de datos abiertos de la Unión Europea, <http://open-data.europa.eu>

## 2.4 Interoperabilidad y enriquecimiento de los datos



**Figura 2.10:** Mapa de datos enlazados abiertos

Dichos esfuerzos, están de este modo fructificando, dando como resultado la publicación de datos bajo la infraestructura de la web semántica [Heath 11]. A continuación se destacan algunos de los conjuntos de datos más importantes:

- ◇ *DBpedia*<sup>13</sup>. Conjunto de datos extraído de Wikipedia. Contiene unos 3,4 millones de conceptos descritos mediante mil millones de tripletas, incluyendo resúmenes en once idiomas.
- ◇ *GeoNames*<sup>14</sup>. Aporta descripciones RDF de más de 7.500.000 de elementos geográficos y geopolíticos de todo el mundo.
- ◇ *FOAF*<sup>15</sup>. *Friend Of A Friend*. Ontología ampliamente referenciada, utilizada para describir a las personas, sus actividades y sus relaciones.
- ◇ *Bio2RDF*<sup>16</sup>. Proyecto de código abierto que pone a disposición del usuario la mayor red de datos enlazados sobre ciencias de la salud y biología.

<sup>13</sup> DBpedia, <http://wiki.dbpedia.org>

<sup>14</sup> GeoNames, <http://www.geonames.org>

<sup>15</sup> FOAF, <http://www.foaf-project.org>

<sup>16</sup> Bio2RDF, <http://bio2rdf.org>

## 2. Estado del arte

---

- ◇ *Data.gov*<sup>17</sup>. Web oficial del gobierno de Estados Unidos desde donde expone más de 1000 conjuntos de datos públicos disponibles como datos enlazados (aproximadamente 6.400 millones de tripletas).
- ◇ *datahub.io*<sup>18</sup>. Registro de datos enlazados abiertos, gestionado y administrado por la *Open Knowledge Foundation*<sup>19</sup>, organización internacional sin ánimo de lucro dedicada a promover la apertura y la divulgación del conocimiento.

### 2.5 Hacia la semantización de datos de transporte

Las capacidades para promover la compartición de conocimiento, la estructuración de dicho conocimiento y la interoperabilidad entre sistemas han favorecido el uso de ontologías y datos enlazados en multitud de ámbitos y dominios de aplicación diferentes (p. ej. medicina [Ashburner 00, Mol 02], educación [Guangzuo 04, Jia 11], logística [Dong 08, Giménez 08] o turismo [Abowd 97, Buján 13]).

Habiendo identificado las necesidades de interoperabilidad en el ámbito de los ITS y las posibilidades de los modelos formales, la web semántica y los datos enlazados de cara a proporcionar una arquitectura base sobre la que desplegar servicios avanzados de consulta de información, no resulta extraño encontrar trabajos similares en el ámbito de la información de transporte. Trabajos motivados en base a diversos objetivos, metodologías o resultados esperados, pero todos ellos bajo la premisa de la modelización semántica de los datos.

Se pretenden mostrar a continuación los trabajos más relevantes en este aspecto. Para ello, se ha realizado una clasificación inicial atendiendo a la finalidad y la cronología de la investigación realizada. Así, los trabajos expuestos inicialmente buscan la modelización de la información geográfica (puntos, líneas, polígonos, etc.) de forma directa, como mecanismo de formalización objetiva. Esto permite realizar búsquedas conceptuales no basadas en palabras clave, eliminando la ambigüedad semántica.

---

<sup>17</sup> Data.gov, <http://catalog.data.gov/dataset>

<sup>18</sup> datahub.io, <http://datahub.io/dataset>

<sup>19</sup> Open Knowledge Foundation, <https://okfn.org>

## **2.5 Hacia la semantización de datos de transporte**

---

Posteriormente, ante los problemas encontrados para modelar una gran cantidad de información geoespacial y, más concretamente, la dificultad para realizar operaciones geométricas complejas inherentes a los GIS (Sistemas de información geográfica) sobre dicho modelo ontológico, se presentan una serie de estudios enfocados en un dominio más concreto y alineado con el desarrollo presentado en esta tesis doctoral, el de la información semántica de transporte multimodal.

### **2.5.1 Ontologías con información geoespacial**

Durante miles de años la población ha estado recogiendo datos geográficos. Hasta hace unas décadas, éstos consistían sobre todo en mapas en papel. Generalmente, una pequeña leyenda era suficiente para explicar cómo leer el mapa.

Con las TICs y la introducción de los mapas digitalizados, almacenados en un ordenador, se dio un importante paso adelante. De este modo los mapas podían ser replicados y compartidos más fácilmente. Los mapas digitalizados, sin embargo, sólo permiten realizar transformaciones simples, como escalados o rotaciones.

Así, después, se empezó a almacenar no ya la información gráfica de los mapas sino directamente la información contenida en ellos. Por ejemplo, en vez de la representación gráfica de una carretera, se almacenaron las coordenadas de los cruces, junto con algo más de información sobre el tipo de carretera. De esta manera, se podían utilizar los datos, no sólo para la producción cartográfica, sino para la provisión de servicios adicionales.

Es a partir de este punto cuando se detecta el problema fundamental ya mencionado con anterioridad, la necesidad de guardar la información en un formato estándar e integrable. Así, la interoperabilidad se convierte en esencial para los sistemas de información geográfica, cuya información se almacena generalmente en bases de datos geoespaciales heterogéneas

Se hace necesario considerar dicha heterogeneidad y favorecer métodos que posibiliten la interoperabilidad entre herramientas geográficas con el fin de satisfacer la creciente demanda en el uso e intercambio de datos geoespaciales [Boucher 09, Vaccari 09, Zhang 10].

## 2. Estado del arte

---

### Sistemas de información geográfica basados en el uso de ontologías

Ya en 1999, Fonseca [Fonseca 99] presenta un diseño conceptual de GIS basado en el uso de ontologías para la descripción de objetos que representan a su vez entidades geográficas. Una de las principales aportaciones de dicho trabajo es la definición de tres tipos de heterogeneidad afectando a los GIS:

- ◇ *Heterogeneidad semántica*, donde se usan diferentes términos refiriéndose a conceptos similares.
- ◇ *Heterogeneidad esquemática*, donde los mismos términos pueden hacer referencia a conceptos diferentes.
- ◇ *Heterogeneidad sintáctica*, mostrando diferencias en cuanto al modelo de datos en sí mismo.

Dicha heterogeneidad se produce al efectuar los sistemas GIS tradicionales consultas espaciales usando un método basado en palabras clave. Este enfoque genera ambigüedad y es incapaz de expresar de forma completa las necesidades del usuario, debido a la falta de conceptos geográficos (semántica) en el conjunto de datos [Fu 05]. En este contexto, el enfoque más prometedor para lograr terminar con esta ambigüedad es la implementación de semántica mediante el uso de ontologías para conjuntos de datos geográficos.

### Ontología para las redes de transporte

Lorenz realiza un exhaustivo análisis en su trabajo *Ontology of Transportation Networks* [Lorenz 05], en donde expone en primer lugar los esfuerzos que se han venido realizando desde las instituciones, tanto internacionales como europeas, dirigidos hacia la estandarización de la información geográfica.

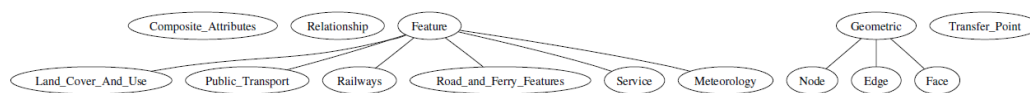
Entre ellos destaca la definición del formato GDF, *Geographic Data File*, desarrollado mediante el proyecto de investigación y desarrollo *European Digital Road Map* [EDRM 94] y siendo el líder en su ámbito desde entonces, con proveedores como NAVTEQ, Bosch, Philips o Volvo usando dichos mapas para sus sistemas de navegación para automóviles.



## 2.5 Hacia la semantización de datos de transporte

En su trabajo toma como referencia dicho modelo de datos para la generación de una ontología para las redes de transporte (OTN, *Ontology of Transportation Networks*). Su objetivo es definir semánticamente (mediante el uso de OWL) los conceptos geográficos establecidos en GDF, favoreciendo la integración de los datos espaciales y la compartición de la información. La figura 2.11 muestra las cinco clases básicas pertenecientes a dicha ontología:

- ◇ *Feature*. Conversión de todos los elementos de GDF a clases en OTN
- ◇ *Geometric*. Definición de la forma geométrica de los elementos
- ◇ *Composite\_Attributes*. Definición de clases compuestas por varios atributos
- ◇ *Relationship*. Descripción de las relaciones no geométricas entre elementos
- ◇ *Transfer\_Point*. Descripción del mecanismo de desplazamiento desde un elemento de la ontología a otro (p. ej. Estación de tren)



**Figura 2.11:** Clases básicas de la ontología OTN

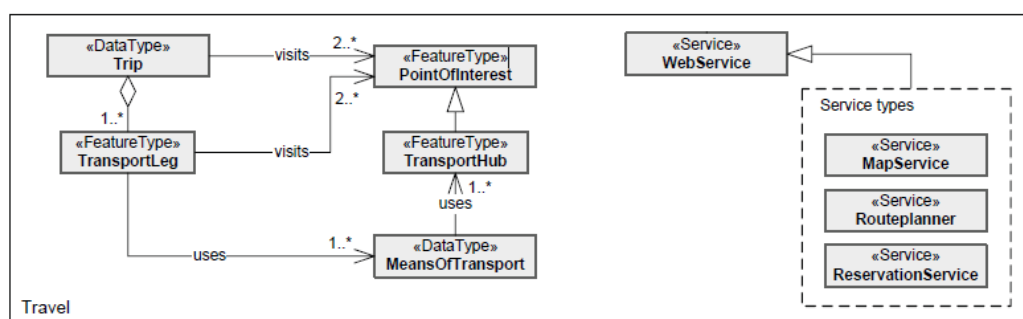
### Interoperabilidad semántica en servicios geospaciales distribuidos

El trabajo de tesis presentado por Lemmens [Lemmens 06] ahonda en la necesidad de buscar la interoperabilidad entre distintos conjuntos de datos y servicios web basados en información geográfica.

En él expone cómo el cambio de paradigma en cuanto a arquitecturas software, de un modelo centralizado a uno distribuido e interconectado, ha influido de modo sustancial en las herramientas y la información geográfica de la que disponemos en la actualidad (mapas interactivos, planificadores de rutas, información estadística, etc.) De este modo surge la necesidad en las organizaciones de integrar y reutilizar la información disponible (tanto interna como externa), apareciendo la problemática asociada a la heterogeneidad de dichos datos.

## 2. Estado del arte

Lemmens se centra en el proceso de integración u orquestación de servicios. Para ello en primer lugar establece la generación de una serie de ontologías, tanto de dominio (la figura 2.12 muestra como ejemplo la ontología diseñada para el dominio del transporte) como operativas, basadas en el estándar ISO GFM (*General Feature Model*). GFM define una serie de metadatos utilizando el esquema UML que permiten identificar conceptualmente elementos geográficos. Es decir, amplía el dominio geoespacial desde elementos geométricos (puntos, líneas, polígonos, etc.) hacia elementos conceptuales (pozo, frontera geológica, medición, etc.)



**Figura 2.12:** Estructura general de la ontología para el dominio del transporte

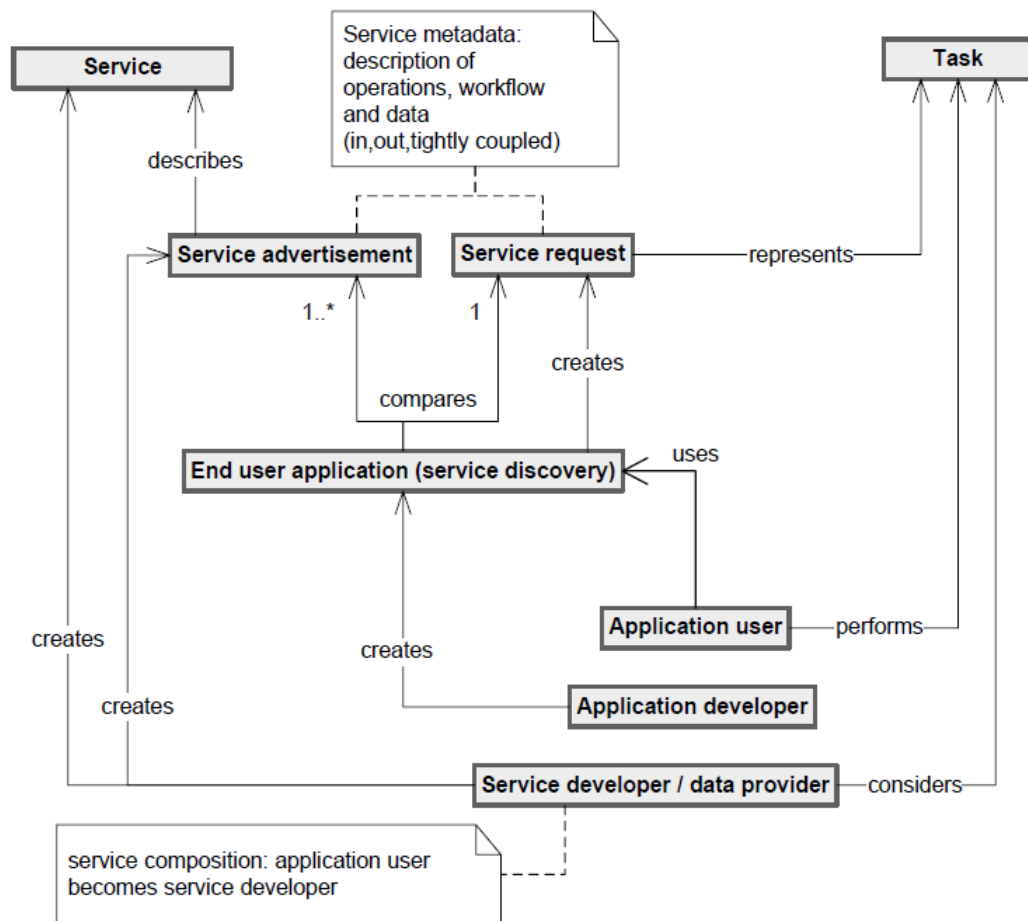
En segundo lugar, genera descriptores de servicios basados en dichas ontologías, de modo que puedan ser descubiertos por aplicaciones externas. Así, define cuatro casos de uso en los que un aplicativo externo puede hacer una petición al sistema y éste le devuelve la información solicitada gestionando la agregación de los datos de forma distribuida. La figura 2.13 presenta un esquema conceptual del sistema de descriptores y servicios, tanto desde el punto de vista del consumidor como desde el punto de vista del desarrollador de servicios y/o gestor de los datos.

Se exponen sin embargo dos limitaciones principales de la arquitectura propuesta: la complejidad en cuanto a la introducción manual de los descriptores para cada servicio a publicar/buscar, y la incompletitud o la rigidez en las asignaciones entre las distintas ontologías de dominio presentes.

### Integración y consulta de datos geoespaciales mediante ontologías

Zhao [Zhao 08] presenta por su parte un artículo donde atiende la necesidad de interoperabilidad en los datos geográficos desde una perspectiva distinta.

## 2.5 Hacia la semantización de datos de transporte



**Figura 2.13:** Contexto de descubrimiento y composición de servicios

Mientras que los autores citados anteriormente focalizan sus esfuerzos en modelar semánticamente la información geográfica, diseñando una o varias ontologías para intentar conseguir una traducción directa desde formatos geoespaciales (como GDF o GFM) hacia un modelo semántico más rico, Zhao busca reutilizar formatos y protocolos geoespaciales estándar ya existentes y potenciarlos mediante el uso de semántica en la fase de consulta.

Así, su investigación se centra en la explotación y mejora mediante técnicas semánticas de formatos y especificaciones definidos por el OGC (*Open Geospatial Consortium*). Consorcio que agrupa más de 400 organizaciones públicas y privadas encargado de definir y fomentar estándares abiertos en el ámbito de los GIS.

## 2. Estado del arte

---

En concreto toma como base el formato GML (*Geographic Markup Language*) que deriva del XML, y es el encargado de codificar la geometría de objetos en base a puntos, líneas o polígonos, y el protocolo WFS (*Web Feature Service*), interfaz de comunicación que permite la consulta de información geográfica (en formato GML) vía web. Ambos formatos han sido ampliamente adoptados por la comunidad GIS debido a la mejora evidente que suponen en cuanto a la compartición y la sincronización de información geoespacial. Sin embargo, siguen sin resolver los problemas de heterogeneidad semántica ya explicitados.

Class	Superclass
<i>Route</i>	<i>Feature</i>
<i>Pattern</i>	<i>Feature</i>
<i>LinkSequence</i>	<i>Feature</i>
<i>Stop</i>	<i>SpatialFeature</i>
<i>Link</i>	<i>SpatialFeature</i>

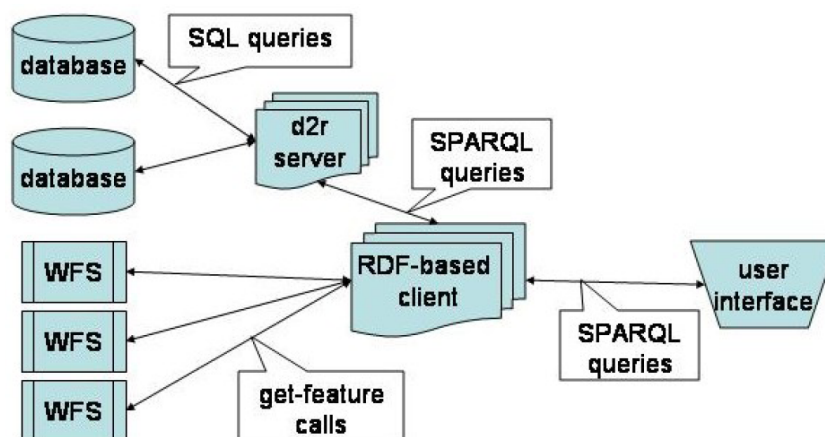
Property	Domain	Range
name	<i>Route</i>	<i>string</i>
intersection	<i>Stop</i>	<i>string</i>
routeID	<i>Pattern</i>	<i>integer</i>
patternID	<i>LinkSequence</i>	<i>integer</i>
linkID	<i>LinkSequence</i>	<i>integer</i>
linkOrder	<i>LinkSequence</i>	<i>integer</i>
route_link	<i>Route</i>	<i>Link</i>
route_stop	<i>Route</i>	<i>Stop</i>
nearby_stop	<i>Stop</i>	<i>Stop</i>

**Figura 2.14:** Ontología puente entre WFS y las consultas SPARQL

La transformación directa de los datos geoespaciales presentes (almacenados en bases de datos geográficas) hacia un modelo ontológico tampoco resulta una alternativa viable. Por un lado el proceso sería prolongado, propenso a la aparición de errores e ineficiente. Por otro, las herramientas existentes para la gestión de ontologías no serían capaces de soportar una carga tan alta de instancias debido principalmente al consumo de memoria exigido. Tampoco el rendimiento sería aceptable si consideramos la inclusión de funcionalidades provistas por las bases de datos geográficas o los servicios WFS como la gestión de transacciones o las consultas espaciales incluyendo operaciones geométricas complejas.

Teniendo esto en cuenta, se propone la inclusión de semántica en una capa superior, mediante un adaptador o interfaz implementado como una ontología RDF (ver figura 2.14) y situado entre las consultas web realizadas por los usuarios y los formatos geoespaciales subyacentes. La arquitectura propuesta (ver figura 2.15) se estructura en tres fases principales:

## 2.5 Hacia la semantización de datos de transporte



**Figura 2.15:** Arquitectura de la interfaz ontológica RDF

1. La creación de una ontología que describa el dominio concreto a soportar por la interfaz (en la figura 2.15)
2. La definición de las relaciones y las reglas de inferencia necesarias para conectar la ontología con los tipos de datos de WFS y los esquemas de BBDD.
3. El algoritmo de reescritura o transformación de las consultas realizadas a la ontología (*SPARQL queries*) hacia el servicio WFS (*get-feature calls*) o la BBDD (*SQL queries*) de cara a obtener los resultados.

### 2.5.2 Modelización semántica de datos de transporte

Hasta este momento se ha venido reflejando cómo varios autores han intentado gestionar mediante ontologías la información geográfica presente en sistemas GIS. Sin embargo, se ha podido comprobar cómo la complejidad intrínseca de este tipo de información [Frank 01] así como las necesidades de capacidad de cómputo [Zhao 08] hacen inviable la aplicación a gran escala de soluciones de este tipo.

Por otra parte, estudios relevantes como el realizado por Agarwal [Agarwal 05], discuten también la viabilidad de un enfoque ontológico unificado, teniendo en cuenta el carácter interdisciplinar de la investigación en información geográfica y las diferentes conceptualizaciones y terminología utilizada para los mismos conceptos geoespaciales.

## 2. Estado del arte

---

El enfoque actual dirige así la investigación hacia dominios o áreas de actuación más concretas dentro de la información de transporte. En ese sentido, se pretende encontrar un modelo ontológico que formalice de algún modo el transporte multimodal [Zhang 11] y permita atender las consultas para los diferentes modos de transporte público y privado en redes de transporte a gran escala.

A continuación se describen una serie de trabajos relevantes bajo ese mismo modelo de actuación. Se especifican también las carencias o los elementos de mejora encontrados por los propios autores ya que se intentarán resolver de forma efectiva mediante la arquitectura propuesta en esta tesis.

### **Ontología para la personalización multi-criterio de rutas de transporte**

Niaraki [Niaraki 09] combina en su trabajo la aplicación de un Proceso Analítico Jerárquico (AHP), una técnica estructurada para tratar con decisiones complejas, junto con la modelización ontológica de la información de transporte (en este caso privado, el estudio no contempla el uso de transporte público) para personalizar rutas en base a diversos criterios cualitativos externos.

Se define para ello una ontología específica para el dominio de la planificación de rutas, combinando información geográfica junto con información turística o contextual. El algoritmo propuesto para la selección de la ruta óptima establece una serie de criterios cualitativos que el autor decide introducir como elementos dentro de la ontología. Esto permite al sistema extender y agregar información desde fuentes externas, favoreciendo la correcta toma de decisiones y mejorando la precisión del algoritmo de búsqueda.

Niaraki destaca la efectiva aplicación de un modelo ontológico en un área, la planificación de rutas personalizadas, donde no había aún soluciones de este tipo. Se comprueban las ventajas aportadas por la modelización semántica en cuanto a reusabilidad y compartición de la información, aunque también se apunta a la necesaria revisión iterativa de la ontología propuesta y la evaluación de esta en escenarios reales.

### Ontología de transporte público para la planificación de rutas

Houda [Houda 10] presenta en su trabajo una ontología de dominio para la gestión de rutas multimodales en la que incorpora elementos turísticos que pueden ser de relevancia para el usuario a la hora de preparar su viaje. Tiene en cuenta para ello la aparición de servicios como restaurantes, bibliotecas, centros comerciales, etc. en las proximidades de la ruta. Se desmarca así de la modelización completa de los datos geoespaciales, generando desde cero una ontología para el transporte público en la que define tanto los conceptos implicados en la realización de las rutas (medios de transporte, agencias, paradas, etc.) como el mecanismo, basado en reglas, necesario para la selección y clasificación de estas. La figura 2.16 muestra diagramas UML de ambas partes.

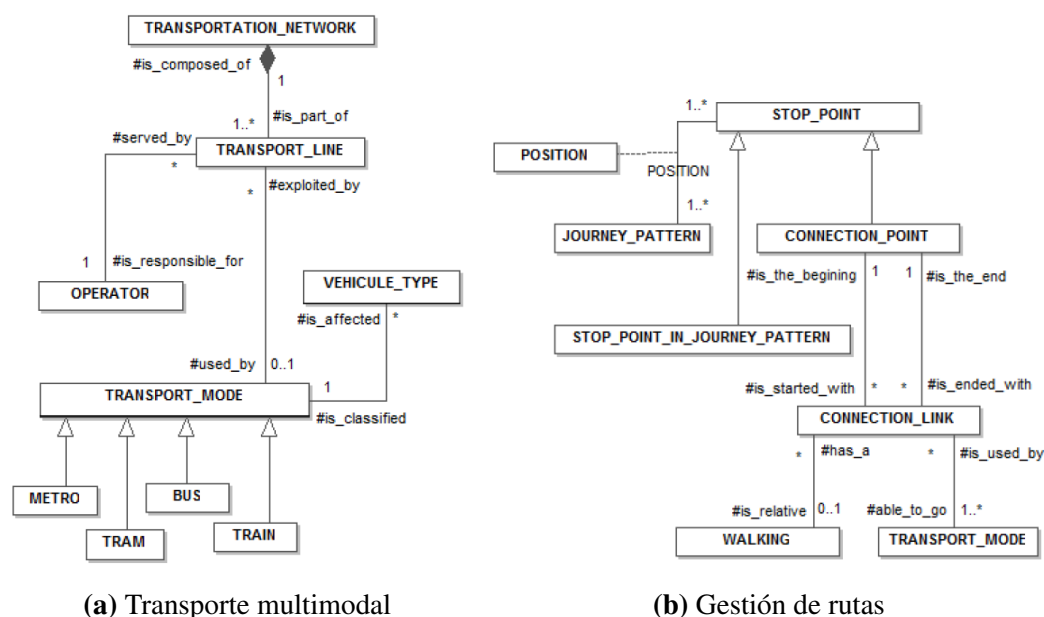
$$\begin{aligned} & CONNECTION\_POINT(?z) \wedge LIBRARY(?f) \wedge is\_encircled\_by(?z, ?f) \wedge \\ & POSITION\_RELATIONSHIP(?b) \wedge is\_designated(?z, ?b) \wedge \\ & JOURNEY\_PATTERN(?v) \wedge Journey\_Pattern\_Position(?b, ?v) \rightarrow \\ & LEISURE\_JOURNEY\_PATTERN(?v) \end{aligned} \quad (2.1)$$

Por otra parte, la expresión 2.1 ejemplifica mediante lenguaje SWRL (*Semantic Web Rule Language*) la clasificación de una entidad ruta (*journey pattern*) como ruta de ocio (*leisure journey pattern*) dado que mantiene una conexión (*connection point*) con una entidad biblioteca (*library*).

El mayor problema de la ontología propuesta es la necesidad de realizar una carga manual tanto de los elementos turísticos a considerar como de las rutas de transporte provistas por las diferentes agencias implicadas. Esta información que será consumida por la ontología deberá ser instanciada previamente de manera manual. También el conjunto de reglas que permiten al sistema realizar la clasificación y la selección de las distintas rutas debe ser especificado a priori.

Sin embargo, el trabajo propuesto, al igual que el de Niaraki, es un ejemplo de cómo ontologías específicas utilizadas para modelar información de un dominio concreto pueden proporcionar características deseables como la formalización de la información o la interoperabilidad suponiendo un avance sobre los formatos y tecnologías recogidos en la sección 2.3.2 de este estado del arte.

## 2. Estado del arte



**Figura 2.16:** Diagramas UML de la ontología

### Ontología de transporte multimodal para la personalización de interfaces

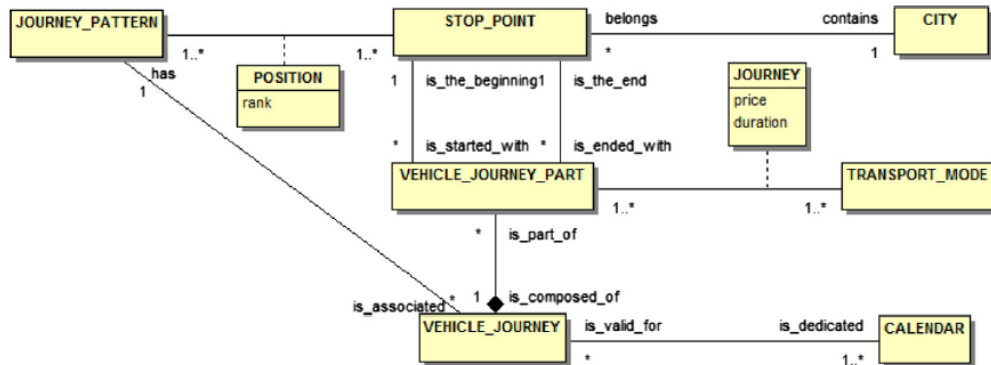
Oliveira presenta en [De Oliveira 13] un trabajo en el que centra la utilización de ontologías en el modelado de información para el transporte multimodal con el objetivo de generar interfaces personalizadas según las características del usuario para el desarrollo de sistemas interactivos de transporte.

Dicha interactividad viene dada por la personalización de la interfaz de usuario que se adaptará al dispositivo (tablet, pc de escritorio o teléfono móvil) y a la necesidad concreta del usuario en cada momento de acuerdo a su contexto. El contexto, compuesto por información de contacto, plataforma utilizada e información del entorno (localización, horario, temperatura, etc.) será mapeado mediante un modelo ontológico de contexto.

Como conclusión, el autor destaca la utilidad de las ontologías de dominio desarrolladas tanto en el proceso de diseño como en el de ejecución de las consultas, facilitando la identificación de la información y la provisión de resultados. Sin embargo también apunta como punto de mejora la preparación de herramientas de soporte que faciliten la introducción de datos. La figura 2.17 muestra el conjunto de propiedades definidas por Oliveira para modelar una ruta de transporte.



## 2.5 Hacia la semantización de datos de transporte



**Figura 2.17:** Propiedades definidas para modelar una ruta de transporte

### Planificador de rutas de transporte público semántico

El trabajo realizado por Gunay en [Gunay 14] presenta una solución más general, consiguiendo modelar semánticamente la información de transporte público en base a la ontología PDTO (*Public Transport Domain Ontology*) incluyendo conceptos, relaciones e individuos y proveyendo de un sistema de búsqueda geoespacial.

Para la generación de la ontología, y a diferencia de otros autores que definen su propio modelo, Gunay decide basarse en el conjunto de directivas INSPIRE<sup>20</sup> (*Infrastructure for Spatial Information in the European Community*) que proponen un modelo de datos común con el que se posibilite la interoperabilidad de datos geográficos entre los miembros de la UE.

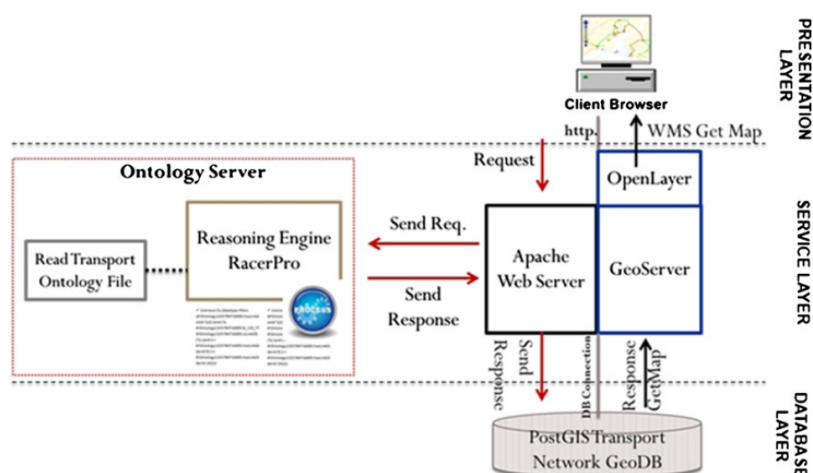
El objetivo del trabajo era investigar cómo el uso de información semántica puede potenciar el enfoque GIS tradicional para lo que se desarrolló un portal para la planificación de rutas de transporte público basado en el modelo semántico diseñado. Dicho portal sigue una arquitectura clásica cliente-servidor de 3 capas (ver figura 2.18) cuyas funciones son las siguientes:

- ◇ *Capa de datos:* Base de datos geográfica (*postgreSQL* y *PostGIS*) con los datos geoespaciales disponibles.
- ◇ *Capa de aplicación:* Servidor web y razonador semántico (*RacerPro Query Language*) encargados de gestionar las peticiones y seleccionar las rutas.
- ◇ *Capa de presentación:* *OpenLayers JavaScript API*, para dibujar el mapa y representar las rutas. Se utiliza *Java* para el desarrollo software.

<sup>20</sup> Directiva INSPIRE, <http://inspire.ec.europa.eu>

## 2. Estado del arte

---



**Figura 2.18:** Arquitectura del portal semántico planteado por A. Gunay

La solución desarrollada se muestra operativa, atendiendo correctamente a las peticiones de los usuarios y gestionando los datos de forma homogénea (resolviendo problemáticas inherentes a las soluciones GIS tradicionales como la sinonimia de conceptos, p. ej. una parada intermodal con distintas referencias).

El autor hace referencia sin embargo a una serie de problemas que imposibilitarían la puesta en marcha a gran escala de la solución propuesta. Por un lado, refiere una carga semiautomática de los datos de transporte ya existentes (presentes en la base de datos geoespacial) hacia la ontología para la generación de individuos (instancias concretas), que después serán utilizados para la obtención de las rutas. El proceso no está lo suficientemente automatizado, llevando mucho tiempo y esfuerzo la carga de datos en la ontología.

Por otra parte, indica el alto coste computacional de la solución y la limitación de medios de transporte incorporados a ésta (autobús, taxi y ferry), recomendando la utilización de técnicas distribuidas como posible solución a dichos problemas.

*No evitéis las dificultades a vuestros hijos, mas bien enseñadles a superarlas.*

Louis Pasteur

# 3

## Oportunidades de mejora y contribuciones

**D**urante el análisis previo del estado del arte, se han venido apuntando los avances realizados en los últimos años en la creación de servicios avanzados de información al viajero; presentando soluciones comerciales como *Google Maps* o soluciones de código abierto desarrolladas por la comunidad *Open Data* que han supuesto un importante avance a la hora de facilitar la gestión y la planificación del transporte. También se han plasmado otros enfoques y trabajos de investigación, derivados de las necesidades de interoperabilidad y apertura de la información, más orientados al análisis del modelo de datos subyacente y la habilitación de servicios.

Algunos de estos trabajos han servido como base a la hora de diseñar diferentes funcionalidades o mecanismos incluidos en la arquitectura presentada en esta tesis doctoral. Otras soluciones descritas, en cambio, han contribuido al desarrollo de la tesis porque presentan limitaciones o exponen necesidades que motivan o justifican en cierta manera su desarrollo e interés.

Se presentan en este capítulo las capacidades, limitaciones y oportunidades de mejora detectadas en el análisis efectuado del estado del arte, haciendo explícitas de este modo las aportaciones realizadas mediante la arquitectura propuesta.

### 3. Oportunidades de mejora y contribuciones

---

A lo largo del segundo capítulo se han venido identificando una serie de características deseables en cuanto a la representación y la provisión de información de transporte. Muchas de estas características se corresponden con limitaciones o puntos de mejora de los trabajos analizados y están directamente relacionadas con el objetivo de esta tesis. Recordamos a continuación el enunciado de dicho objetivo:

*Diseñar y desarrollar una arquitectura software distribuida que permita, gracias a la estructuración de datos de transporte provenientes de fuentes heterogéneas junto con la integración de información relevante relacionada, ofrecer una plataforma interoperable para la habilitación de servicios software avanzados relacionados con la movilidad multimodal.*

Se presentan a continuación dichas características, derivadas del análisis realizado sobre los distintos enfoques para la gestión de la información de transporte, y clasificadas de acuerdo a áreas temáticas concretas. En primer lugar se exponen las capacidades que un modelo de datos enfocado a la provisión de información de tránsito debiera poseer, comparándose con los formatos analizados previamente. En segundo lugar, se exponen las oportunidades en cuanto al enriquecimiento de la información provista, mediante la integración, la compartición y la publicación de la misma como datos enlazados abiertos. Por último, se desarrolla el planteamiento de una arquitectura distribuida que soporte la explotación y la provisión de información de transporte heterogénea en un escenario de actuación real.

#### 3.1 Modelización de los datos

Tal y como se detalla en el apartado 2.3.2, diversos sistemas coexisten a la hora de representar la información de tránsito. Sin embargo, las soluciones existentes no consiguen satisfacer completamente el conjunto de objetivos fundamentales planteados en este trabajo, presentando una serie de limitaciones.

Se considera que las soluciones existentes no son lo suficientemente interoperables ni fácilmente consultables o reutilizables, lo que se traduce en información inconexa, desactualizada y aislada. Por otro lado, el acceso a dicha información es también en la mayor parte de los casos un problema, ya que o bien no está disponible en un formato que permita su consulta efectiva, o directamente no es pública, haciendo muy costoso el acceso a los datos.

#### 3.1.1 Caracterización del modelo de datos

La tabla 3.1, es una extensión de la comparativa presentada en el capítulo 2, exponiendo un conjunto de características útiles para la correcta integración y compartición de la información. Se agrega en este caso la columna MTO (*Multimodal Transport Ontology*), siendo dicha ontología para el transporte multimodal (cuyo diseño se estudiará con detalle durante el capítulo 4) una de las aportaciones destacadas de esta tesis.

	DATEX II	GTFS	WFS	MTO
<b>Clasificación</b>	<i>Open Data</i>	<i>Open Data</i>	<i>Open Data</i>	<i>Open Data</i>
<b>Estructura</b>	XML	CSV	GML (XML)	Ontología formal
<b>Extensibilidad</b>	Sí	No	No	Sí
<b>Enlazabilidad</b>	No	No	No	Sí
<b>Consultabilidad</b>	Servicio Web	Programática	Servicio Web	Directa (SPARQL)
<b>Acceso a datos</b>	Limitado	Completo	Limitado	Completo

**Tabla 3.1:** Formatos para la representación de la información de transporte (b)

A continuación se muestra de forma teórica cómo mediante el enfoque ontológico propuesto se pretenden resolver los problemas encontrados en la modelización de la información de transporte, dando soporte efectivo a cada una de las características y capacidades expuestas.

- ◇ *Clasificación y acceso a datos.* Se propone un uso libre, abierto y completamente redistribuible de la información provista, permitiendo un acceso completo a los datos. El uso de una arquitectura basada en protocolos y servicios web estándar (como HTTP o REST) habilitará el acceso universal y multi-plataforma a la información disponible.
- ◇ *Estructura.* Se considera necesario establecer un modelo estructurado que facilite la compartición y la reutilización del conocimiento. El uso de ontologías formales, tal y como se ha relatado en la sección 2.4 del estado del arte,

### 3. Oportunidades de mejora y contribuciones

---

ha supuesto un importante avance en este aspecto. Es por ello que la arquitectura propuesta hará uso de dicho sistema como mecanismo para modelar la información.

- ◇ *Extensibilidad y enlazabilidad.* El uso de un modelo ontológico para la representación de la información permitirá a la arquitectura propuesta aprovecharse de las capacidades de agregación, integración y consulta aportadas por la pila de protocolos de la web semántica. Así, la información provista como datos enlazados abiertos (LOD) tendrá la capacidad de ser extendida mediante su conexión con otros conjuntos de datos relevantes.
- ◇ *Consultabilidad.* Al igual que para el punto anterior, el uso de protocolos de la web semántica favorece la consulta de los datos almacenados. Se utilizará *SPARQL*, un lenguaje de consulta sobre recursos ontológicos (similar a SQL en el caso de almacenamiento en bases de datos relacionales), como mecanismo para la consulta de la información.

#### 3.1.2 Diseño del modelo ontológico

Una de las problemáticas detectadas en el análisis previo de trabajos en los que se utilizan ontologías para modelar información de diversos dominios, se corresponde con la generalización y la formalización de dicha ontología, atendiendo a la especificidad de los conceptos y la reutilización de vocabularios existentes.

Estudios como el de Agarwal en [Agarwal 05] discuten la viabilidad de un enfoque ontológico unificado, sobretodo para dominios muy amplios, debido a las diferentes conceptualizaciones y terminología utilizada para los mismos conceptos, lo que puede derivar en inconsistencias y heterogeneidad semántica.

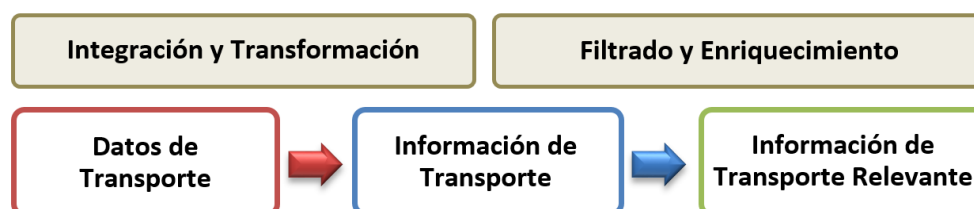
En el caso de MTO, y de cara a solventar dichas dificultades, se tendrá en cuenta el conjunto de buenas prácticas definidas por [Berners-Lee 09a] a la hora de especificar el modelado del dominio. Dado que GTFS es actualmente el estándar de facto para la representación de datos de tránsito se ha optado por realizar una reformulación de sus ficheros CSV a entidades dentro de la ontología. Dicha característica supone la adopción de un modelo de datos consolidado, otorgando a la ontología resultante de una ventaja competitiva importante sobre modelos creados al efecto y sin un proceso de validación previa.

El enfoque a seguir para la construcción de la ontología MTO persigue por tanto la definición de un alcance muy concreto, adscrito a una representación completa pero no superflua de los conceptos y atributos del transporte multimodal.

### 3.2 Interoperabilidad y enriquecimiento

Otra de las carencias encontradas en las arquitecturas para la provisión de información de transporte es que no tienen en cuenta factores relevantes como el contexto o la agregación de información relacionada a la hora de proveer los resultados. Tomando como ejemplo fundamental los sistemas para la planificación de rutas, dichas limitaciones afectan a la utilidad final de la información provista, ofreciendo resultados en base a parámetros como la distancia o la duración del viaje, no siendo posible combinar este tipo de consultas con otros parámetros relacionados, lo que dotaría de mayor relevancia y precisión a dicha información.

Se ha realizado un trabajo de análisis en cuanto a las posibilidades de enriquecimiento y mejora de la información disponible. La figura 3.1 muestra la evolución deseable de dicha información, transitando por distintas fases y procesos, donde se ve enriquecida y potenciada buscando dotarla de relevancia para el usuario.



**Figura 3.1:** Evolución de la información de transporte

A continuación se detallan dichas fases, mostrando cómo colabora cada una de ellas en la provisión de información más relevante:

◇ *Fase de integración y transformación*

En esta fase se realiza el proceso de agregación e integración de los datos ya existentes. Partirá de los datos de transporte disponibles y los combinará para generar nueva información enlazada e integrada de acuerdo a un modelo formal estructurado. Los datos de transporte de partida podrán proveer de distintas fuentes de información o encontrarse en distintos formatos.

### 3. Oportunidades de mejora y contribuciones

---

#### ◇ *Fase de filtrado y enriquecimiento*

En esta fase se cumple con el objetivo de dotar de valor añadido a la información de transporte resultante de la fase anterior mediante el uso del componente contextual. Dicha característica permitirá a los usuarios filtrar la información de acuerdo a sus intereses o necesidades particulares, personalizando de este modo la información y dotando de relevancia a la misma.

La ejecución de dichas fases se basa en el uso de un modelo semántico como plataforma para la estructuración y la publicación de la información de transporte. Como se ha comentado anteriormente, el uso de modelos formales, como el planteado con la ontología MTO, aporta una serie de características deseables como la formalización y la interoperabilidad de la información, en este caso en el ámbito del transporte multimodal. Del mismo modo, el uso de la web semántica proporciona una infraestructura estable sobre la que construir el mecanismo de gestión y provisión de dichos datos, haciendo un uso extensivo del conjunto de protocolos provisto (desde HTTP a OWL o SPARQL) y basándose en la metodología *Linked Open Data* para la publicación de información integrada y enlazada con otros conjuntos de datos.

Se ha realizado un estudio de los vocabularios existentes y que se podrían reutilizar para el diseño de la ontología. A este respecto se ha previsto la integración de vocabularios ampliamente soportados por la comunidad *Linked Data* en el ámbito de la información geográfica, enlazando y/o extendiendo dichos vocabularios y favoreciendo de este modo la alineación de la ontología con el paradigma LOD.

### 3.3 Explotación y publicación de la información

De acuerdo a lo expuesto anteriormente, se ha optado por un enfoque ontológico basado en el uso de LOD para la modelización y la publicación de información semántica de transporte multimodal. Dicho enfoque facilitará la evolución necesaria de los datos en bruto a la información relevante, y ya ha sido utilizado en este y otros ámbitos de actuación (ver sección 2.5).

Precisamente, de dichos trabajos relacionados, en los que se utilizan ontologías para modelar información de diversos dominios, se extraen una serie de problemáticas comunes derivadas de la puesta en marcha de la solución final.



#### 3.3.1 Datos reales de transporte multimodal

Otra de las problemáticas analizadas deriva de la necesidad de inicializar el modelo ontológico generado con instancias válidas y, preferiblemente, reales, que representen de este modo fielmente el dominio que se quiere modelar [Houda 10, De Oliveira 13]. Esto permitirá la validación efectiva de dicho modelo ontológico. Este proceso normalmente no se encuentra automatizado con lo que la carga de datos en la ontología puede llevar mucho tiempo y esfuerzo.

De acuerdo a este aspecto, y teniendo en cuenta que ya existe una gran cantidad de información de transporte multimodal en formato GTFS, se ha planteado el desarrollo de un adaptador para la conversión automatizada de dicho formato al modelo de datos MTO. Mediante dicho proceso se propicia de igual modo la transformación de recursos no ontológicos (NOR), como GTFS, a un modelo integrable y enlazado, concepto primordial en el ámbito de la publicación de LOD.

El uso de dicho adaptador favorecerá también la instanciación transparente de datos reales de transporte multimodal ya existentes, pasando a formar parte de la arquitectura propuesta. Así, se prevé la carga de datos de transporte publicados por el Gobierno Vasco, utilizados en herramientas previamente analizadas como *Moveuskadi* (ver sección 2.2.3) y la generación del contenido semántico correspondiente.

Esto permitirá realizar un despliegue de la solución propuesta bajo un escenario real de uso del transporte público y multimodal en un área lo suficientemente amplia como para poder validar la arquitectura de forma eficaz.

#### 3.3.2 Arquitectura distribuida

La manera en la que se representa y se estructura la información de transporte, en este caso mediante la ontología MTO, supone parte de la innovación presentada en este trabajo. Sin embargo, dicho modelo de datos debe tener un soporte que habilite su consulta, definiendo a su vez opciones de búsqueda y filtrado de la información.

Del mismo modo, la arquitectura propuesta persigue solventar algunas de las limitaciones encontradas en los sistemas existentes. El despliegue en un escenario real conllevará la gestión de un gran volumen de información ontológica. Dicha característica ha sido identificada ya por varios investigadores [Zhao 08, Gunay 14] como un problema de los sistemas de modelización semántica, derivando en incompletitud o falta de información por limitaciones de cómputo.

### 3. Oportunidades de mejora y contribuciones

---

Esta problemática será abordada por la arquitectura propuesta mediante la distribución de la información semántica en servidores SPARQL organizados geográficamente. Cada servidor mantendrá información de transporte propia y será administrado localmente, pero facilitará la interoperabilidad gracias a su interconexión con el resto de servidores de la red mediante URIs dinámicas definidas como metadatos dentro de la propia ontología, planteando de este modo un mecanismo para la realización de consultas distribuidas y transparentes para el usuario.

Es decir, cuando un usuario realice una petición acerca de una ruta concreta intermodal a uno de los servidores, éste facilitará la misma consulta a los servidores que tiene registrados (realizando una búsqueda federada hacia cada uno de ellos) y así sucesivamente hasta que la consulta sea atendida, favoreciendo así el acceso a información actual e integrada.

#### 3.4 Análisis y diseño conceptual

Durante este capítulo se han recogido las capacidades, limitaciones y oportunidades de mejora detectadas en el análisis efectuado del estado del arte. También se han introducido de manera teórica las contribuciones planteadas en el presente trabajo de tesis con el objetivo de resolver o mitigar dichas carencias.

Así, se ha realizado una caracterización del modelo de datos para la representación de la información de transporte multimodal, se ha definido la ontología MTO como modelo ontológico soporte de dicha representación y se ha previsto la reutilización de vocabularios relevantes con el fin de enriquecer la información provista. Por último, se ha planteado la utilización de una arquitectura distribuida, enfocada a la publicación de la información como LOD y garante del correcto funcionamiento del sistema bajo condiciones de gran volumen de datos como las producidas en escenarios de actuación reales.

Se presenta a continuación la tabla 3.2 que pretende sintetizar dicho análisis. Así mismo, se muestra la figura 3.2 a modo de punto de partida y esquema conceptual de la solución propuesta, cuya descripción completa se llevará a cabo en los siguientes capítulos.

### 3.4 Análisis y diseño conceptual

Limitaciones	Contribuciones
<b>Modelización de los datos</b>	
Acceso y consulta de la información	Modelo ontológico estructurado Información abierta, extensible y enlazable Acceso completo mediante consultas SPARQL
Especificación del dominio y del modelo	Ontología de dominio para el transporte multimodal Modelo de datos consolidado basado en la reformulación del formato GTFS
<b>Interoperabilidad y enriquecimiento</b>	
Agregación e integración de la información	Personalización y filtrado de la información de acuerdo con el contexto del usuario Web Semántica como infraestructura para la integración y la provisión de la información
Reutilización de información existente	Metodología <i>Linked Open Data</i> para la generación de información enlazada Integración de vocabularios en el ámbito de la información geográfica y geopolítica
<b>Explotación y publicación de la información</b>	
Carga de instancias y publicación de los datos	Adaptador para la conversión automatizada desde el formato GTFS a instancias MTO Arquitectura completamente distribuida para la provisión de información semántica
Rendimiento ante un gran volumen de datos	Despliegue de varios servidores distribuidos jerárquicamente e interoperables Mecanismo para la realización de consultas distribuidas transparentes para el usuario
Validación del sistema	Instanciación de datos reales de transporte multimodal en Euskadi Despliegue de la solución bajo un escenario real de uso del transporte multimodal

**Tabla 3.2:** Análisis de limitaciones y contribuciones planteadas

3. Oportunidades de mejora y contribuciones

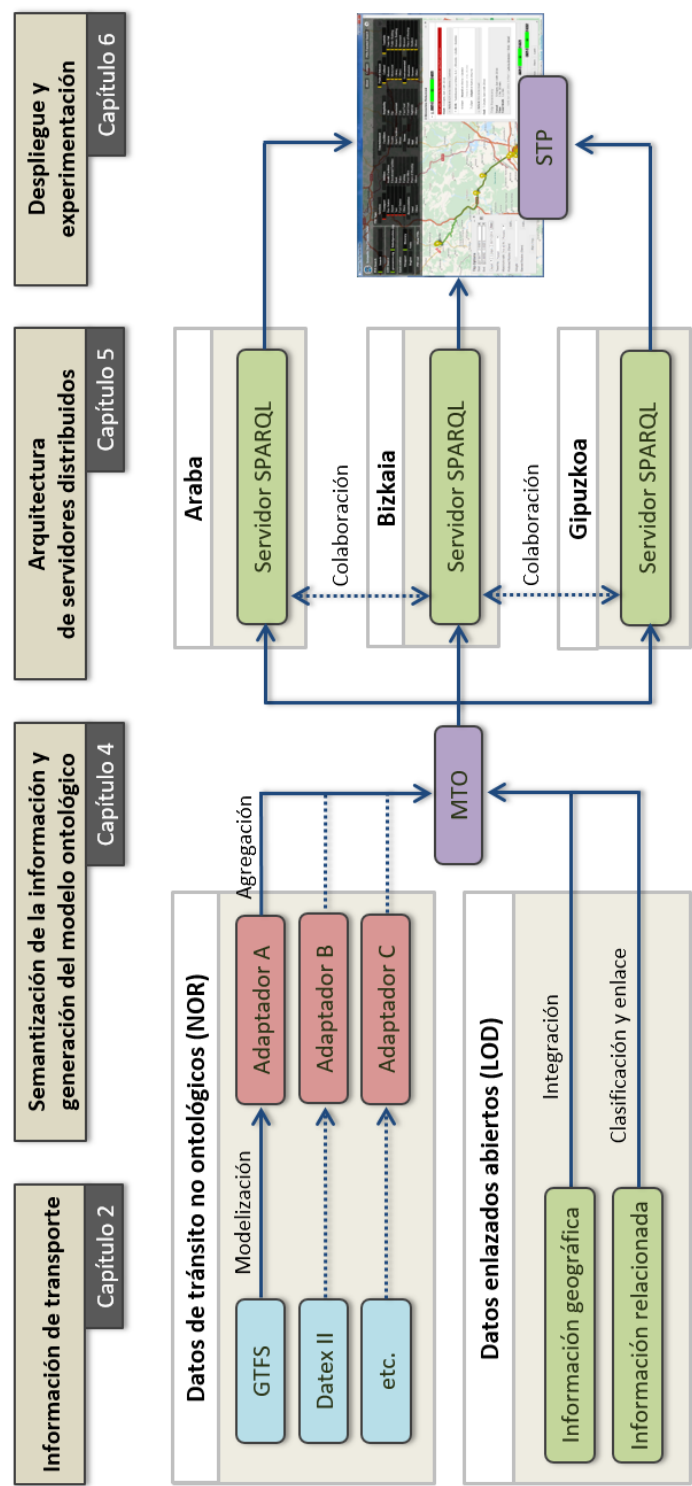


Figura 3.2: Esquema conceptual de la solución propuesta

*El asunto es el problema; la forma,  
la solución.*

Friedrich Hebbel

# 4

## Ontología para el transporte multimodal

**D**urante el capítulo precedente se ha realizado una contextualización en el ámbito de los sistemas de gestión y provisión de información de transporte, exponiendo las limitaciones, capacidades y oportunidades de mejora detectadas e introduciendo de forma teórica las contribuciones planteadas en el presente trabajo de tesis. En vista de dichas limitaciones, relacionadas fundamentalmente con la definición de un lenguaje formal para la integración y la provisión de información de transporte, y dado que son resueltas satisfactoriamente mediante el uso de nuevos enfoques como el modelado ontológico o paradigmas como *Linked Open Data*, se decide primeramente la creación de una ontología de dominio como formato para la gestión y la futura provisión de información semántica de transporte.

Con un especial énfasis en el diseño del modelo de datos, se pretende integrar y estructurar la información disponible, obteniendo de este modo un modelo semántico para la representación de la información. Dicho modelo tendrá capacidad para extender y enlazar información relacionada relevante y estará soportado por herramientas que permitan la consulta efectiva de sus datos, atendiendo al contexto de interés y habilitando la creación de servicios avanzados de información.

## 4. Ontología para el transporte multimodal

---

Se ha definido la ontología MTO (*Multimodal Transport Ontology*) como modelo ontológico soporte de dicha representación y se ha previsto la reutilización de vocabularios de interés dentro del dominio de transporte con el fin de enriquecer la información. El presente capítulo cumple con el objetivo de explicitar las decisiones de diseño y la especificación final de dicha ontología, elemento básico en el proceso de generación y enriquecimiento de la información provista.

### 4.1 Metodología de diseño

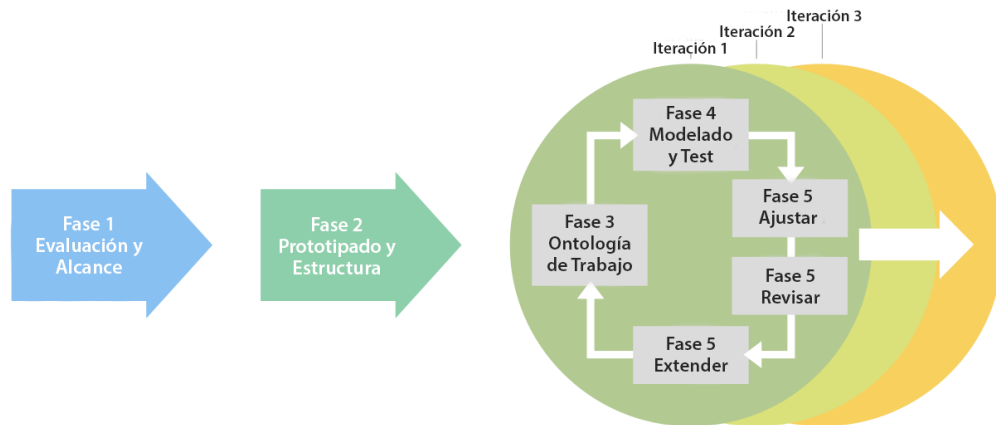
La representación formal del conocimiento se basa en el uso de conceptualizaciones de objetos, atributos y otras entidades pertenecientes a un área de interés y las relaciones que mantienen entre ellas. Una conceptualización es una visión abstracta y simplificada del mundo que queremos representar por algún propósito. Una ontología es una especificación explícita de esta conceptualización. Es por ello que en el diseño de ontologías se debiera seguir un proceso para guiar este modelo.

Varios estudios se pueden encontrar a este respecto, algunos de ellos descritos en el apartado 2.4.1.1. No hay un consenso en cuanto a la utilización de una metodología común para el desarrollo de ontologías; sin embargo, teniendo en cuenta las características del dominio a modelar o la problemática que se quiere abordar mediante el diseño ontológico, se puede decidir el uso de una o varias metodologías acordes al escenario elegido y los resultados previstos. Si bien no existe unanimidad en la comunidad científica en cuanto a qué metodología usar, sí que existe un cierto consenso en cuanto a los objetivos específicos de diseño. Este consenso general, indica que las ontologías deben ser:

- |                              |                  |
|------------------------------|------------------|
| ◇ Orientadas a dominio       | ◇ Colaborativas  |
| ◇ Ligeras (en tamaño)        | ◇ Incrementales  |
| ◇ Coherentes y reutilizables | ◇ Interoperables |

Así, para MTO se ha optado por seguir un proceso iterativo como el propuesto por Noy [Noy 01] y recogido por Mike Bergman en la figura 4.1, que establece la definición del dominio y el alcance, y la reutilización de vocabularios existentes teniendo en cuenta para ello el conjunto de criterios identificados por Gruber [Gruber 95] en cuanto a la claridad, la coherencia o la extensibilidad.

Un enfoque ontológico puede cumplir de este modo con los requisitos de compartición e integración de información de forma muy satisfactoria, en contraste con otras alternativas que caen fácilmente en la ambigüedad y por tanto dificultan la interoperabilidad.



**Figura 4.1:** Enfoque incremental por fases para el desarrollo de ontologías

El diseño de MTO también se ha basado en algunos de los conceptos clave de la metodología NeOn [Suárez 09], un conjunto de nueve escenarios para la construcción de ontologías. Principalmente, debido al hecho de que se centra en la reutilización y transformación de recursos no ontológicos (NOR, *Non Ontological Resources*) como los archivos CSV que componen GTFS. El desarrollo seguido se adscribe principalmente al segundo escenario de la metodología, utilizando adaptadores que permiten la generación de información semántica desde fuentes de datos disponibles actualmente en Internet.

## 4.2 Formalización de la ontología

El objetivo perseguido con la definición de MTO, mediante la adopción de la citada metodología, es integrar los datos de tránsito no estructurados disponibles, enlazando vocabularios existentes que enriquezcan la información provista y generando un modelo semántico para la representación de la información de transporte. Según lo especificado anteriormente, el desarrollo ha estado soportado en ficheros OWL2 [W3C 09], una extensión semántica sobre RDF (estándar XML para la descripción de recursos) generada en 2009 por el W3C.

## 4. Ontología para el transporte multimodal

---

Para facilitar la gestión de la ontología, esta se publica como dos archivos OWL separados. Ambos archivos son necesarios para el funcionamiento de la ontología, por lo que su distribución como paquete unitario es obligatoria:

- ◇ *mto-core.owl* contine el modelo ontológico base; estableciendo las entidades, propiedades y restricciones de la ontología pero sin enlaces a otros vocabularios, individuos o instancias concretas. Importa el archivo *mto-top.owl*.

```
1 <owl:Ontology rdf:about="http://www.mto.org/mto/">
2   <dc:title>Multimodal Transport Ontology</dc:title>
3   <owl:versionInfo>2.0</owl:versionInfo>
4   <dc:creator>Asier Moreno</dc:creator>
5   <dc:institution>DeustoTech Mobility</dc:institution>
6   <owl:imports rdf:resource="&mto;imports/mto-top.owl"/>
7 </owl:Ontology>
```

**Código 4.1:** Definición del fichero mto-core.owl

- ◇ *mto-top.owl* con todas las referencias externas necesarias para las entidades y propiedades de la ontología, importadas de otros vocabularios y vinculadas a la ontología base. *mto-core.owl* importa este archivo.

```
1 <owl:Ontology rdf:about="http://www.mto.org/mto/">
2   <dc:title>Multimodal Transport Ontology</dc:title>
3   <owl:versionInfo>2.0</owl:versionInfo>
4   <dc:creator>Asier Moreno</dc:creator>
5   <dc:institution>DeustoTech Mobility</dc:institution>
6   <owl:imports rdf:resource="www.geonames.org/ontology"/>
7   <owl:imports rdf:resource="www.opengis.net/ont/geosparql"/>
8   <owl:imports rdf:resource="www.w3.org/2006/time"/>
9 </owl:Ontology>
```

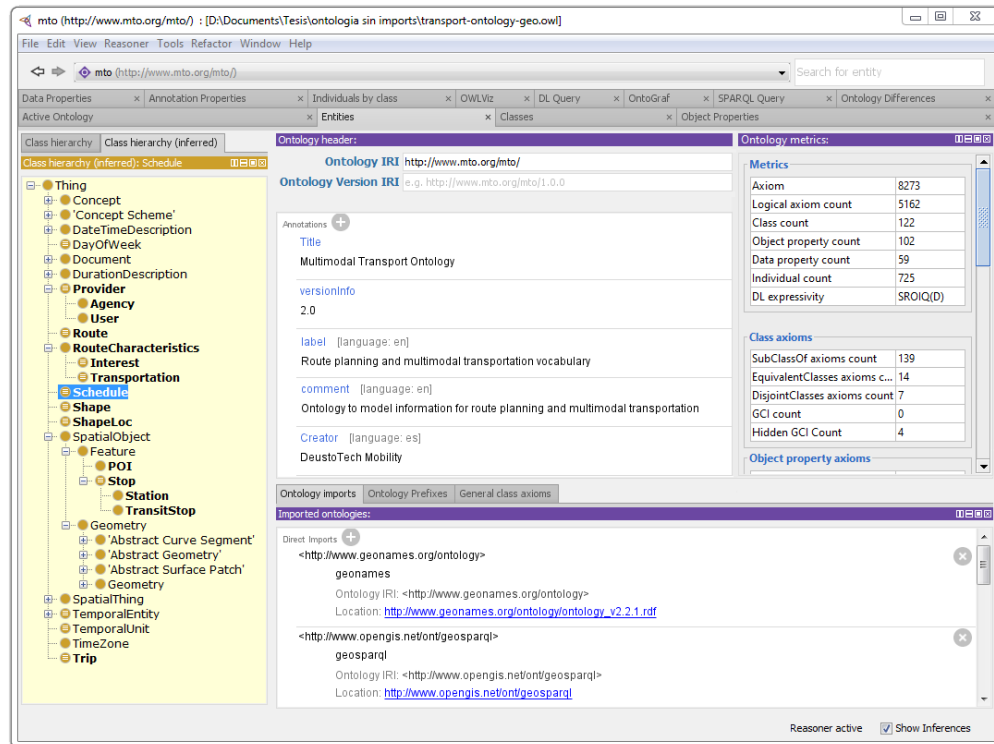
**Código 4.2:** Definición del fichero mto-top.owl

Esta organización facilita un mayor control sobre los vocabularios importados al centralizar en un único fichero todo el contenido externo vinculado a la ontología base, permitiendo una sustitución transparente de dichos vocabularios en caso de cambios o actualizaciones.



## 4.2 Formalización de la ontología

Para el desarrollo de MTO se ha utilizado la herramienta Protégé<sup>1</sup>, un editor ontológico y framework para el desarrollo de sistemas inteligentes desarrollado por la Universidad de Stanford y actualmente utilizado por más de 100.000 usuarios. La figura 4.2 muestra la interfaz principal del editor con la ontología cargada.



**Figura 4.2:** Fichero OWL de MTO cargado en Protégé

En primer plano se puede observar la descripción general de la ontología, con el título, el IRI y la versión actual, entre otros datos. En la parte inferior de la interfaz de usuario se listan los vocabularios importados; en este caso *GeoNames* y *GeoSPARQL*. La funcionalidad de dichos vocabularios se detalla en la sección 4.2.2. En la parte derecha se muestran estadísticas de la ontología con el número de axiomas, clases, relaciones e instancias generadas.

El panel amarillo de la izquierda muestra la jerarquía de clases de la ontología, con las entidades correspondientes a MTO resaltadas. Los detalles del vocabulario MTO, así como de los vocabularios externos importados para la generación del modelo de datos se exponen en las siguientes subsecciones.

<sup>1</sup> Protégé, <http://protege.stanford.edu>

## 4. Ontología para el transporte multimodal

---

### 4.2.1 Vocabulario

Dado que GTFS es actualmente el estándar *de facto* para la representación de datos de tránsito multimodal, se ha decidido tomar como base dicha especificación para definir y construir el vocabulario de MTO, reformulando los archivos CSV de dicho formato a entidades y relaciones dentro de la ontología resultante. Cada archivo CSV del formato GTFS se ha modelado como una (o varias, dependiendo de la complejidad y del diseño escogido) entidades de la ontología MTO, definiendo de igual modo las relaciones entre dichas entidades y las reglas lógicas que las soportan y que se detallarán a continuación. Esto permite generar un modelo semántico para un dominio concreto y con un alcance definido, ya analizado y validado, en el que la definición de las clases y sus propiedades son conocidas, favoreciendo la comprensión y facilitando la adopción del mismo.

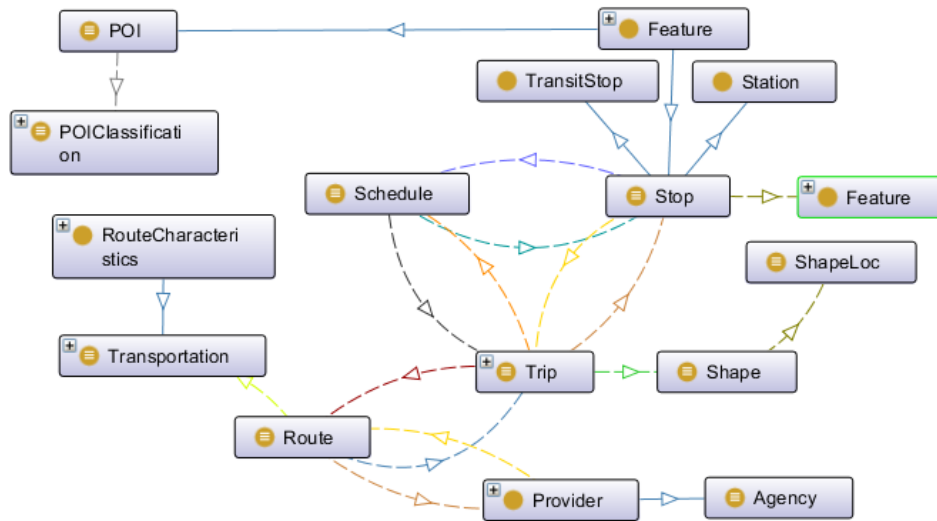
La sección 2.3.2 del presente documento describe la citada especificación GTFS, adjuntando también algunos ejemplos de los ficheros CSV utilizados por la misma. El hecho de reformular un formato consolidado como GTFS favorecerá la integración de grandes volúmenes de datos, posibilitando la generación semiautomatizada a partir de los ficheros CSV ya existentes y del adaptador desarrollado (descrito en el apartado 4.2.3) de un completo *dataset* de información semántica de transporte.

Se ha utilizado el componente *OntoGraf* de la herramienta *Protégé* para generar el siguiente diagrama de entidades (ver figura 4.3) donde se muestran los elementos y las relaciones principales de la ontología propuesta. Las relaciones representadas mediante líneas discontinuas describen propiedades funcionales asociadas al dominio (p. ej. una ruta (*Route*) representa un conjunto de viajes (*Trips*) compuestos por una serie de paradas (*Stops*)). Por otro lado, las relaciones representadas mediante líneas continuas en color azul describen relaciones de herencia entre conceptos (p. ej. una agencia de transporte (*Agency*) hereda del concepto general Proveedor (*Provider*), definiendo así una estructura para las entidades que pudieran generar información de tránsito).

Se detallan a continuación las clases o entidades que componen el vocabulario, junto con una descripción funcional, su especificación dentro de la ontología mediante lógica de primer orden<sup>2</sup> (específica del modelo) y su equivalencia, si la hubiera, con los ficheros CSV antes citados, provistos en el formato GTFS.

---

<sup>2</sup> En esta tesis se usa el mismo formato encontrado en [Anand 12] para presentar los axiomas



**Figura 4.3:** Entidades y relaciones principales de la ontología MTO

Se ha utilizado la sintaxis Manchester OWL [Horridge 06] para describir cada una de las clases. Dicha sintaxis permite reconocer de forma natural las relaciones entre entidades y su cardinalidad, así como los atributos principales de éstas.

- ◇ *Provider*. Clase abstracta que actúa como clase padre para todos los proveedores de datos. Permite establecer enlaces con entidades externas proveedoras de información de tránsito. No tiene equivalencia en GTFS.
- ◇ *Agency*. Empresa que proporciona datos de transporte multimodal. Hereda de *Provider*. Equivalente al archivo *agency.txt*, presente en la especificación GTFS.

```

1 (isProviderOf some Route)
2 and (isProviderOf only Route)
3 and (hasTimeZone some string)
4 and (hasURL some string)

```

**Código 4.3:** Descripción lógica de la clase Agency

- ◇ *RouteCharacteristics*. Clase abstracta que actúa como contenedor de opciones para las rutas, permitiendo su clasificación en función del tipo (turística, comercial, etc.) o del medio de transporte utilizado. No tiene equivalencia en GTFS.
- ◇ *Transportation*. Medio de transporte utilizado en una ruta. Hereda de *RouteCharacteristics*. No tiene equivalencia en GTFS.

## 4. Ontología para el transporte multimodal

---

- ◇ *Route*. Ruta de transporte público. Una ruta es un conjunto de viajes (*Trips*) suministrados por un mismo proveedor (*Provider*) y que se muestran a los usuarios como servicios independientes. Se contemplan varios modos de transporte. Equivalente al archivo *routes.txt*, presente en la especificación GTFS.

```
1 (hasTrip some Trip)
2 and (isProvidedBy some Provider)
3 and (typeOfTransportation some Transportation)
4 and (hasTrip only Trip)
5 and (isProvidedBy only Provider)
6 and (typeOfTransportation only Transportation)
7 and (hasName exactly 1 string)
```

**Código 4.4:** Descripción lógica de la clase Route

- ◇ *Trip*. Viajes para cada ruta (*Route*). Un viaje es una secuencia definida de dos o más paradas (*Stops*) que se producen en un horario (*Schedule*) específico. Contiene también información para la representación gráfica del recorrido y permite el cálculo de rutas en un planificador multimodal. Equivalente al archivo *trips.txt*, presente en la especificación GTFS.

```
1 (hasStop some Stop)
2 and (partOfRoute some Route)
3 and (hasSchedule only Schedule)
4 and (hasServiceOn only DayOfWeek)
5 and (hasShape only Shape)
6 and (partOfRoute only Route)
7 and (hasDirection only boolean)
8 and (hasEndDate exactly 1 dateTime)
9 and (hasStartDate exactly 1 dateTime)
```

**Código 4.5:** Descripción lógica de la clase Trip

- ◇ *Stop*. Clase abstracta que actúa como clase padre para todos los elementos con ubicación geográfica concreta. Hereda de *Feature* (ontología GeoSPARQL). Contiene dos propiedades especialmente relevantes: *hasGeometry*, que establece un formato estándar para definir la geometría del punto concreto en base al vocabulario de GeoSPARQL, y *belongsToPlace*, que vincula la posición del punto con la entidad geopolítica relacionada de la ontología GeoNames.

## 4.2 Formalización de la ontología

En la sección 4.2.2 se detalla el diseño utilizado para la gestión de la información geográfica (mediante el vínculo de esta entidad con vocabularios externos). Equivalente al archivo *stops.txt*, presente en la especificación GTFS.

```
1 (belongsToPlace some Feature)
2 and (hasTimetable only Schedule)
3 and (partOfTrip only Trip)
4 and (hasGeometry some Feature)
5 and (hasLatitude exactly 1 lat)
6 and (hasLongitude exactly 1 long)
7 and (hasName exactly 1 string)
```

**Código 4.6:** Descripción lógica de la clase Stop

- ◇ *TransitStop*. Especialización de la clase *Stop* para las paradas en ruta que forman parte de un viaje (*Trip*). Hereda de *Stop*. No tiene equivalencia en GTFS.
- ◇ *Station*. Especialización de la clase *Stop* para las paradas consideradas como estaciones, entidades desde donde se inician o terminan los distintos viajes (*Trip*). Hereda de *Stop*. No tiene equivalencia en GTFS.
- ◇ *POIClassification*. Extensión de la ontología base de transporte. Establece una clasificación temática para los puntos de interés turístico o POIs (*Points of Interest*). No tiene equivalencia en GTFS.
- ◇ *POI*. Puntos de interés turístico. Extensión de la ontología base de transporte basada en el enlace de ésta con la plataforma de datos colaborativos *Linked-GeoData*. En la sección 4.3.1 se detalla el diseño utilizado. Hereda de *Feature* (GeoSPARQL) ya que define una localización geográfica concreta para los puntos de interés enlazados. No tiene equivalencia en GTFS.

```
1 (hasClassification some POIClassification)
2 and (hasLatitude exactly 1 lat)
3 and (hasLongitude exactly 1 long)
4 and (hasName exactly 1 string)
5 and (providedBy exactly 1 string)
```

**Código 4.7:** Descripción lógica de la clase POI

## 4. Ontología para el transporte multimodal

---

- ◇ *Schedule*. Planificación horaria establecida para un determinado vehículo y ruta. Horarios en los que un vehículo llega a una parada (*Stop*) concreta y sale de ella en cada viaje (*Trip*). Permite el cálculo de rutas en un planificador multimodal. Equivalente al archivo *stoptimes.txt*, presente en la especificación GTFS.

```
1 (appliedToStop some Stop)
2 and (appliedToTrip some Trip)
3 and (appliedToStop only Stop)
4 and (appliedToTrip only Trip)
5 and (distanceTraveled only double)
6 and (hasArrivalTime exactly 1 dateTime)
7 and (hasDepartureTime exactly 1 dateTime)
8 and (hasStopNumber exactly 1 int)
```

**Código 4.8:** Descripción lógica de la clase Schedule

- ◇ *Shape*. Conjunto de reglas para el trazado de líneas en un mapa o grafo representando las rutas de un servicio de transporte. Asociado a los viajes *Trips* permite visualizar gráficamente el recorrido de los mismos en un planificador multimodal. Equivalente al archivo *shapes.txt*, presente en la especificación GTFS.

```
1 (hasShapeLocation some ShapeLoc)
2 and (hasShapeLocation only ShapeLoc)
```

**Código 4.9:** Descripción lógica de la clase Shape

- ◇ *ShapeLoc*. Datos concretos de localización enfocados a la representación en un mapa o grafo de un recorrido de transporte público. Permite el cálculo de rutas en un planificador multimodal. Junto con los datos de *Shape*, equivalente al archivo *shapes.txt*, presente en la especificación GTFS.

```
1 (distanceTraveled only double)
2 and (hasLatitude exactly 1 lat)
3 and (hasLongitude exactly 1 long)
4 and (hasSequenceNumber exactly 1 int)
```

**Código 4.10:** Descripción lógica de la clase ShapeLoc

### 4.2.2 Gestión de la información geográfica

Además del modelo generado a partir de la transformación de entidades provenientes de GTFS también se utilizan otros vocabularios externos para la gestión de información temporal, geográfica y geopolítica, todos ellos necesarios para la correcta modelización del dominio de transporte.

A este respecto, una de las características fundamentales de un sistema que pretende modelar información de transporte es el diseño utilizado para almacenar la información geográfica. En este caso las decisiones se han basado en la estandarización y la futura extensibilidad y/o reusabilidad de la ontología diseñada, como se justificará en detalle a lo largo de los apartados siguientes.

#### 4.2.2.1 Gestión de la información geoespacial

En cuanto a la estandarización se ha seguido una doble vertiente: los puntos georeferenciados se han definido mediante las propiedades *latitude* y *longitude*, ampliamente soportadas en el modelado de ontologías con elementos de este tipo, siguiendo el estándar WGS84 lat/long<sup>3</sup>. Sin embargo, dicho modelado no permite realizar consultas geográficas sobre los puntos de la ontología (distancia entre paradas, localización de un punto de interés respecto a una ruta, etc.) por lo que también se ha definido la propiedad *hasGeometry* como atributo de la clase *Stop*, extendiendo y heredando de la ontología GeoSPARQL<sup>4</sup>.

El estándar OGC GeoSPARQL soporta la representación y consulta de datos geoespaciales en la Web Semántica mediante la definición de un vocabulario en RDF y la extensión del lenguaje de consulta SPARQL para el procesamiento de dichos datos geoespaciales. El fragmento 4.11 muestra un ejemplo de cómo dotar de información de geolocalización a un elemento dentro de la ontología.

```
1 mto:Stop a geo:Feature;  
2   rdfs:label "Vitoria Avenida";  
3   geo:hasGeometry ex:Point.  
4 ex:Point a geo:Point;  
5   geo:asWKT "POINT(-2.681353 42.849506)"^^geo:WktLiteral.
```

**Código 4.11:** Definición de un elemento geolocalizado mediante GeoSPARQL

<sup>3</sup> Basic Geo (WGS84 lat/long), <http://www.w3.org/2003/01/geo>

<sup>4</sup> GeoSPARQL, <http://www.opengeospatial.org/standards/geosparql>

## 4. Ontología para el transporte multimodal

---

Siguiendo el ejemplo anterior, la primera línea indica que el recurso *mto:Stop* hereda del tipo *geo:Feature* de GeoSPARQL, un concepto del mundo real con una ubicación determinada. La segunda línea le da a este nuevo recurso un nombre o descripción. La tercera línea le da a la instancia una ubicación espacial mediante *geo:hasGeometry*. La cuarta línea establece que dicha ubicación espacial es un punto (*geo:Point*), y la quinta línea define el punto con sus coordenadas.

Gracias a la incorporación de elementos geoposicionados en la ontología MTO es posible la ejecución de cálculos geoespaciales avanzados, como comparaciones entre geometrías; definidas en este caso para paradas, puntos de interés (POIs) o rutas dentro del modelo de tránsito. Las funciones de filtrado proporcionadas por GeoSPARQL habilitan la comparación entre varias geometrías resultado de una consulta o bien la comparación directa entre el resultado de una consulta y una geometría explícita. Los filtros proporcionados permiten, por ejemplo, calcular distancias entre puntos, definir la situación de un punto con respecto a un área u obtener puntos de intersección entre segmentos, entre otros cálculos geométricos.

```
1 SELECT DISTINCT ?poi ?nompoi ?lat ?lon
2 WHERE {
3   ?poi a mto:POI; mto:hasName ?nompoi.
4   ?poi mto:hasLatitude ?lat; mto:hasLongitude ?lon.
5   ?poi geo:hasGeometry ?g.
6   ?g geo:asWKT ?wkt.
7   FILTER (geof:distance(?wkt, "LINESTRING(-2.50 42.50, -2.60
8     42.60)"^^geo:wktLiteral, units:metre) < 5000)
```

**Código 4.12:** Consulta SPARQL. Selección de POIs a 5km de una recta dada

A modo de ejemplo, en el fragmento de código 4.12 se muestra una consulta SPARQL para la selección de un conjunto de POIs, incluyendo su nombre y posición. Para filtrar los resultados de la consulta se utiliza la función *geof:distance* estableciendo una distancia de 5km en relación a una recta geométrica prefijada.

El uso de este tipo de filtros, derivados de la inclusión de GeoSPARQL en el diseño de MTO, junto con el despliegue de un servidor de contenido semántico geoespacial, permite dotar a la ontología de funcionalidades avanzadas de búsqueda y filtrado inasumibles por otro tipo de formatos para la representación de información de transporte, habilitando el acceso completo y directo a los datos almacenados.



### 4.2.2.2 Gestión de la información geopolítica

A nivel de extensibilidad se ha decidido enlazar la ontología con GeoNames<sup>5</sup>, una base de datos geográfica gratuita y accesible a través de Internet bajo una licencia *Creative Commons* enfocada al almacenamiento de información geopolítica.

Dicha base de datos contiene más de 10 millones de nombres geográficos que corresponden a más de 7,5 millones de lugares o *Features*. Estos nombres están organizados en 9 categorías y 645 subcategorías, atendiendo a parámetros de organización administrativa como la comunidad autónoma, provincia o población a la que pertenecen. Dicha información se completa con datos como la latitud, la longitud, la altitud o el código postal disponibles en varios idiomas para cada ubicación.

La información de GeoNames se encuentra accesible por varios medios (página web dedicada, API, servicios web, etc.) En el caso de MTO se utilizará el enlace semántico provisto mediante recursos RDF. Cada elemento o *Feature* se representa como un recurso web identificado por un URI estable. Este URI permite acceder a la descripción RDF del elemento de la ontología GeoNames. Esta descripción incluye además enlaces semánticos con DBpedia<sup>6</sup> (proyecto para generar información semántica a partir del contenido de Wikipedia) y otros vocabularios enlazados.

En el caso de MTO, el enlace con GeoNames se realiza mediante la propiedad *belongsToPlace* referenciando la URI de GeoNames que corresponde al territorio en el que se encuentra el punto concreto en la ontología. El fragmento 4.13 muestra un ejemplo de cómo dotar de información geopolítica a un elemento, en este caso la parada Deusto, utilizando este enlace con GeoNames.

```
1 <owl:NamedIndividual rdf:about="mto:Stop/885/3688">
2   <rdf:type rdf:resource="mto:Stop"/>
3   <hasLongitude rdf:datatype="wgs84#long">-2.94990</hasLongitude>
4   <hasLatitude rdf:datatype="wgs84#lat">43.26127</hasLatitude>
5   <hasName rdf:datatype="xsd:string">Deusto</hasName>
6   <belongsToPlace rdf:resource="http://sws.geonames.org/3123912">
7   <geo:hasGeometry rdf:resource="mto:Point/885/3684"/>
8 </owl:NamedIndividual>
```

**Código 4.13:** Especificación de una parada (*Stop*). Enlace con GeoNames

<sup>5</sup> GeoNames Ontology, <http://www.geonames.org/ontology>

<sup>6</sup> DBpedia, <http://wiki.dbpedia.org>

## 4. Ontología para el transporte multimodal

---

Al igual que en el punto anterior, el uso de este tipo de información geográfica compleja, derivada de la inclusión de GeoNames en el diseño de MTO, permite dotar a la ontología de funcionalidades de consulta basadas en conceptos semánticos. Estos conceptos, como la provincia en la cual se efectúa una determinada ruta, ofrecen de forma directa una capa adicional de filtrado, dotando de mayor exactitud a los datos provistos. El fragmento de código 4.14 muestra un ejemplo de consulta utilizando este enlace con GeoNames.

```
1 SELECT DISTINCT ?route_long_name ?nommuni ?prov
2 WHERE {
3   ?route a mto:Route; mto:hasName ?route_long_name.
4   ?route mto:hasTrip ?trip.
5   ?trip mto:hasStop ?stop.
6   ?stop mto:belongsToPlace ?muni.
7   ?muni gn:name ?nommuni; gn:parentADM2 ?prov.
8   ?prov gn:officialName ?nameprov
9   FILTER(regex(str(?nameprov), 'Alava', 'i'))
10 } ORDER BY ?route
```

**Código 4.14:** Consulta SPARQL. Selección de rutas para una provincia dada

En este caso la consulta realiza una selección de rutas de transporte, incluyendo su nombre y su ubicación. Para filtrar los datos resultantes, se utiliza información semántica sobre la provincia en la que se lleva a cabo la ruta (línea 9, filtro por nombre). La propiedad *mto:belongsToPlace* (línea 6) es responsable de recopilar dicha información mediante la vinculación de la ontología con los datos de GeoNames (líneas 7 y 8, prefijo *gn:*).

### 4.2.3 Adaptador GTFS

Tal y como se ha apuntado al inicio del capítulo, se ha decidido tomar como base la especificación GTFS para definir y construir el vocabulario de MTO, reformulando los archivos CSV a entidades y relaciones dentro de la ontología. GTFS es uno de los formatos más extendidos para la representación de información de tránsito, con lo que la conversión a MTO de los datos ya existentes en este formato permitirá la creación de un completo *dataset* de información semántica de transporte multimodal.

## 4.2 Formalización de la ontología

---

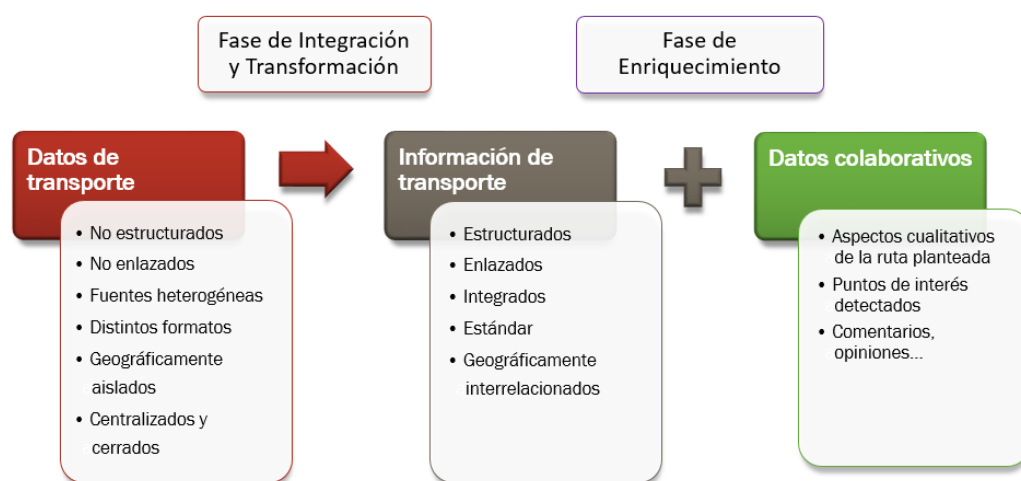
De acuerdo también con la metodología NeOn descrita con anterioridad, que focaliza el desarrollo en la reutilización y la transformación de recursos no ontológicos, se ha desarrollado un adaptador para la conversión automatizada de dichos recursos. El adaptador es una aplicación de escritorio portable y multiplataforma desarrollada en Java. Su funcionalidad, secuencialmente descrita, es la siguiente:

1. Carga y validación inicial de los ficheros GTFS de las distintas compañías de transporte. Se comprueba que estén bien formados de tal modo que la conversión pueda completarse correctamente. Posteriormente se procede a la extracción de los ficheros CSV que forman parte del contenedor GTFS.
2. Carga del fichero OWL base con el modelo ontológico desarrollado y descrito durante el presente capítulo. La gestión del modelo dentro del entorno de desarrollo del adaptador se ha efectuado mediante OWL-API, implementación Java de referencia para la manipulación y la serialización de ontologías.
3. Transformación de los recursos no ontológicos obtenidos desde los ficheros CSV: creación de las distintas instancias, con sus correspondientes propiedades, relaciones y restricciones, de acuerdo con el vocabulario y las equivalencias descritas en el apartado 4.2.1. Generación de índices únicos para cada agencia de transporte.
4. Generación de los enlaces semánticos y la jerarquía de clases correspondiente, incluyendo los enlaces con vocabularios como GeoNames y GeoSPARQL. En este paso se vinculan los conceptos provenientes de GTFS con vocabularios relacionados relevantes siguiendo la filosofía LOD.
5. Serialización de los ficheros OWL resultantes con los datos de transporte semantizados y enlazados. Para facilitar su carga en el almacén semántico correspondiente el número de ficheros generados dependerá del tamaño del fichero original.

El proceso de carga de los ficheros GTFS y generación del contenido ontológico se encuentra completamente automatizado. La herramienta provista realiza el proceso de validación y descompresión de los ficheros fuente, y mediante un proceso de *scraping* o parseo efectúa la conversión, incluyendo las referencias semánticas a los vocabularios descritos en la sección 4.2.2.

### 4.3 Enriquecimiento de la información

Uno de los conceptos principales sobre los que se fundamenta la arquitectura propuesta es la mejora de la información de transporte que actualmente se provee al usuario. El gráfico 4.4, mostrado a continuación, presenta un esquema ampliado del conjunto de estados en los que se puede encontrar dicha información, así como las distintas fases planteadas. La primera fase, de integración y transformación, se corresponde con la primera parte del presente capítulo, donde se muestra el diseño de la ontología MTO. La segunda, la fase de enriquecimiento, se detallará a continuación.



**Figura 4.4:** Enriquecimiento de la información de transporte

Una de las carencias encontradas en el estudio de las arquitecturas para la provisión de información de transporte es el no tener en cuenta factores relevantes como el contexto o la agregación de información relacionada a la hora de proveer los resultados al usuario. Esta carencia limita la utilidad de la información provista al encontrarse aislada y limitada al conjunto de datos utilizado en la búsqueda.

La utilización en este trabajo de un modelo formal, como las ontologías, para la representación de la información de transporte, junto con las capacidades aportadas por la infraestructura global de la web semántica y el conjunto de datos enlazados establecen las bases para el enriquecimiento efectivo de la información de transporte generada, ya semantizada y disponible.

### 4.3.1 Vocabularios enlazados

El hecho de representar la información de transporte como una ontología facilita el aprovechamiento de características ventajosas como la posibilidad de enlazar los datos de tránsito con otros conjuntos de datos existentes en Internet y de interés para el dominio concreto del transporte multimodal. Tal y como se describe en la sección 2.4.3.1 del estado del arte, durante los últimos años se esta produciendo un auge, con el apoyo tanto de usuarios como de instituciones y gobiernos, en cuanto a la publicación de datos enlazados abiertos o LOD. Se produce así una colaboración efectiva, donde usuarios e instituciones participan para completar y mejorar los datos existentes, enriqueciendo la información y habilitando la creación de servicios en torno a ella.

La fase de enriquecimiento cumple por tanto con el objetivo de dotar de valor añadido a la información de transporte resultante de la fase anterior mediante el uso de dicho componente colaborativo, partiendo de la capacidad de la arquitectura y de la ontología propuesta para enlazar dichos datos. Esto supone un cambio en cuanto a la calidad de la información provista, al integrar en ésta aspectos no cuantitativos, aportados directamente por los usuarios e imposibles de adquirir de otro modo. Se ha realizado un estudio de los vocabularios existentes relacionados con el ámbito del transporte y que podrían enriquecer la ontología desarrollada, con la idea también de ejemplificar cómo la arquitectura y el modelo ontológico propuestos permiten lograr un enriquecimiento de los datos disponibles mediante su enlace con vocabularios externos afines, como el escogido en este caso.

```
1 <rdf:Description rdf:about="http://linkedgeodata.org/258768315">
2   <rdf:type rdf:resource="geo:Feature" />
3   <rdf:type rdf:resource="mto:POI" />
4   <mto:hasClassification rdf:resource="mto:Airport" />
5   <mto:hasClassification rdf:resource="mto:Services" />
6   <mto:hasClassification rdf:resource="mto:Transport" />
7   <mto:hasLatitude>42.8802</mto:hasLatitude>
8   <mto:hasLongitude>-2.7345</mto:hasLongitude>
9   <mto:hasName>Aeropuerto de Vitoria-Gasteiz</mto:hasName>
10  <mto:providedBy>linkedgeodata.org</mto:providedBy>
11  <geo:hasGeometry rdf:nodeID="b10251" />
12 </rdf:Description>
```

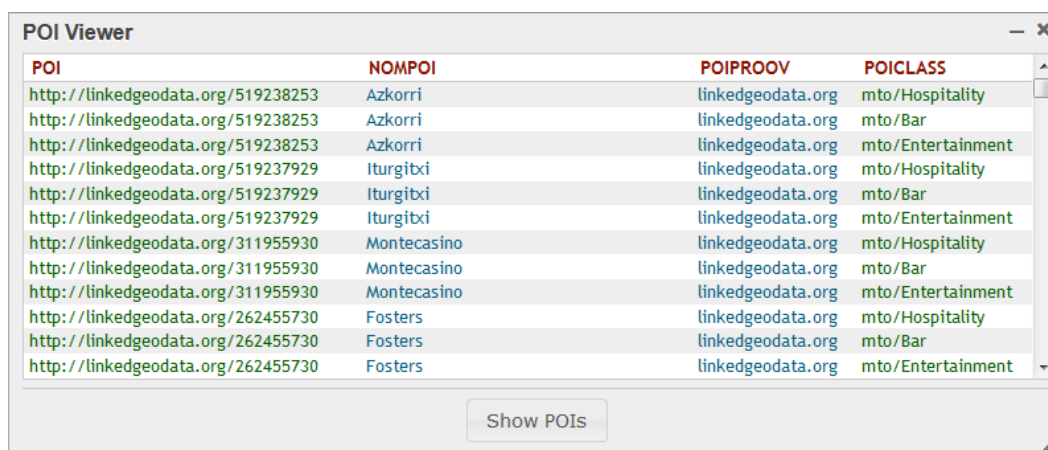
**Código 4.15:** Especificación de un POI proveniente de LinkedGeoData

#### 4. Ontología para el transporte multimodal

Se ha optado por buscar una extensión de MTO enlazando y clasificando los puntos de interés (POIs) colaborativos expuestos por la ontología LinkedGeoData. Dicha ontología utiliza la información recopilada por el proyecto OpenStreet-Map (en el que los usuarios colaboran definiendo elementos geoespaciales), semantizándola y generando *datasets* RDF según los principios de LOD. El hecho de utilizar este conjunto de información es un ejemplo de las capacidades de extensibilidad y reusabilidad del modelo propuesto de cara a proveer información relacionada al usuario según su interés.

La fórmula elegida para relacionar e integrar la ontología MTO junto con los datos provistos por LinkedGeoData<sup>7</sup> se basa en la definición de las clases *POI* y *POIClassification*. La entidad POI es un contenedor para puntos de interés turístico colaborativos que hereda de la clase *Feature* (GeoSPARQL) para obtener la posición geográfica (latitud y longitud) de dichos puntos. También almacena el URI original del elemento, su nombre y descripción, el proveedor de la información (además de POIs de LinkedGeoData también se gestionan POIs de GeoNames) y una clasificación (*POIClassification*) asignada de acuerdo a su utilidad y función: servicios, entretenimiento, turismo, etc.

El fragmento de código 4.15 muestra la descripción RDF de un elemento de la ontología MTO derivado del enlace con LinkedGeoData, en este caso la instancia correspondiente al Aeropuerto de Vitoria-Gasteiz.



POI	NOMPOI	POIPROOV	POICLASS
<a href="http://linkedgeodata.org/519238253">http://linkedgeodata.org/519238253</a>	Azkorri	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Hospitality
<a href="http://linkedgeodata.org/519238253">http://linkedgeodata.org/519238253</a>	Azkorri	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Bar
<a href="http://linkedgeodata.org/519238253">http://linkedgeodata.org/519238253</a>	Azkorri	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Entertainment
<a href="http://linkedgeodata.org/519237929">http://linkedgeodata.org/519237929</a>	Iturgitxi	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Hospitality
<a href="http://linkedgeodata.org/519237929">http://linkedgeodata.org/519237929</a>	Iturgitxi	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Bar
<a href="http://linkedgeodata.org/519237929">http://linkedgeodata.org/519237929</a>	Iturgitxi	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Entertainment
<a href="http://linkedgeodata.org/311955930">http://linkedgeodata.org/311955930</a>	Montecasino	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Hospitality
<a href="http://linkedgeodata.org/311955930">http://linkedgeodata.org/311955930</a>	Montecasino	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Bar
<a href="http://linkedgeodata.org/311955930">http://linkedgeodata.org/311955930</a>	Montecasino	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Entertainment
<a href="http://linkedgeodata.org/262455730">http://linkedgeodata.org/262455730</a>	Fosters	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Hospitality
<a href="http://linkedgeodata.org/262455730">http://linkedgeodata.org/262455730</a>	Fosters	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Bar
<a href="http://linkedgeodata.org/262455730">http://linkedgeodata.org/262455730</a>	Fosters	<a href="http://linkedgeodata.org">linkedgeodata.org</a>	mto/Entertainment

Figura 4.5: Resultados de una consulta SPARQL: relación de restaurantes

<sup>7</sup> LinkedGeoData, <http://linkedgeodata.org>

Siguiendo el ejemplo, la línea 1 apunta a la URI de LinkedGeoData, esto permite el enlace entre ambos vocabularios. En las líneas 2 y 3 se indica el tipo del elemento: en este caso una *Feature* de GeoSPARQL y un *POI* en MTO. Las líneas 4 a 6 especifican la clasificación del elemento en MTO, favoreciendo posteriores búsquedas y la línea 11 relaciona el elemento con GeoSPARQL habilitando también las búsquedas geoespaciales. En relación a dichas búsquedas atendiendo a la clasificación de los elementos enlazados así como a su ubicación geográfica, la figura 4.5 muestra los resultados de una consulta SPARQL realizada sobre MTO en la que se solicitan los POIs (en este caso los restaurantes) provistos por LinkedGeoData en la provincia de Bizkaia.

#### 4.3.2 Información contextual

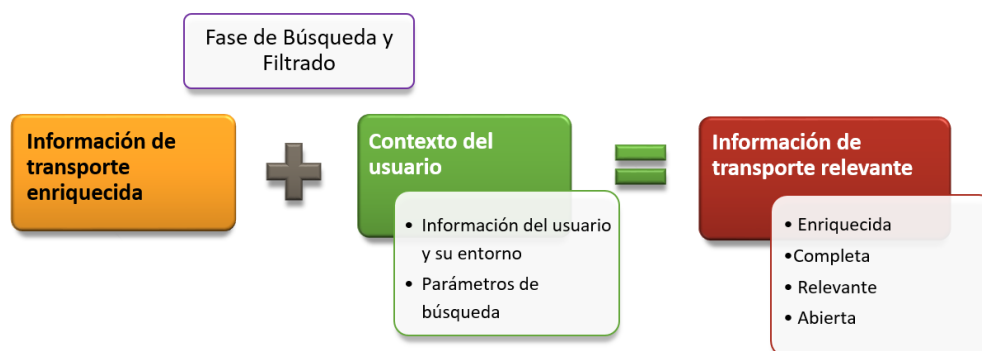
Además de asegurarse de contar con información estructurada, integrada y enlazada con otros datos relacionados, otra de las cuestiones importantes a la hora de generar información más precisa y relevante es tener en cuenta el contexto concreto del usuario. Sin él, los datos ofrecidos perderían interés al no aportar información personalizada.

El término “*context-aware*” fue introducido por primera vez por Schilit en [Schilit 94] y se refería al software que se adaptaba a su localización, a la colección de objetos y personas que lo rodean, y a los cambios sufridos por estos actores. Dey introduce en [Dey 01] una definición de contexto más amplia y más comúnmente aplicada en la actualidad que es la siguiente:

*“El contexto es cualquier información que puede ser utilizada para caracterizar una situación de una entidad. Una entidad es una persona, lugar u objeto que es considerado relevante a la interacción entre un usuario y una aplicación, incluyendo al usuario y a la aplicación como tal”.*

En base a esta definición, deducimos que la inclusión de información contextual dentro de un sistema software, asumiendo e incorporando las necesidades concretas del usuario, permite precisar y personalizar la información. Es por ello que se ha buscado la forma de recoger la información contextual en el ámbito de los sistemas de planificación de rutas con el objetivo de mejorar la información provista. Se ha definido una fase de búsqueda y filtrado (ver figura 4.6) en la que recoger tanto los datos básicos del usuario como sus intereses en forma de parámetros de búsqueda.

## 4. Ontología para el transporte multimodal



**Figura 4.6:** Evolución de los datos de transporte: información contextual

Para cada consulta realizada, el sistema almacena datos como la localización o el historial de búsqueda, de modo que puede establecer el contexto concreto de interés del usuario. Dicho contexto, una vez modelado, se utiliza para filtrar la información de acuerdo a sus necesidades y circunstancias concretas. La arquitectura propuesta define además, para complementar dicho modelo, una búsqueda facetada a partir de las siguientes propiedades referidas a los puntos de interés en ruta:

- ◇ Nombre del punto de interés concreto a buscar.
- ◇ Fuente de información a consultar (GeoNames y/o LinkedGeoData).
- ◇ Localización geográfica de la consulta:
  - A nivel de organización administrativa (ej. POIs en Álava, código 4.14)
  - A nivel de ubicación geoespacial (ej. POIs a 5 km en ruta, código 4.12)
- ◇ Clasificación jerárquica del punto de interés:
  - Servicios: Transporte, Alojamiento, Utilidades
  - Entretenimiento: Ocio, Tiendas, Hostelería
  - Puntos de interés: Cultural, Medioambiental, Turístico

Se cumple así con el objetivo de dotar de valor añadido a la información de transporte resultante de la fase anterior mediante el uso del componente contextual. Dicha característica permite a los usuarios filtrar la información de acuerdo a sus intereses o necesidades particulares, personalizando de este modo la información provista y dotando de relevancia a la misma.



*Si tu intención es describir la  
verdad, hazlo con sencillez y la  
elegancia déjasela al sastre.*

Albert Einstein

# 5

## Arquitectura del sistema

**L**a manera en la que se representa y se estructura la información de transporte, en este caso mediante la creación de la ontología MTO, supone parte de la innovación realizada en el presente trabajo. Sin embargo, dicho modelo de datos debe tener un soporte que permita su consulta, habilitando a su vez todas las características avanzadas de búsqueda y filtrado de la información disponible. Del mismo modo, la arquitectura propuesta persigue solventar algunas de las limitaciones encontradas en los sistemas de gestión de información de transporte existentes, en muchas ocasiones convertidos en silos de datos, impidiendo el acceso universal a información actual e integrada.

En este capítulo se describe el proceso de diseño y desarrollo de la arquitectura distribuida propuesta, cuyo despliegue está permitiendo la publicación efectiva de la información semántica generada con MTO. En primer lugar se exponen las capacidades principales de dicha arquitectura, atendiendo a características como la consulta de los datos y la provisión y explotación de la información disponible. A continuación se realiza un análisis de los componentes software utilizados en el desarrollo de la misma, para después profundizar en su diseño y funcionalidad, así como en las herramientas y mecanismos empleados para dotarla de su naturaleza distribuida.

### 5.1 Capacidades y características

En capítulos previos se han detallado las capacidades y características positivas aportadas por la modelización semántica de la información de transporte y la alineación del modelo propuesto con paradigmas como *Linked Open Data*. La definición de MTO como modelo ontológico con un dominio y alcance concretos, su vinculación con vocabularios relacionados así como la gestión de la información geográfica y la inclusión del contexto, suponen claras contribuciones del modelo propuesto en términos de estructuración y clasificación de la información, interoperabilidad, extensibilidad, enriquecimiento y mejora de la información.

En este punto, quedan aún al menos dos características consideradas relevantes para la gestión de la información de transporte disponible, apuntadas como tal en capítulos previos (ver tabla comparativa 3.1): la consulta efectiva y completa de los datos, y la accesibilidad y posibilidad de explotación de la información disponible.

En una arquitectura enfocada a la provisión de información interoperable e integrada resulta fundamental contar con un protocolo de acceso a datos que garantice tanto la completitud de la información proporcionada como su apertura y disponibilidad universal, todo ello en relación con el formato y el mecanismo de consulta de los datos almacenados. Es decir, la arquitectura debe garantizar que el usuario pueda acceder a todos los datos disponibles (ya enlazados e integrados mediante la ontología) devolviendo resultados agregados en un formato adecuado (atendiendo a las características de los formatos para la web establecidas en la tabla 2.2), fácilmente explotable y, a poder ser, personalizable.

De acuerdo a estas características, formatos como GTFS no proveen de ningún mecanismo para facilitar el acceso a la información. De esta forma, dicha información se encuentra alojada en servidores dispersos no interrelacionados. El formato provisto, aunque estructurado en primera instancia (CSV), llega al usuario final como fichero comprimido (ZIP) por lo que no es directamente tratable ni consultable. Otros formatos más enfocados a la web como WFS sí habilitan un mecanismo abierto de acceso a los datos. Sin embargo este acceso se encuentra limitado en cuanto a contenido y formato, dependiendo de las características que el desarrollador haya desplegado. Su utilización además es compleja debido al uso de ficheros GML para gestionar las peticiones, lo que añade mucha información de cabecera innecesaria, ofuscando los datos y penalizando el rendimiento.

## 5.1 Capacidades y características

De acuerdo a lo expuesto anteriormente, la arquitectura propuesta plantea el uso de la web semántica como infraestructura estable sobre la que construir el mecanismo de gestión y provisión de la información, haciendo un uso extensivo del conjunto de protocolos provisto (desde HTTP a OWL o SPARQL) y basándose en la metodología LOD para la publicación de información abierta, integrada y enlazada con otros conjuntos de datos.

En la sección 5.3 del presente capítulo se describe con mayor detalle la arquitectura propuesta, así como los elementos software que la componen. Sin embargo, antes de indagar en el diseño y desarrollo de la misma, se quiere mostrar mediante un ejemplo concreto cómo el diseño planteado para la provisión de la información, mediante el uso de servidores SPARQL accediendo a datos semánticos en formato MTO, consigue cumplir satisfactoriamente con las capacidades de consulta y acceso a datos anteriormente citadas.

Así, el fragmento de código 5.1 muestra una consulta SPARQL sobre uno de los servidores desplegados, que contienen los datos de transporte público de la Comunidad Autónoma Vasca (CAV), semantizados y cargados mediante el adaptador ontológico desarrollado (descrito en la sección 4.2.3 del capítulo previo).

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema>
2 PREFIX mto: <http://www.mto.org/mto/>
3
4 SELECT ?agency_id ?agency_name ?agency_url ?agency_timezone ?
   agency_lang
5 WHERE {
6   ?agency_id a mto:Agency.
7   ?agency_id mto:hasName ?agency_name.
8   ?agency_id mto:hasURL ?agency_url.
9   ?agency_id mto:hasTimeZone ?agency_timezone
10 } ORDER BY ?agency_id LIMIT 10
```

**Código 5.1:** Consulta SPARQL. Selección de agencias y exportación a CSV

Esta consulta en concreto, solicitando información sobre las agencias de transporte y resuelta en menos de 1 segundo por el servidor, exporta los datos directamente en formato CSV siguiendo las directrices de la especificación GTFS. Esto es debido a las variables personalizadas utilizadas (línea 4 del código) y a las capacidades de exportación y agregación del servidor utilizado.

## 5. Arquitectura del sistema

---

Lo relevante de la consulta y los resultados es mostrar cómo mediante la solución propuesta para la provisión de información semántica es posible generar múltiples formatos de salida de datos, como el utilizado en la especificación GTFS, de forma rápida y eficaz, obteniendo resultados directamente procesables e integrados, y todo ello utilizando un único punto de acceso a la información disponible.

Se pueden comparar los resultados generados por la consulta (fragmento de código 5.2) con el fichero CSV de agencias (fragmento 2.2) presentado en la descripción del formato GTFS en el capítulo 2. De este modo, el formato y los ficheros utilizados en GTFS pasan a ser un subconjunto de datos reproducible, incluido dentro de la información que la arquitectura propuesta es capaz de proveer.

```
1 agency_id,agency_name,agency_url,agency_timezone,agency_lang
2 http://www.mto.org/agency/002,Alavabus,www.alava.net/alavabus,Eu,
3 http://www.mto.org/agency/025,Tuvisa,www.vitoria-gasteiz.org,Eu,
4 http://www.mto.org/agency/103,Lurraldebus Pesa,www.pesa.net,Eu,
5 http://www.mto.org/agency/173,MetroBilbao,www.metrobilbao.com,Eu,
6 http://www.mto.org/agency/387,Bilbobus,bilbao.net/bilbobus,Eu,
7 http://www.mto.org/agency/402,Renfe,www.renfe.com,Eu,
8 http://www.mto.org/agency/472,LaBurundesa,www.laburundesa.com,Eu,
9 http://www.mto.org/agency/498,Euskotren,www.euskotren.es,Eu,
10 http://www.mto.org/agency/508,DonostiaBus,www.dbus.es,Eu,
11 http://www.mto.org/agency/546,Herm.Berrotaran,www.pasaia.net,Eu,
12 http://www.mto.org/agency/667,Etxebarribus,www.leintz.com,Eu,
13 http://www.mto.org/agency/767,Funicular Artxanda,bilbao.net,Eu,
14 http://www.mto.org/agency/885,La Union,www.autoslaunion.com,Eu,
15 http://www.mto.org/agency/961,Bizkaibus,www.bizkaia.net,Eu,
```

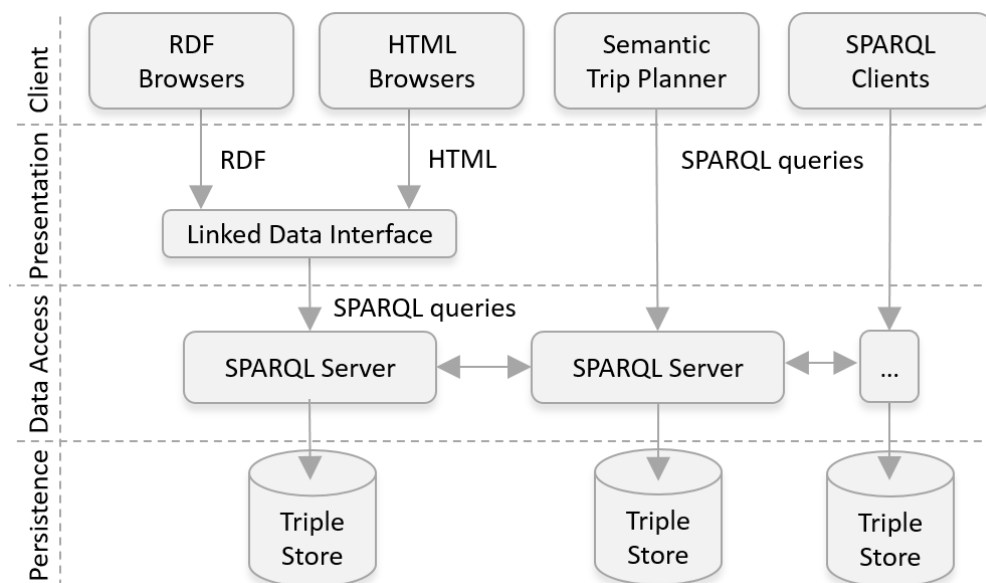
**Código 5.2:** Conjunto de agencias en formato CSV compatible con GTFS

El hecho de poder acceder a un conjunto de datos extenso (con más de 10 agencias de transporte y 1500 rutas distintas dentro de la CAV en el despliegue realizado) y complejo (con soporte para consultas semánticas y geoespaciales avanzadas y múltiples formatos de salida de datos) desde un único lugar asociado a una URL y con una arquitectura estable basada en tecnologías web, habilitando además el diseño distribuido y el balanceo de carga entre servidores como se justificará en detalle en la sección 5.3 del presente capítulo, supone otra de las aportaciones relevantes realizadas en el presente trabajo de tesis.

## 5.2 Diseño y componentes software

Como se ha adelantado, la arquitectura propuesta para la gestión de la información de transporte se fundamenta en el uso de servicios y herramientas de la web semántica, como infraestructura estable sobre la que construir el mecanismo de almacenamiento y provisión de la información generada.

Así, la información semántica disponible, en este caso las tripletas que conforman los datos en la ontología MTO, es almacenada en servidores SPARQL distribuidos e interrelacionados que publican sus datos mediante el despliegue de *end-points* (servicios web que permiten realizar consultas SPARQL sobre un grafo de entrada compuesto por tripletas RDF) disponibles en la web de forma pública. Dicha información podrá ser accedida desde múltiples dispositivos y plataformas y, por tanto, deberá poder ser consultada en diversos formatos y capacidades. Se han desplegado para ello servidores encargados de realizar esta tarea, sirviendo el contenido semántico de forma distinta dependiendo del cliente que solicite los datos.



**Figura 5.1:** Arquitectura para la provisión de información semántica de transporte

La figura 5.1 muestra un esquema con la arquitectura lógica del sistema, definiendo los elementos principales que componen cada una de las capas funcionales (persistencia, acceso a datos y presentación) y la interacción entre los mismos.

## 5. Arquitectura del sistema

---

De esta manera, la arquitectura se puede dividir así en 3 capas funcionales horizontales, además de los dispositivos y aplicaciones cliente, usuarios de la misma. La descripción de cada uno de los componentes presentes en dichas capas se detalla a continuación:

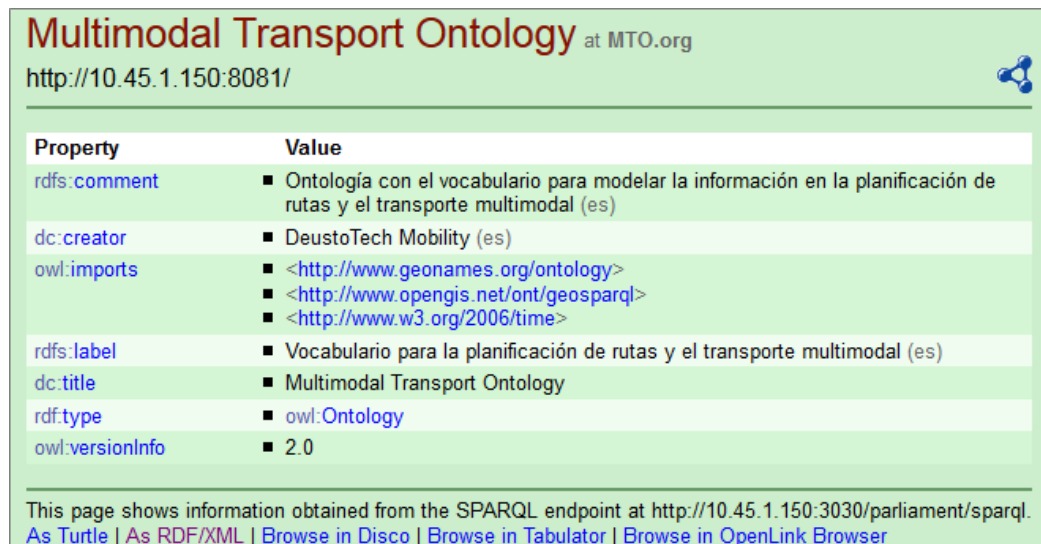
- ◇ *Client Devices & Applications*. Diferentes tipos de dispositivos o aplicaciones cliente que pueden solicitar la información de transporte proporcionada, como un navegador semántico, un navegador web, un planificador multimodal o un servidor de consultas SPARQL. Deberán tenerse en cuenta las capacidades y/o requisitos del dispositivo o aplicación cliente a la hora de proveer información relevante y en un formato adecuado.
- ◇ *Linked Data Interface*. Elemento encargado de realizar dicha negociación de contenidos, es decir, de proveer una interfaz entre el servidor SPARQL y el cliente, sirviendo contenido adecuado a cada aplicación: contenido RDF a navegadores y buscadores semánticos y HTML simple con la información de las entidades almacenadas a navegadores web convencionales.
- ◇ *SPARQL Servers*. Conjunto distribuido e interoperable de servidores SPARQL, con funcionalidades avanzadas de consulta y filtrado de la información y capacidad para devolver los datos solicitados en múltiples formatos abiertos. En este caso compatibles con el estándar GeoSPARQL para consultas geoespaciales y con el estándar SPARQL 1.1 que permite la federación de consultas y habilita el comportamiento distribuido.
- ◇ *Triple Store*. Base de datos especialmente diseñada para el almacenamiento y recuperación de contenido semántico estructurado en tripletas (entidad de datos basada en vocabularios ontológicos y compuesta de sujeto-predicado-objeto). Utilizado en este caso para gestionar las entidades e instancias de la ontología MTO y proveer de datos al servidor de consulta.

En las siguientes subsecciones se detalla la selección realizada de herramientas software para el despliegue de cada uno de los componentes de la arquitectura.

### 5.2.1 *Linked Data Interface* y servidor RDF

Se ha seleccionado el proyecto de código abierto *Pubby*<sup>1</sup> como herramienta para independizar la provisión de datos de transporte de las características técnicas del dispositivo o el cliente que solicite la información. *Pubby* facilita un interfaz web *Linked Data* a *endpoints* SPARQL ya existentes. Muchas bases de datos ontológicas y *endpoints* SPARQL sólo pueden ser accedidos mediante aplicaciones cliente que utilicen el protocolo SPARQL directamente, impidiendo así el acceso desde aplicaciones de usuario tradicionales no semánticas.

Uno de los objetivos perseguidos con la arquitectura propuesta es permitir el acceso universal a los datos de transporte multimodal disponibles. Este objetivo no sería adecuadamente satisfecho si la información fuese sólo accesible a través de consultas SPARQL o el uso de aplicaciones semánticas; es por ello por lo que el despliegue de una solución como *Pubby* se hace necesaria. El servidor desplegado se encargará de traducir peticiones HTTP mapeando cada entidad de la ontología con una URI y realizando las consultas SPARQL a la base de datos para después mostrar los datos recibidos en el formato correspondiente, realizando una negociación de contenidos entre HTML y RDF/XML. La figura 5.2 muestra el despliegue realizado para la provisión de información desde los servidores SPARQL de MTO.



Property	Value
<a href="#">rdfs:comment</a>	■ Ontología con el vocabulario para modelar la información en la planificación de rutas y el transporte multimodal (es)
<a href="#">dc:creator</a>	■ DeustoTech Mobility (es)
<a href="#">owl:imports</a>	■ <a href="http://www.geonames.org/ontology">http://www.geonames.org/ontology</a> ■ <a href="http://www.opengis.net/ont/geosparql">http://www.opengis.net/ont/geosparql</a> ■ <a href="http://www.w3.org/2006/time">http://www.w3.org/2006/time</a>
<a href="#">rdfs:label</a>	■ Vocabulario para la planificación de rutas y el transporte multimodal (es)
<a href="#">dc:title</a>	■ Multimodal Transport Ontology
<a href="#">rdf:type</a>	■ <a href="#">owl:Ontology</a>
<a href="#">owl:versionInfo</a>	■ 2.0

This page shows information obtained from the SPARQL endpoint at <http://10.45.1.150:3030/parliament/sparql>.  
[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

**Figura 5.2:** Despliegue de Pubby sobre los datos semánticos de transporte

<sup>1</sup> Pubby, <http://wifo5-03.informatik.uni-mannheim.de/pubby>

## 5. Arquitectura del sistema

---

### 5.2.2 *Triple store* y servidor SPARQL

La elección del componente software encargado de soportar el almacenamiento y la consulta de los datos semánticos generados se ha basado en el adecuado cumplimiento de dos características fundamentales, requisito indispensable de la arquitectura propuesta: la capacidad para resolver consultas geoespaciales, es decir, soportar el estándar GeoSPARQL, y la posibilidad de atender a consultas federadas, un concepto que se describirá con detalle en la sección 5.3.2 de este capítulo.

Ambas características, junto con otras también beneficiosas en el modelado del dominio de transporte, son implementadas por el proyecto *Parliament*<sup>2</sup>, una base de datos de contenido ontológico y servidor SPARQL diseñado para la web semántica y compatible con RDF, RDFS, OWL, SPARQL, y las normas GeoSPARQL. Fue lanzado como proyecto de código abierto en junio de 2009 y ofrece interesantes características altamente indicadas para el escenario del transporte:

- ◇ Emplea un innovador sistema de almacenamiento de datos que entrelaza los datos con un índice único. Debido a esto, puede responder a las consultas de manera eficiente, reordenando la ejecución de la consulta para que las partes más restrictivas se ejecuten primero. Esta es una característica valiosa para la solución propuesta debido a la ejecución de consultas complejas relacionadas con datos geoespaciales y contextuales y atendiendo a múltiples variables.
- ◇ Soporta GeoSPARQL, norma aprobada en 2012 por el OGC (*Open Geospatial Consortium*) para la modelización de datos semánticos con contenido geoespacial [OGC 12]. Del mismo modo en que el conjunto de vocabularios semánticos y datos enlazados ha venido aumentando considerablemente durante los últimos años, también lo ha hecho la cantidad de información con un inherente contexto espacial. Sin un razonamiento espacial específico, sin embargo, el valor de esta información queda completamente limitado.

*Parliament* aborda esta cuestión con la implementación de un índice geoespacial dedicado, compatible con el estándar definido y asociado a las tripletas con datos geográficos, mediante el que responde de manera eficiente a este tipo de consultas. *p. ej. "listar elementos situados dentro de la región X"*.

---

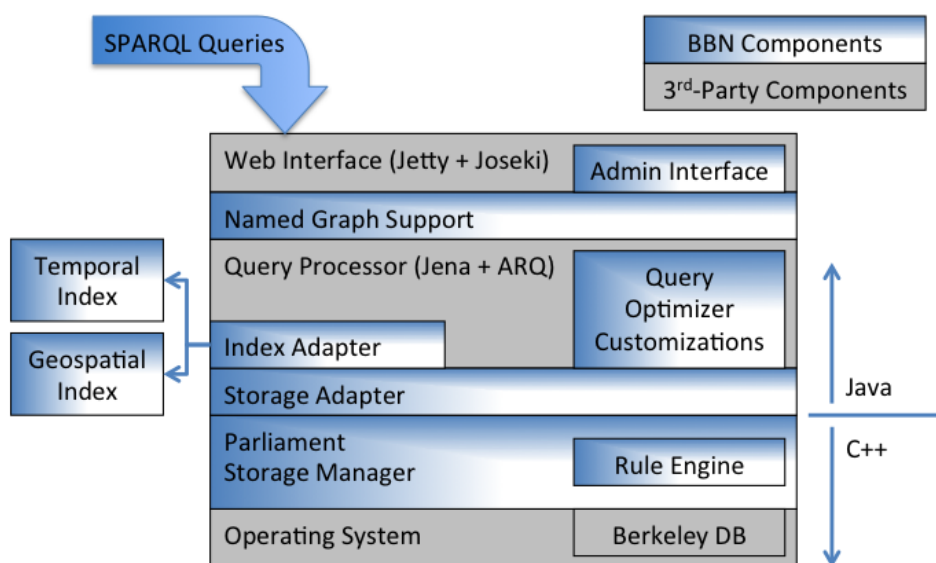
<sup>2</sup> Parliament, <http://parliament.semwebcentral.org>



## 5.2 Diseño y componentes software

- ◇ Tiene un índice temporal, análogo en cuanto a funcionalidad al anterior, de manera que puede responder eficientemente búsquedas vinculadas a intervalos de tiempo. *p. ej. "listar todos los eventos producidos entre X e Y fechas"*.
- ◇ Incluye un motor de reglas semánticas de alto rendimiento. Esto permite inferir automáticamente y de forma transparente hechos y relaciones adicionales a los datos, enriqueciendo los resultados de la consulta. Implementa inferencias RDFS junto con algunos elementos de OWL.

*Parliament*, como proyecto que engloba un almacén para datos semánticos, un procesador avanzado de consultas y un servidor web encargado de publicar *end-points* SPARQL para el acceso a datos, está basado en varios proyectos de código abierto que estructuran el sistema. La figura 5.3 obtenida de la documentación del proyecto muestra la arquitectura y el conjunto de servicios y herramientas software que lo componen.



**Figura 5.3:** Parliament: arquitectura software y componentes

La arquitectura y las características descritas, así como el soporte a la directiva *Service* y la federación de consultas, novedades de la normativa SPARQL 1.1<sup>3</sup> de 2013 hacen de este servidor de información semántica una opción óptima para el almacenamiento y la gestión de los datos de transporte disponibles.

<sup>3</sup> SPARQL 1.1, <http://www.w3.org/TR/sparql11-query>

### 5.3 Integración de datos de transporte distribuidos

Tal y como se ha indicado a lo largo de este capítulo, se ha establecido la utilización de una arquitectura distribuida, basada en la interoperabilidad y la colaboración de varios servidores SPARQL interrelacionados, y enfocada a la publicación de los datos como LOD, garante del correcto funcionamiento del sistema bajo condiciones de gran volumen de información como las producidas en escenarios de actuación reales en el ámbito del transporte multimodal.

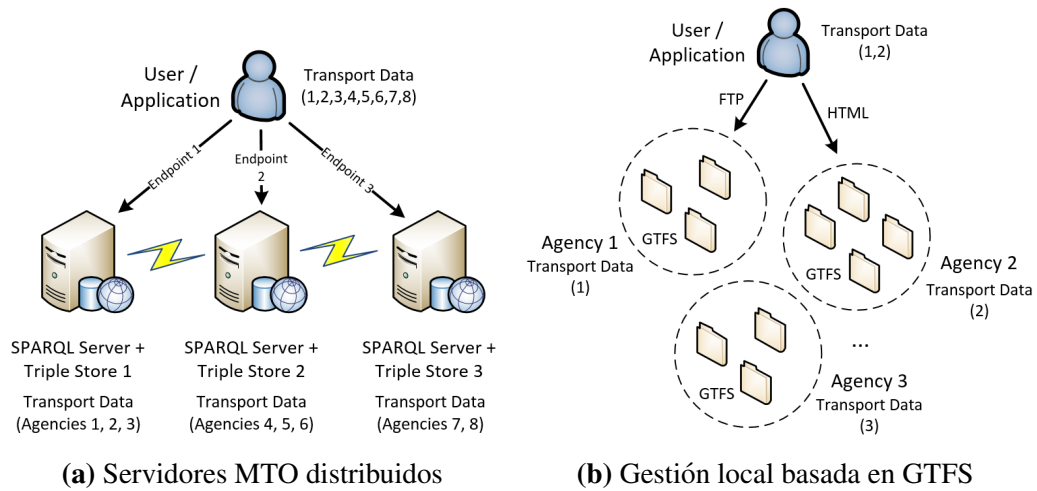
#### 5.3.1 Motivaciones y diseño

La motivación principal que ha llevado al planteamiento distribuido descrito es abordar la problemática existente en cuanto al tratamiento de grandes volúmenes de información ontológica mediante las herramientas y protocolos de la web semántica, teniendo en cuenta además que la validación y el despliegue de la arquitectura propuesta en un escenario real conllevará enfrentarse directamente con dicha limitación. Varios de los autores citados en el capítulo 2, [Zhao 08, De Oliveira 13], ponen también de manifiesto dicho problema. Gunay, en un desarrollo similar al propuesto para la semantización de la información de tránsito [Gunay 14], llega a recomendar la utilización de técnicas distribuidas como posible solución al mismo.

La arquitectura diseñada posibilita la compartición de los datos de transporte, almacenados en bases de datos semánticas o *Triple Stores* independientes. Dichos datos (en formato MTO) son accedidos por el servidor SPARQL correspondiente (habitualmente desplegado en la misma máquina) que, haciendo uso de sus capacidades de agregación de datos, será el encargado de integrarlos junto con otros provenientes de servidores afines remotos, distribuyendo así la carga de procesamiento entre varios servidores que colaboran entre sí (ver figura 5.4a). Se puede definir de este modo una integración de la información a nivel de acceso a datos, tal y como se mostraba en el diagrama de la figura 5.1.

Cada uno de los servidores SPARQL desplegados mantendrá así información de transporte propia y será administrado localmente, pero facilitará la interoperabilidad gracias a su interconexión con el resto de servidores del sistema. Dicha funcionalidad se consigue gracias a la gestión dinámica realizada por cada servidor de las direcciones IP de los *endpoints* remotos desplegados, habilitando de este modo un mecanismo para la realización de consultas distribuidas transparentes.

### 5.3 Integración de datos de transporte distribuidos



**Figura 5.4:** Arquitecturas para la provisión de información de transporte

En cuanto a la publicación y el acceso a los datos, la figura 5.4 establece una comparativa entre el modelo utilizado en la arquitectura propuesta y el mayoritariamente utilizado en la actualidad, basado en la compartición de ficheros GTFS.

Es destacable reseñar, tal y como se ha apuntado anteriormente, que no existe un sistema estandarizado para compartir la información GTFS. La publicación de los datos de transporte por parte de las agencias se realiza en muchos casos simplemente alojando los ficheros en un servidor web o FTP accesible desde Internet. Esto conlleva evidentes problemas a la hora de gestionar dicha información. Así, el usuario deberá conocer de antemano la ubicación de los ficheros y su estado (si se encuentran actualizados o no) para después integrarlos manualmente, si quisiera operar con información de varias agencias de transporte.

En la figura 5.4b puede observarse, a modo de ejemplo, cómo un usuario accede por distintas vías a la información de transporte público de varias agencias. Sin embargo, puede existir más información disponible (Agencias 4,5,6,7,8) de la cual no tiene constancia, o información conocida (Agencia 3) a la que no tiene acceso en ese momento. Así, este modelo de gestión local de los datos, no integrado, está favoreciendo la creación de silos de información con datos desactualizados, aislados y no interoperables.

En lo que respecta a la arquitectura distribuida propuesta, representada en la figura 5.4a, la información se almacena en servidores gestionados localmente de cara a mantener los datos actualizados. Dichos servidores se encuentran interrela-

## 5. Arquitectura del sistema

---

cionados, distribuyendo los datos disponibles y optimizando la carga de proceso, al encargarse cada servidor de parte de la consulta cuando ésta afecta alguno de sus datos locales. La adopción de este modelo permite que desde cualquier servidor se acceda a toda la información disponible, integrando los datos y gestionando el direccionamiento de forma transparente para el usuario final.

Cuando un usuario y/o aplicación realice una petición al sistema (p. ej. acerca de una ruta intermodal) dirigida a uno de los servidores SPARQL disponibles, este facilitará la misma consulta a los servidores que tiene registrados (realizando una búsqueda federada hacia cada uno de ellos). Dichos servidores efectuarán la misma operación (pudiéndose establecer una organización jerárquica) y así sucesivamente hasta que la consulta sea atendida. Una vez los datos son devueltos al servidor principal, éste realiza la agregación de los mismos y los presenta al usuario. La arquitectura es similar a la utilizada por DNS para resolver nombres de dominio en Internet, facilitando la distribución de la información de forma transparente.

En el caso de la información de transporte multimodal, ámbito concreto de aplicación de la arquitectura desarrollada, se ha realizado un diseño jerárquico en base a elementos geopolíticos o administrativos. Así, se ha establecido el despliegue de un *endpoint* SPARQL para cada provincia de la cual se dispongan datos de transporte multimodal. A partir de ahí se define una organización jerárquica donde se establece un *endpoint* por cada comunidad autónoma, enlazando los servidores anteriormente descritos, y así sucesivamente siguiendo la estructura administrativa correspondiente de abajo a arriba.

### 5.3.2 Federación de consultas SPARQL

La arquitectura diseñada establece la distribución de la información de transporte a nivel de acceso a datos, o dicho de otra manera, delega en los servidores SPARQL, encargados de procesar las consultas y acceder a los *Triple Stores* la gestión de la compartición de los datos y de este modo la distribución de los mismos. Cada uno de los servidores desplegados facilitará la interoperabilidad gracias a su interconexión con el resto de servidores que compongan el sistema. Dicha funcionalidad se consigue gracias a la llamada *federación de consultas*, especificación introducida en la versión 1.1 de SPARQL por el W3C con la idea de habilitar un mecanismo para la realización de consultas distribuidas entre servidores remotos.

### 5.3 Integración de datos de transporte distribuidos

La especificación desarrollada define la sintaxis y la semántica a utilizar para ejecutar consultas distribuidas sobre diferentes *endpoints* SPARQL. La palabra clave *Service*, extensión de SPARQL 1.1, será la utilizada para soportar la ejecución de consultas remotas, fusionando datos distribuidos a través de la web.

El fragmento de código 5.3 muestra un ejemplo de consulta en el que se utiliza la directiva *Service* para completar la consulta al servidor local con información remota proporcionada por otro servidor SPARQL en <http://people.example.org>

```
1 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
2 SELECT ?me ?name
3 WHERE {
4   ?me foaf:knows ?person.
5   SERVICE <http://people.example.org/sparql> {
6     ?person foaf:name ?name. }
7 }
```

**Código 5.3:** Consulta SPARQL. Ejemplo directiva remota *Service*

La arquitectura desarrollada hace un uso extensivo de esta directiva para soportar la agregación de información de transporte distribuida sobre varios servidores interoperables organizados jerárquicamente. Para ello, se han definido un conjunto de entidades dentro de la ontología MTO (bajo la ruta *fed* de *federated*) denominadas *Datasets*. Cada entidad *Dataset* contiene información dinámica acerca de un *endpoint* SPARQL remoto asociado con el servidor actual, incluyendo su URL. Esta información permitirá realizar consultas remotas hacia dicho servidor, habilitando así el funcionamiento distribuido del sistema.

El fragmento de código 5.4 muestra la descripción del *Dataset* para la región de Araba, incluyendo el enlace a su *endpoint* en la línea 4.

```
1 <rdf:RDF xmlns:mtof="http://www.mto.org/fed/">
2   <rdf:Description rdf:about="mtof:Dataset/araba">
3     <rdf:type rdf:resource="mtof:Dataset"/>
4     <mtof:endpoint rdf:resource="http://localhost:3030/sparql"/>
5     <mtof:webview rdf:resource="http://localhost:8081"/>
6   </rdf:Description>
7 </rdf:RDF>
```

**Código 5.4:** Descripción de la entidad *Dataset*. URL al *endpoint* SPARQL

## 5. Arquitectura del sistema

---

Tal y como se ha descrito anteriormente, se ha definido una organización jerárquica asociada al nivel geopolítico administrativo, permitiendo distribuir la información multimodal en servidores SPARQL clasificados por área geográfica. Así, el servidor asociado a la Comunidad Autónoma Vasca contendrá la información con los *Datasets* de las 3 provincias vascas: Araba, Bizkaia y Gipuzkoa.

Una vez que se han establecido las referencias, las consultas al servidor autonómico se distribuirán en consecuencia, y la información relacionada con un área específica se recogerá de forma distribuida y transparente. El fragmento 5.5 muestra un ejemplo de consulta federada que recibe toda la información sobre las paradas (*Stops*) contenidas en los servidores SPARQL asociados.

Para ello se ha usado la variable *?ep* (línea 9) que representa un *mtof:endpoint* como el descrito anteriormente en 5.4. Esta solución permite establecer las URLs de los distintos *endpoints* como argumentos de la directiva *Service* (línea 10), a diferencia de la URL estática del ejemplo 5.3, permitiendo la generación dinámica de consultas remotas en base a metadatos de la propia ontología. El servidor iterará a través de todos los *endpoints* provistos y, utilizando el protocolo *Service*, llevará a cabo consultas remotas a todos ellos, agregando sus datos y presentando la información al usuario final de forma integrada.

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX mto: <http://www.mto.org/mto/>
4 PREFIX gn: <http://www.geonames.org/ontology#>
5 PREFIX mtof: <http://www.mto.org/fed/>
6
7 SELECT ?stop_id ?stop_name ?stop_lat ?stop_lon
8 WHERE {
9   ?ds mtof:endpoint ?ep.
10  SERVICE ?ep {
11    ?stop_id a mto:Stop.
12    ?stop_id mto:hasName ?stop_name.
13    ?stop_id mto:hasLatitude ?stop_lat.
14    ?stop_id mto:hasLongitude ?stop_lon.
15  }
16 } ORDER BY ?stop_id
```

**Código 5.5:** Consulta federada. Listado de paradas

### 5.3 Integración de datos de transporte distribuidos

El resultado de dicha consulta, realizada sobre el servidor autonómico que da soporte a la CAV, es un conjunto de casi 5000 paradas (*Stops*) de transporte público: 1150 de Araba, 2700 de Bizkaia y 790 de Gipuzkoa, atendiendo a los datos disponibles provistos por las agencias de transporte y cargados en cada servidor provincial (en el listado 5.6 se muestran a modo de ejemplo las 10 primeras).

La consulta al *endpoint* SPARQL autonómico se resuelve así de forma distribuida, agregando los datos provistos por los servidores de cada provincia, en un tiempo total de aproximadamente 5 segundos.

```
1 stop_id, stop_name, stop_lat, stop_lon
2 http://www.mto.org/stop/568/1106, "Orixe 54", 43.2837, -2.9625,
3 http://www.mto.org/stop/568/1201, "Elorrieta", 43.2840, -2.9654,
4 http://www.mto.org/stop/568/1302, "Don Bosco", 43.2733, -2.9553,
5 http://www.mto.org/stop/568/1406, "Euskalduna", 43.2668, -2.9499,
6 http://www.mto.org/stop/568/1409, "Buenavista", 43.2739, -2.9432,
7 http://www.mto.org/stop/568/1418, "Madariaga", 43.2707, -2.9568,
8 http://www.mto.org/stop/568/1423, "San Pedro", 43.2719, -2.9450,
9 http://www.mto.org/stop/568/1511, "Zorrotzaurre", 43.2746, -2.9670,
10 http://www.mto.org/stop/568/1601, "Araneko 17", 43.2740, -2.9484,
```

**Código 5.6:** Listado de paradas (*Stops*). Se muestran las 10 primeras





*El experimentador que no sabe lo  
que está buscando no comprenderá  
lo que encuentra.*

Claude Bernard

# 6

## Experimentación y resultados

Una de las etapas fundamentales de una metodología de investigación es la experimentación, necesaria para realizar una validación rigurosa de la hipótesis planteada. Así, la naturaleza de la hipótesis y del propio trabajo de investigación determinará el empleo de unos medios de validación u otros. En muchos casos, la realización de experimentos, considerando diversas modalidades (con o sin participación de los usuarios, basados o no en análisis estadísticos, formales o no formales, en laboratorio o en entorno real, etc.), constituirá una opción adecuada conducente a dicha validación formal.

El propósito del presente capítulo es, precisamente, mostrar los instrumentos que han sido empleados para la validación de la hipótesis inicialmente planteada en esta tesis doctoral, justificando las estrategias de verificación seleccionadas y describiendo los desarrollos y experimentos efectuados con tal fin. Una vez presentado el escenario de pruebas a utilizar y ejecutada la citada experimentación sobre él, se procederá a realizar una evaluación crítica de los resultados obtenidos, atendiendo al cumplimiento de los objetivos y concluyendo así la veracidad o no de la hipótesis enunciada.

### 6.1 Estrategia de validación de la hipótesis

Dado que el conjunto de aportaciones realizadas en esta tesis se adscriben al diseño y desarrollo de elementos software en el ámbito de los ITS, se considera que la opción de validación de la hipótesis más adecuada es la experimentación funcional en laboratorio, asegurando la utilización de datos de transporte multimodal provenientes de entornos de actuación reales en dicho ámbito.

Los experimentos en laboratorio constituyen una técnica cuantitativa aplicada para comprobar el comportamiento de un determinado fenómeno bajo unas condiciones establecidas. Su principal función es verificar una teoría o hipótesis o bien deducir nuevo conocimiento a partir de una muestra particular extraída en unas condiciones específicas.

Con el objetivo de definir en detalle cómo se va a proceder a la experimentación, se muestra nuevamente la hipótesis planteada en el capítulo 1, alrededor de la cual gira la presente tesis.

«Es posible desarrollar una arquitectura software que, partiendo de un conjunto heterogéneo de fuentes de datos de transporte e incorporando información relacionada relevante, sea capaz de generar información de transporte multimodal estructurada e integrada, habilitando de este modo la creación de servicios avanzados de consulta mediante la publicación distribuida e interoperable de dicha información»

La hipótesis presentada se puede disgregar en dos ideas principales. Por un lado, ésta enuncia que es posible desarrollar un procedimiento capaz de generar información de transporte multimodal estructurada e integrada, enlazando datos desde diversas fuentes y agregando información relevante relacionada.

Durante el capítulo 4 se ha visto cómo mediante el diseño de un modelo ontológico para la representación de la información de transporte es posible extender, integrar y enlazar los datos de tránsito disponibles. En este capítulo, además, se pretenden validar dichas capacidades, directamente relacionadas con los objetivos de esta tesis doctoral, mediante la puesta en marcha de la arquitectura propuesta en un escenario de actuación real (con datos de transporte de la CAV).

## 6.2 Despliegue del sistema y escenario de pruebas

---

Con esta idea se acude a la segunda parte de la hipótesis, que sostiene que el despliegue de dicha arquitectura permitirá la creación de herramientas y servicios avanzados de consulta de información, basados en la publicación distribuida, abierta e interoperable de los datos semánticos de transporte generados.

En este caso, el detalle de dicha contribución, en forma de arquitectura distribuida basada en LOD para el despliegue de servidores semánticos interoperables, se puede encontrar en el capítulo 5 del presente trabajo. En este capítulo, se quiere ir más allá, demostrando la validez de esta premisa mediante el desarrollo de un planificador multimodal semántico como herramienta avanzada de información al viajero, análogo a los descritos en el capítulo 2 y basado en la gestión de los datos de tránsito provistos por la arquitectura desplegada. Esto permitirá evaluar la funcionalidad, el rendimiento y la adaptabilidad de dicha solución ante herramientas complejas de consulta de información, como el citado planificador multimodal.

Los resultados obtenidos tras el despliegue y ejecución de dicha experimentación servirán para validar si las aportaciones realizadas resuelven o no las limitaciones encontradas, cumpliendo con los objetivos iniciales fijados y demostrando así la validez de la hipótesis.

## 6.2 Despliegue del sistema y escenario de pruebas

A través del despliegue del sistema propuesto corriendo bajo circunstancias preestablecidas, se evaluará el grado de cumplimiento de los objetivos inicialmente planteados. De acuerdo con las particularidades del ámbito de actuación, en este caso se propone utilizar información de tránsito actual de la CAV.

Se establecen para ello las siguientes fases, cuyo desarrollo se describirá en las siguientes subsecciones y finalizará con la puesta en marcha de la arquitectura:

1. La obtención y clasificación de los datos de transporte a utilizar en la experimentación, así como la organización de servidores de acuerdo al diseño distribuido propuesto.
2. La carga de los datos de transporte disponibles y la generación del contenido semántico correspondiente de acuerdo con la ontología MTO desarrollada.
3. La puesta en marcha de 3 servidores semánticos, publicando como LOD la información de transporte multimodal referente a las 3 provincias de la CAV.

## 6. Experimentación y resultados

### 6.2.1 Presentación del escenario de pruebas

En este caso se utilizarán los datos de tránsito provistos por las agencias de transporte que operan dentro de la CAV como base sobre la que desplegar la solución. A nivel arquitectónico, se ha establecido un despliegue distribuido de servidores interoperables basado en el área geográfica de actuación. En concreto se ha realizado una división administrativa a nivel provincial, de modo que la solución a desplegar contará con 3 servidores, gestionando la información de cada una de las 3 provincias vascas respectivamente.

Las tablas 6.1 y 6.2 muestran a modo de referencia información acerca del conjunto de datos utilizado en la experimentación.

Serv. Araba		Serv. Bizkaia		Serv. Gipuzkoa	
Agencia	Modo	Agencia	Modo	Agencia	Modo
Alavabus	Bus	Bilbobus	Bus	DonostiaBus	Bus
La Union	Bus	BizkaiBus	Bus	La Burundesa	Bus
Tuvisa	Bus, Tran	Euskotren	Tren, Tran	Renfe	Tren
		MetroBilbao	Metro	Pesa	Bus

**Tabla 6.1:** Conjunto de datos de tránsito. Distribución por provincias y modos

La tabla 6.1 presenta en primer lugar la distribución provincial realizada a nivel de agencias de transporte para la gestión de los datos disponibles, así como los modos de transporte provistos por cada una de ellas.

Se gestionan en total datos de tránsito de 11 agencias de transporte que operan en la CAV. Algunas de estas agencias, como Euskotren o Renfe, gestionan información referente a varias provincias ya que realizan recorridos intra e interprovinciales. Con objeto de no duplicar la información, cada una de estas agencias ha sido asignada a la provincia desde la que opera preferentemente, atendiendo también a criterios de balanceo de carga entre los servidores.

Puede observarse cómo se provee información acerca de 4 modos de transporte (autobús, tren, tranvía y metro) aunque no todos se encuentran disponibles para todas las provincias. El único modo de transporte presente en las 3 provincias es el autobús con un porcentaje de aparición del 68,2 % en relación al total de agencias.

## 6.2 Despliegue del sistema y escenario de pruebas

Por otro lado, la tabla 6.2 presenta un análisis respecto a la carga de procesamiento que deberá ser asumida por cada servidor de acuerdo a la arquitectura distribuida planteada. Se muestra a modo de resumen el número de agencias, rutas, paradas y horarios gestionados por cada servidor provincial.

Servidor	Agencias	Rutas	Paradas	Horarios
Araba	3	144	866	142.590
Bizkaia	4	155	2.962	410.863
Gipuzkoa	4	67	1.170	763.114
Total	11	366	4.998	1.316.567

**Tabla 6.2:** Conjunto de datos de tránsito. Distribución de la carga de proceso

Se puede observar una distribución uniforme en cuanto a agencias, sin embargo la diferencia en la operativa acorde al modo de transporte utilizado (p. ej. en el metro, con rutas y paradas fijas, pero con amplios horarios) hace que la divergencia en cuanto a rutas, paradas y horarios sea en algunos casos sustancial. Se estima en todo caso que la distribución de carga entre servidores derivada de esta organización provincial, aún con dichas diferencias, permitirá evaluar de manera veraz la funcionalidad y el rendimiento de la arquitectura desplegada.

### 6.2.2 Carga de datos y generación del contenido ontológico

Los datos de tránsito de las agencias de transporte de la CAV, en formato GTFS, utilizados para generar la información semántica que almacenará la ontología desarrollada, se han obtenido del servidor FTP geoEuskadi<sup>1</sup>, iniciativa pública del Gobierno Vasco para el acceso a la información geográfica.

Partiendo de dicho conjunto de datos y utilizando el adaptador desarrollado para la conversión desde el formato GTFS (descrito en la sección 4.2.3) se ha obtenido un *dataset* completo de transporte público para la CAV, incluyendo información geopolítica administrativa y datos geoespaciales mediante el enlace de la ontología con otros conjuntos de datos, habilitando la consulta de información semántica de transporte multimodal en base a dichas características geográficas.

<sup>1</sup> FTP geoEuskadi, <ftp://ftp.geo.euskadi.net/cartografia/>

## 6. Experimentación y resultados

El proceso de carga de los ficheros GTFS y generación del contenido ontológico se encuentra completamente automatizado. La herramienta provista realiza el proceso de validación de los ficheros fuente y efectúa la conversión incluyendo las referencias semánticas a los vocabularios externos. La tabla 6.3 muestra a modo de comparativa el detalle en cuanto a tamaños y tiempos del proceso de transformación o semantización de la información para el total de agencias de transporte utilizadas en el escenario de pruebas propuesto.

	<b>Tamaño GTFS<sup>2</sup></b>	<b><i>Stops</i></b>	<b><i>Schedules</i></b>	<b>Tamaño OWL<sup>2</sup></b>	<b>Ratio</b>	<b>Tiempo Total<sup>2</sup></b>
<b>Alavabus</b>	38	353	4.493	206	-	-
<b>Bilbobus</b>	546	561	96.337	3.343	6,12	88,3
<b>Bizkaibus</b>	516	2.137	62.163	2.715	5,26	170,9
<b>Euskotren</b>	502	80	107.040	3.505	6,98	71,6
<b>La Union</b>	25	162	1.917	89	-	-
<b>La Burundesa</b>	33	181	3.503	102	-	-
<b>Metro Bilbao</b>	1.197	184	145.323	5.551	4,64	97,0
<b>Renfe</b>	1.250	73	203.370	7.603	6,08	116,9
<b>Tuvisa</b>	842	351	136.180	4.523	5,37	92,8
<b>Pesa</b>	1.042	402	251.275	4.350	7,37	98,8
<b>DonostiaBus</b>	3.243	514	304.966	15.633	-	-
<b>Media</b>	839,45	463,45	119.687,91	4.329,09	5,97	105,19

**Tabla 6.3:** Proceso de semantización de la información de transporte

La tabla muestra una serie de aspectos relevantes a la hora de plantear la semantización de la información de transporte ya disponible. En cuanto al espacio necesario para almacenar la información semántica, se observa un ratio medio relativo de 5,97 veces el espacio ocupado por los ficheros GTFS originales. Este ratio se considera ajustado ya que supone un tamaño medio real de los ficheros OWL de 4,23 MB, tamaño completamente asumible y que además supone la provisión de información enlazada y, por tanto, más completa.

<sup>2</sup> Tamaño de los ficheros GTFS y OWL en Kilobytes (KB). Tiempo en segundos (s)

## 6.2 Despliegue del sistema y escenario de pruebas

---

Por otro lado, en cuanto al tiempo necesario para efectuar la conversión, se observa una duración promedio del algoritmo de 105,2 segundos. Se valora como una duración aceptable teniendo en cuenta la carga totalmente automatizada de los datos y la generación de las instancias semánticas correspondientes, una de las problemáticas más frecuentes, como se indica en [Boucher 09, Agarwal 05], en la adopción de ontologías como mecanismo para la representación de la información. Cabe destacar que el tiempo necesario para la conversión se ve afectado mayoritariamente, tal y como se puede observar en la tabla, por el número de *Stops* contenidos en el fichero CSV. Esto es debido al enlace realizado para cada *Stop* con las ontologías geográficas descritas en la sección 4.2.2.1.

La utilización de dicho conjunto de datos, utilizado también en herramientas como *Moveuskadi* (ver sección 2.2.3) para la generación del contenido ontológico, permitirá testar el funcionamiento de la arquitectura bajo un escenario real de uso de datos de transporte multimodal en un área lo suficientemente amplia como para poder validar la arquitectura de forma eficaz.

### 6.2.3 Despliegue de servidores distribuidos y publicación de la información semántica

Una vez realizado el proceso de semantización del conjunto de datos de prueba de acuerdo con la ontología MTO, se procede a desplegar la arquitectura según el escenario definido anteriormente, esto es, mediante la puesta en marcha de 3 servidores semánticos publicando como LOD la información ontológica generada. Se ha seguido la distribución de agencias de transporte por provincias definida en la tabla 6.1. Esta organización distribuida de los datos de tránsito disponibles permitirá evaluar el correcto funcionamiento y las capacidades de la arquitectura diseñada, estableciendo la web semántica como infraestructura sobre la que construir el mecanismo de provisión y consulta de información.

Además de asegurar el correcto funcionamiento de la arquitectura comprobando la publicación efectiva de datos de tránsito distribuidos, se ha querido evaluar a su vez el rendimiento en cuanto a la gestión de grandes volúmenes de información, problemática común en la representación de información semántica. Así, partiendo del despliegue realizado con los datos de transporte de la CAV se ha realizado un análisis de carga con el que medir el rendimiento global del sistema.

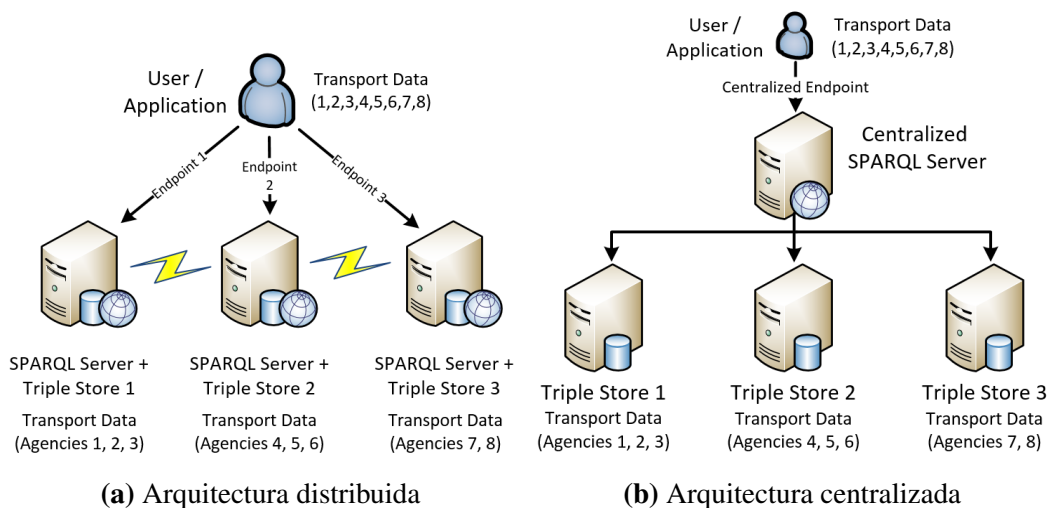
## 6. Experimentación y resultados

La tabla 6.4 establece una comparativa mostrando el número de entidades y el tiempo utilizado en la ejecución de consultas SPARQL sobre la ontología MTO para cada uno de los servidores provinciales desplegados.

	Agencias		Paradas		Horarios	
	Número	Tiempo	Número	Tiempo	Número	Tiempo
<b>Araba</b>	3	0,2 s	866	2,5 s	142.590	25,0 s
<b>Bizkaia</b>	4	0,2 s	2.962	5,1 s	410.863	45,7 s
<b>Gipuzkoa</b>	4	0,2 s	1.170	3,8 s	763.114	80,6 s
<b>Centralizado</b>	11	0,2 s	4.998	10,1 s	1.316.567	148,6 s
<b>Distribuido</b>	11	0,3 s	4.998	5,2 s	1.316.567	83,5 s
<b>Diferencia</b>		+0,1 s		-4,9 s		-65,1 s

**Tabla 6.4:** Comparativa de arquitecturas: modelos centralizado y distribuido

De cara a la realización de la comparativa se ha desplegado tanto un modelo semántico distribuido como un modelo centralizado, donde un único servidor gestiona la información de los tres almacenes de datos disponibles. La tabla muestra tanto el total como la diferencia de tiempo utilizado por cada uno de ellos en la resolución de consultas afectando el total de datos disponibles.



**Figura 6.1:** Arquitecturas para la provisión de información semántica



### 6.3 STP: Planificador Semántico de Transporte

---

El modelo distribuido (ver figura 6.1a), utilizado para el despliegue del escenario de pruebas ha sido descrito durante el capítulo 5, estableciendo un modelo en el que varios servidores interrelacionados colaboran en la ejecución de consultas y proveen información de transporte integrada. El modelo centralizado (ver figura 6.1b) despliega por su parte un único servidor SPARQL gestionando las consultas mediante el acceso consecutivo a los 3 almacenes provinciales.

Como puede observarse en la tabla 6.4, ante un aumento en el número de entidades y, por ende, de la carga de procesamiento (p. ej. en la consulta de horarios), la diferencia en cuanto a tiempo de proceso entre la utilización de un modelo distribuido y un modelo centralizado aumenta notablemente. Este hecho reafirma la idoneidad de un modelo como el propuesto en escenarios de actuación reales, en los que el volumen de información de transporte haría muy difícil la adopción de un modelo centralizado.

### 6.3 STP: Planificador Semántico de Transporte

Del análisis efectuado del estado del arte se puede extraer que una de las herramientas software más avanzadas en el ámbito de los ITS y los sistemas de ayuda al viajero, particularmente en lo referente a la gestión de información de transporte y su representación, otorgando un valor añadido al usuario en cuanto a la planificación y la ejecución de sus viajes, son los llamados planificadores multimodales.

Un planificador de rutas multimodal es un sistema informático que puede proporcionar al viajero de forma automatizada un itinerario de transporte optimizado para la ejecución de sus desplazamientos. Estas soluciones software, ya analizadas en detalle en la sección 2.2, utilizan las TICs para recopilar, procesar y distribuir información actualizada al usuario sobre el tráfico, la duración del viaje, retrasos esperados, rutas alternativas, etc. dando a los viajeros la oportunidad de tomar decisiones informadas sobre sus viajes.

Para ello, tal y como se ha resaltado durante todo este trabajo, es condición indispensable que exista información ampliamente disponible sobre todos los modos de transporte y las posibilidades de su uso combinado. Es en este punto donde el trabajo realizado en esta tesis doctoral, definiendo una arquitectura integradora capaz de gestionar y proveer datos de tránsito de manera abierta e interoperable, cobra fundada importancia.

## 6. Experimentación y resultados

---

Se ha considerado por tanto adecuado, de cara a validar el desempeño de la arquitectura desarrollada, la puesta en marcha de un planificador de rutas semántico, basado en la integración de datos enlazados abiertos junto con información colaborativa, y funcionando de este modo sobre la arquitectura propuesta. Esto permitirá evaluar la funcionalidad, el rendimiento y la adaptabilidad de dicha solución ante herramientas complejas de consulta de información de tránsito. Dicho planificador multimodal semántico o STP (del inglés, *Semantic Trip Planner*) hará uso del modelo de datos diseñado y del conjunto de servidores distribuidos desplegados para ofrecer al usuario información relevante en relación al transporte multimodal.

Para la implementación de STP y los test funcionales posteriores se ha utilizado un PC de escritorio con un procesador de 64 bits y 4 núcleos corriendo a 2.66 GHz, junto con 8 GB de RAM. Como Sistema Operativo se ha utilizado Windows 7 ejecutando la maquina virtual de Java 7 también en su versión de 64 bits. En cuanto a los servidores semánticos distribuidos que gestionan la información de cada provincia, se ha dedicado un servidor de maquinas virtuales para su gestión, proveyendo de 3 máquinas idénticas, basadas en el mismo software (Windows 7 64 bits y Java 7) con CPUs corriendo a 2.5 GHz y 6 GB de RAM dedicados respectivamente.

A nivel general, para el despliegue de todos los componentes de la arquitectura (Parliament, Pubby y STP) es recomendable contar al menos con 4 GB de RAM, sin restricciones en cuanto a plataforma, teniendo en cuenta que la cantidad de memoria necesaria dependerá de manera directa de la cantidad de información semántica a gestionar, debiendo optar por un diseño distribuido (como el planteado para la CAV) en caso de necesidad. El rendimiento global de la solución en el entorno desplegado se expondrá durante el presente capítulo mediante diversas pruebas de consulta y acceso a la información semántica realizadas sobre STP.

### 6.3.1 Consideraciones iniciales: objetivos y alcance

El objetivo perseguido con el desarrollo del planificador semántico STP es ilustrar cómo la arquitectura planteada, accediendo a información semántica de transporte de forma distribuida e interoperable, puede incorporarse a soluciones avanzadas de consulta ya existentes. Se mostrará así la aplicabilidad real del desarrollo realizado, habilitando a su vez la generación de nuevas herramientas en base a la información provista.

### 6.3 STP: Planificador Semántico de Transporte

La herramienta deberá ser capaz de proveer información semántica de transporte geolocalizada, de acuerdo con el contexto y las preferencias del usuario, y representada gráficamente mediante una interfaz basada en el uso de mapas. Se hace necesario aclarar que para el proceso de búsqueda de rutas de transporte (algoritmia de planificación), intrínseco a este tipo de soluciones, la herramienta se apoyará en el uso de OpenTripPlanner (OTP), plataforma alternativa a Google Transit para la planificación de viajes de carácter multimodal.

Dicha plataforma, ya expuesta en el análisis previo realizado del estado del arte (ver sección 2.2.4) es un proyecto de software libre y código abierto bajo licencia LGPL (*GNU Lesser General Public License*) lo cual implica que puede ser modificado y distribuido libremente. Esta capacidad de libre modificación y acceso al código fuente ha permitido definir la herramienta OTP como base para el desarrollo de la solución semántica propuesta (STP, ver figura 6.2). Cabe destacar que esto ha conllevado una extensiva modificación de dicha herramienta, derivada de su necesaria adaptación a la arquitectura desplegada, tomando por ejemplo como fuente de datos los *endpoints* SPARQL habilitados a tal fin.

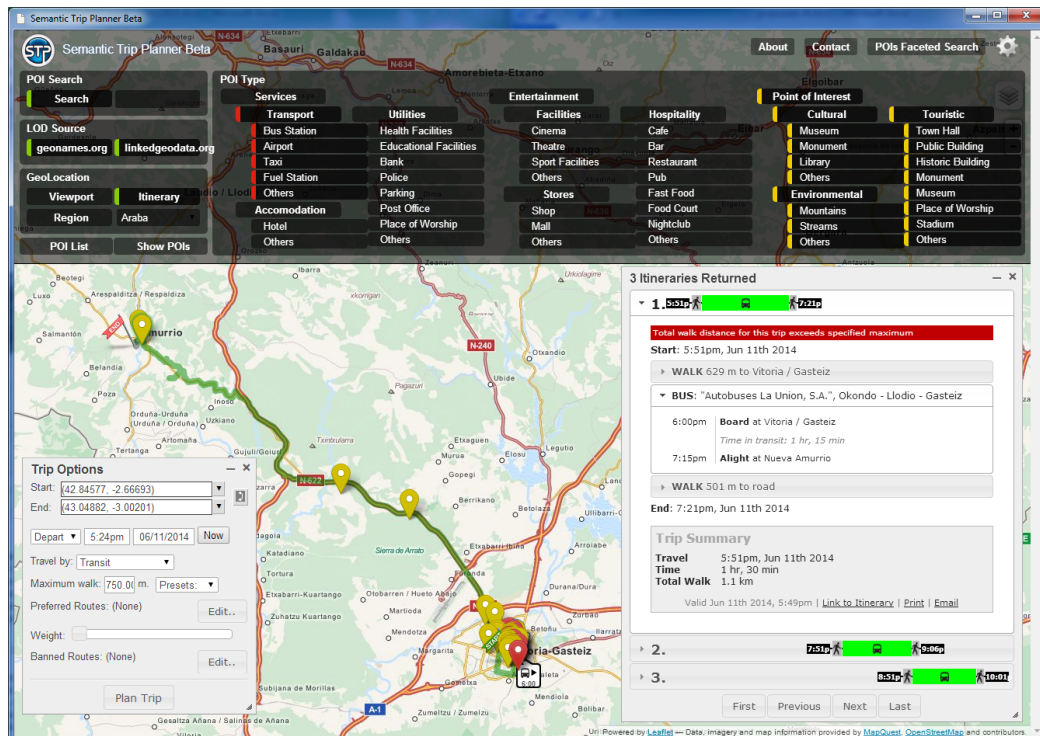


Figura 6.2: Interfaz web de STP (*Semantic Trip Planner*)

### 6.3.2 Modificación del proyecto OTP

La plataforma OTP proporciona un conjunto de servicios relacionados con la gestión y la provisión de información multimodal de tránsito. En relación con la planificación de rutas, proporciona varias interfaces web basadas en mapas así como una API REST destinada a su utilización por aplicaciones de terceros. Se basa en estándares abiertos incluyendo GTFS para datos de tránsito y OpenStreetMap (OSM) para el geoposicionamiento y la gestión de información cartográfica.

La solución presentada utilizará algunas de estas características, como la interfaz web (con las modificaciones pertinentes para adecuarla a las nuevas funcionalidades) y el motor de enrutamiento, de cara a proporcionar la información semántica generada. Así, para la implementación de STP, además del despliegue y la configuración inicial del proyecto OTP, aspecto fuera del alcance objetivo de esta tesis, se han llevado a cabo las siguientes acciones:

1. La carga de los datos de transporte publicados por el Gobierno Vasco y la generación del contenido semántico correspondiente (sección 6.2.2).
2. La puesta en marcha de 3 servidores SPARQL, publicando como LOD la información de transporte multimodal semantizada referente a las 3 provincias vascas (sección 6.2.3).
3. La modificación del proyecto (OTP) para que incluya los datos ofrecidos por la arquitectura desplegada:
  - (a) Consumiendo el modelo ontológico desarrollado (sección 6.3.2.1).
  - (b) Accediendo a los datos colaborativos existentes (sección 6.3.2.2).
  - (c) Proveyendo al usuario información contextual (sección 6.3.2.2).

Los primeros dos puntos ya se han abordado durante el presente capítulo, estableciendo el entorno de pruebas donde desplegar y validar la arquitectura. Por otra parte, el tercer punto hace referencia al desarrollo específico del planificador semántico, lo que incluye varias modificaciones sobre el proyecto OTP relacionadas con la naturaleza semántica de los nuevos datos disponibles. Dichas modificaciones son descritas con detalle en las siguientes subsecciones.

### 6.3.2.1 Configuración de las fuentes de datos

OTP trabaja con una estructura llamada *Graph* que contiene toda la información sobre la topología de la zona en la que opera el algoritmo de enrutamiento. Este grafo se obtiene de los datos de tránsito disponibles y su generación es la base del funcionamiento de la herramienta. Para proporcionar a ésta los datos sobre los que se generará el grafo se debe editar un archivo XML en el que se establecerá una propiedad apuntando a la ruta local de cada fichero GTFS que el sistema vaya a consumir. El fragmento de código 6.1 muestra un ejemplo.

```
1 <bean class="org.opentripplanner.model.GtfsBundles">
2   <property name="gtfsBundles">
3     <bean class="org.opentripplanner.model.GtfsBundle">
4       <property value="D:\STP\data\gtfs\Araba\alegria.zip"/>
5     </bean>
6     <bean class="org.opentripplanner.model.GtfsBundle">
7       <property value="D:\STP\data\gtfs\Araba\la_union.zip"/>
8     </bean>
9     <bean class="org.opentripplanner.model.GtfsBundle">
10      <propertyvalue="D:\STP\data\gtfs\Araba\vitoriana.zip"/>
11    </bean>
12    <bean class="org.opentripplanner.model.GtfsBundle">
13      <property value="D:\STP\data\gtfs\Araba\tuvisa.zip"/>
14    </bean>
15    ...
16  </property>
17 </bean>
```

**Código 6.1:** Fichero de configuración de OTP. Fuentes de datos GTFS

Esta operativa presenta ciertas carencias relacionadas con la gestión de las fuentes de datos, las cuales se han resuelto satisfactoriamente con el uso de la arquitectura propuesta al utilizar servidores SPARQL distribuidos como mecanismo para almacenar la información de tránsito. Por un lado, los archivos GTFS debían descargarse manualmente desde la fuente proporcionada por cada uno de los proveedores de información de tránsito que se quisiesen incluir (agencias, administraciones, etc.) y, una vez descargados, modificar el fichero XML de configuración para incluir la ruta concreta al fichero almacenado. En el fragmento 6.1, acortado para mostrar cuatro agencias, se puede observar esto en las líneas 4, 7, 10 y 13.

## 6. Experimentación y resultados

---

También hay que señalar que dicha información debía ser actualizada de forma manual y con regularidad por parte del administrador del planificador para mantener su precisión. Con el despliegue de STP en cambio, el archivo XML sólo almacena una dirección URL a un único servidor SPARQL (como se muestra en el fragmento de código 6.2) que, como se mencionó anteriormente, contiene una clasificación jerárquica del resto de servidores, pudiendo realizar así consultas distribuidas e integrando los datos bajo demanda.

```
1 <bean class="org.opentripplanner.model.SparqlEndpoints">
2   <property name="endpoints">
3     <bean class="org.opentripplanner.model.SparqlEndpoint">
4       <property value="http://localhost:3030/parliament/sparql"/>
5     </bean>
6   </property>
7 </bean>
```

**Código 6.2:** Fichero de configuración de STP. Fuente *endpoint* SPARQL

Cada servidor mantiene información sobre las agencias locales, por lo que la gestión y actualización de su contenido será responsabilidad de dichas agencias y se podrá hacer rápidamente a través de consultas *update* del propio lenguaje<sup>2</sup>. Este sistema, además de mejorar la precisión de la información provista, manteniéndola más actualizada, elimina la necesidad de almacenar o descargar archivos de datos de tránsito y limita al máximo la gestión de la configuración.

Con la finalidad de dar soporte a esta nueva operativa, ha sido necesaria la realización de una serie de cambios en el código de OTP, donde además de adaptar los ficheros de configuración, se ha generado una nueva clase heredando de *GraphBuilder*. Esta clase controla la forma en la que se procesa la información de tránsito para la construcción del grafo con la topología final. Con la intención de facilitar la integración se ha implementado una solución para la transformación de datos en formato OWL (procedentes de los servidores SPARQL) a datos CSV (como los provistos en GTFS), los cuales OTP puede procesar de forma nativa.

Se han elaborado un conjunto de consultas estandarizadas (una para cada archivo presente en el formato GTFS) que, partiendo de la información distribuida contenida en los *triple stores* semánticos, generan respuestas en formato CSV con toda la información necesaria, de manera directa y transparente.

---

<sup>2</sup> SPARQL 1.1 Update, <http://www.w3.org/TR/sparql11-update>

## 6.3 STP: Planificador Semántico de Transporte

---

El fragmento de código 5.1 presente en el capítulo 5 es un ejemplo de este tipo de consulta, en este caso obteniendo datos sobre las paradas en ruta y publicándolos en formato CSV. La utilización de dicho sistema, que hace uso de la flexibilidad de SPARQL y de su capacidad para generar respuestas en distintos formatos, ha permitido la rápida puesta en marcha de la solución de planificador semántico, aprovechando la algoritmia de enrutamiento provista por el proyecto OTP.

### 6.3.2.2 Gestión de contexto y vocabularios enlazados

Un aspecto importante a tener en cuenta de cara a generar información más precisa y rica es la utilización del contexto del usuario. Sin la participación de este componente contextual los datos ofrecidos podrían perder interés al no poder incorporar información personalizada. Del mismo modo, el uso de vocabularios enlazados, como GeoNames o LinkedGeoData, permite dotar a la solución de funcionalidades avanzadas de búsqueda, filtrado y consulta basadas en conceptos semánticos, inasumibles para las herramientas de planificación multimodal tradicionales.

Para cada consulta que llega a STP, el contexto del usuario es calculado y modelado con el fin de filtrar los datos de acuerdo a sus circunstancias específicas. La arquitectura define, con el fin de construir dicho modelo, una búsqueda facetada con un conjunto de propiedades opcionales referidas, de acuerdo al ejemplo de enriquecimiento ya descrito en la sección 4.3.1, a los POIs o puntos de interés en ruta. Dichas opciones de búsqueda y filtrado son:

- ◇ Nombre del punto de interés concreto a buscar
- ◇ Fuente de información enlazada a consultar (GeoNames y/o LinkedGeoData)
- ◇ Localización geográfica de la consulta
  - A nivel de organización administrativa (ej. POIs en Álava)
  - A nivel de ubicación geoespacial (ej. POIs a 5 km en ruta)
- ◇ Clasificación jerárquica del punto de interés (según *POIClassification*)
  - Servicios: Transporte, Alojamiento, Utilidades
  - Entretenimiento: Ocio, Tiendas, Hostelería
  - Puntos de interés: Cultural, Medioambiental, Turístico

## 6. Experimentación y resultados

A nivel técnico, en función del modelo de contexto proporcionado (proveniente de las selecciones realizadas) se generará dinámicamente una consulta SPARQL personalizada. La consulta puede combinar todos los parámetros enumerados anteriormente, como la geolocalización del usuario, el nombre del punto de interés o su clasificación, la proximidad de dicho punto a la ruta óptima, etc.

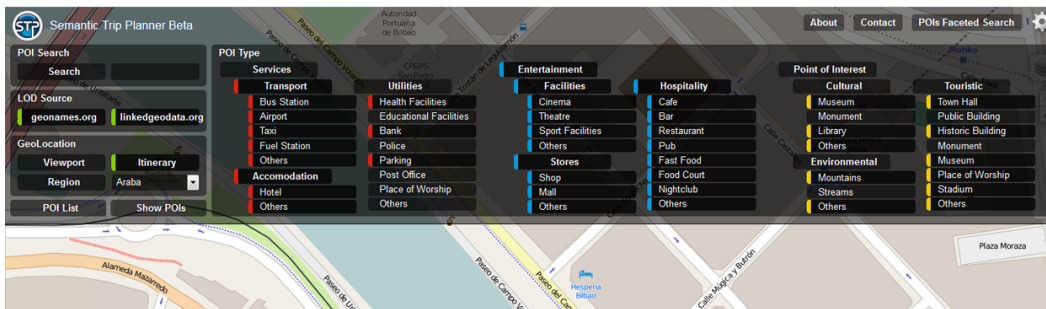


Figura 6.3: Interfaz web para la búsqueda facetada y de POIs

Esta consulta se genera de forma automatizada y transparente para el usuario, que simplemente escogerá entre las diferentes opciones de que dispone mediante el uso de la interfaz web provista y que se muestra en la figura 6.3. El fragmento de código 6.3 muestra la consulta base utilizada, donde en las líneas 6 y 7 se define el proveedor y la clasificación del POI. La línea 8 recoge información opcional del punto (como su enlace con la Wikipedia, p. ej.). Las líneas 9 a 11 filtran la información de acuerdo al nombre del POI o a su localización geográfica, bien por su proximidad a una ruta, bien por su existencia dentro de un área determinada.

```
1 SELECT DISTINCT ?poi ?nompoi ?lat ?lon ?info
2 WHERE {
3   ?poi a mto:POI; mto:hasName ?nompoi.
4   ?poi mto:hasLatitude ?lat; mto:hasLongitude ?lon.
5   ?poi geo:hasGeometry ?g.
6   ?poi mto:providedBy + data[provider][i].
7   ?poi mto:hasClassification mto: + data[class][i].
8   OPTIONAL ?poi mto:hasInfo ?info.
9   FILTER (regex(str(?nompoi), 'data["poiName"]', 'i'))
10  FILTER (geof:distance(?wkt, 'LINESTRING(legs.from, legs.to)'))
11  FILTER (geof:sfWithin(?wkt, 'POLYGON(point.A, point.B, ...)'))
12 }
```

Código 6.3: Consulta SPARQL dinámica en función del contexto de búsqueda



### 6.3.3 Consulta y acceso a la información semántica

Completado el proceso de modificación del proyecto OTP para adaptarlo a la arquitectura desplegada y con ello la generación del planificador multimodal semántico STP, se procede a comprobar el rendimiento y la funcionalidad aportada por éste. Para ello, se ha realizado en primera instancia una comparativa atendiendo al procedimiento de generación del grafo sobre el que se basa la ejecución del algoritmo de rutado. Posteriormente, en base a dicho grafo y a la arquitectura desplegada, se realiza un estudio funcional mediante la ejecución de una serie de consultas sobre la plataforma, atendiendo a las características de la información provista.

El objetivo perseguido con el desarrollo del planificador semántico es el de ilustrar cómo la arquitectura desplegada puede incorporarse a soluciones avanzadas de consulta ya existentes, asegurando así la aplicabilidad real del desarrollo realizado. Dicho objetivo ha sido satisfecho mediante la adaptación realizada del proyecto OTP. Con el fin de explicitar las diferencias y mostrar más detalles de la solución desarrollada, se quiere evaluar comparativamente tanto el rendimiento como la funcionalidad extra aportada por dicha solución frente al proyecto OTP original.

Características	OTP	STP
<b>Descarga automática de los datos de tránsito</b>	✗	✓
<b>Generación dinámica del grafo de enrutamiento</b>	✗	✓
- Filtrando por modos de transporte disponibles	✗	✓
- Filtrando por área geográfica a gestionar	✗	✓
<b>Tiempo utilizado para la generación del grafo</b>	321 s	453 s
- Procesamiento de los datos de tránsito	11 s	127 s
- Algoritmia combinatoria multimodal	310 s	326 s
<b>Acceso a información de transporte multimodal</b>	✓	✓
<b>Agregación de información relacionada</b>	✗	✓
<b>Acceso a información complementaria</b>	✗	✓
<b>Filtrado avanzado basado en conceptos semánticos</b>	✗	✓

**Tabla 6.5:** Comparativa entre OTP y el planificador semántico desarrollado

## 6. Experimentación y resultados

---

Dicha comparativa, detallada en la tabla 6.5, muestra, por un lado, el proceso y el tiempo necesarios para la generación del grafo de enrutamiento y, por otro, las características diferenciadoras de STP, atendiendo a la información proporcionada.

En cuanto al proceso de generación del grafo de enrutamiento, pieza clave en el desarrollo de la herramienta de planificación, se puede observar cómo STP proporciona características como la descarga automática de los datos de tránsito desde los servidores distribuidos desplegados o la generación dinámica del grafo de enrutamiento en base a la consulta SPARQL especificada, permitiendo así personalizar la información de tránsito provista por la herramienta de forma rápida y eficaz.

Esta configuración dinámica aumenta sin embargo el tiempo requerido para la generación del grafo. Esto es debido a que la herramienta descarga los datos bajo demanda y no se basa en ficheros ya descargados por el usuario y almacenados en local. Dicho aumento está, por lo tanto, completamente justificado ya que permite al usuario desentenderse de la localización y la gestión periódica del conjunto de ficheros de tránsito.

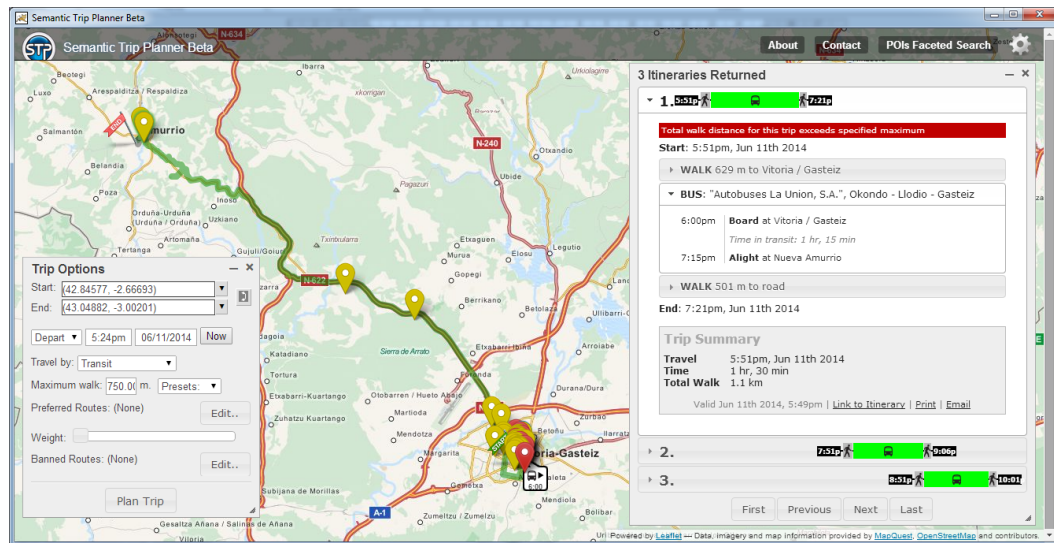
En cuanto a las características diferenciadoras de STP, relacionadas con la calidad de la información provista, puede establecerse de acuerdo con los datos recogidos en la tabla 6.5 que la información proporcionada por la solución desarrollada es más precisa y completa. La estructuración y clasificación de los datos gracias a su organización semántica, la agregación de información enlazada desde vocabularios externos y datasets colaborativos, junto con la adopción de estándares para la consulta de información geoespacial, ha permitido por un lado enriquecer y precisar la información provista y, por otro, adecuarla a las necesidades concretas del usuario, personalizando la respuesta y dotando de relevancia a la misma.

Se ha querido verificar esta afirmación mediante la resolución de dos consultas concretas, escogidas específicamente de modo que se utilice el mayor número de características posibles en su resolución. A continuación se presentan dichas consultas junto con la resolución proporcionada por la herramienta STP:

Consulta 1: “*Localiza las gasolineras cercanas disponibles en la ruta multimodal desde Vitoria-Gasteiz a Amurrio (Araba)*” (ver figura 6.4)

Consulta 2: “*Muestra localizaciones turísticas de entretenimiento en el área de Bilbao (adjuntar información relacionada)*” (ver figura 6.5)

### 6.3 STP: Planificador Semántico de Transporte



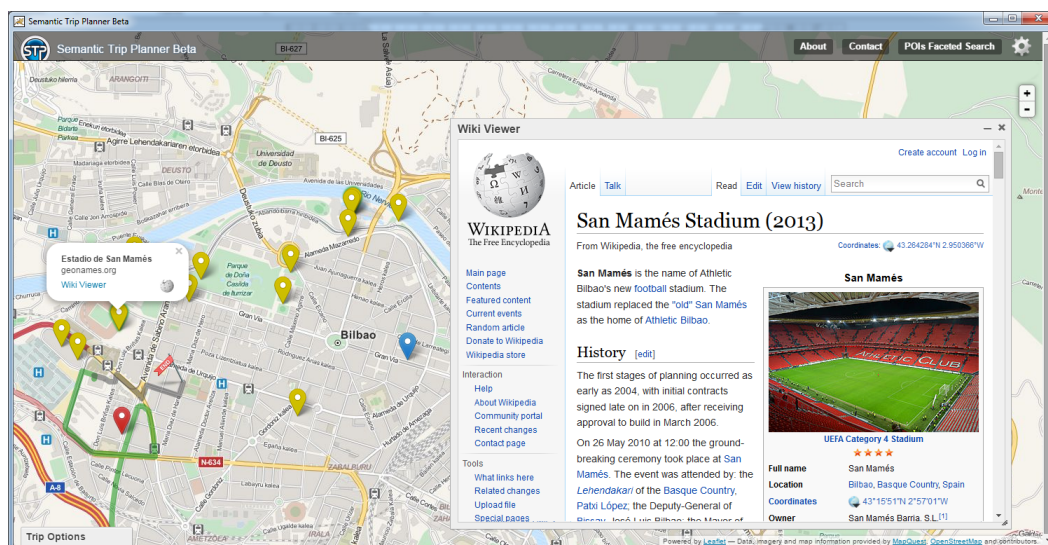
**Figura 6.4:** Agregación de información externa en relación a la ruta generada

La resolución de la primera consulta ejemplifica los beneficios aportados al usuario derivados de la agregación de información contextual externa relacionada (en este caso las gasolineras cercanas), proveniente del vocabulario colaborativo *LinkedGeoData* enlazado en la ontología mediante la clase POI (ver sección 4.3).

Mientras que OTP limita su funcionalidad a generar y visualizar la ruta óptima, el planificador semántico desarrollado es capaz de relacionar dicha ruta con la presencia de gasolineras mediante la utilización de funciones geográficas aplicadas a la resolución de consultas, como el filtrado por distancias de GeoSPARQL (ver sección 4.2.2.1) permitiendo generar información contextual (las gasolineras en ruta) de forma dinámica, agregando información externa y filtrándola en base a las necesidades concretas del usuario.

En relación con la segunda consulta, ésta evidencia las posibilidades de filtrado semántico y acceso a información complementaria de la arquitectura. En primer lugar, la solución clasifica tanto la localización (Bilbao) vía organización administrativa (ver sección 4.2.2.2), como el tipo de POI (entretenimiento) de cara a mostrar información contextual precisa. Posteriormente, haciendo uso de las posibilidades de enlace con otros vocabularios, proporciona marcadores para los POIs con información extra (proveniente de DBpedia). Así, se ha definido una nueva vista en STP para mostrar la información específica de los elementos seleccionados.

## 6. Experimentación y resultados



**Figura 6.5:** Clasificación, filtrado y agregación de información complementaria

## 6.4 Resultados de la experimentación

Las pruebas descritas a lo largo de este capítulo detallan el proceso de experimentación realizado. Tomando como referencia la estrategia de validación de la hipótesis definida en el apartado 6.1 y a través del análisis de los resultados de las pruebas, es posible comprobar el cumplimiento de los objetivos funcionales de la arquitectura fruto de este trabajo, demostrando finalmente la veracidad de la hipótesis.

### 6.4.1 Modelización de los datos y enriquecimiento

La hipótesis presentada enuncia que es posible desarrollar un procedimiento capaz de generar información de transporte multimodal estructurada e integrada, enlazando datos desde diversas fuentes y agregando información relevante relacionada.

En el capítulo 4 se ha explicado dicho procedimiento, detallando el diseño de un modelo ontológico integrado para la representación de la información de transporte, con el que es posible extender y enlazar los datos de tránsito disponibles, así como incorporar información relacionada relevante. Durante el presente capítulo se han validado dichas capacidades mediante la definición de un escenario de pruebas fidedigno, utilizando datos de tránsito provistos por las agencias de transporte de la CAV como base sobre la que generar la nueva información semántica.

Partiendo de dicho escenario y utilizando el adaptador desarrollado para la conversión desde el formato GTFS se ha obtenido un *dataset* completo de transporte público (ver sección 6.2.2), incluyendo información geopolítica y geoespacial mediante el enlace de la ontología con otros conjuntos de datos, habilitando la consulta de información en base a dichas características geográficas.

Con objeto de validar las características de integración, agregación y enriquecimiento de la información se ha procedido también al despliegue de un planificador semántico basado en dicho modelo de datos semántico. El análisis funcional de las características aportadas por dicho planificador en comparación con el proyecto OTP (ver sección 6.3.3) ha demostrado que la herramienta STP es capaz de proveer información más precisa y completa. La estructuración y clasificación de los datos gracias a su organización semántica, la agregación de información enlazada desde vocabularios externos y datasets colaborativos, junto con la adopción de estándares para la consulta de información geográfica, ha permitido enriquecer la información provista, adecuándola además a las necesidades concretas del usuario.

### 6.4.2 Explotación y publicación de la información

La segunda parte de la hipótesis sostiene que el despliegue de la arquitectura permitirá la creación de servicios avanzados de consulta de información, basados en la publicación distribuida, abierta e interoperable de los datos de transporte.

En el capítulo 5 se describe el proceso de diseño de dicha arquitectura, haciendo uso de la web semántica como infraestructura sobre la que construir el mecanismo de gestión de la información generada, permitiendo a su vez la publicación abierta de la misma. Se detalla también el proceso de despliegue de la arquitectura, en forma de servidores semánticos distribuidos e interoperables colaborando en la resolución de consultas. En este capítulo se ha querido evidenciar la validez de la premisa fundamental de la hipótesis, estableciendo el desarrollo de un planificador multimodal semántico como herramienta avanzada de consulta de información basado en la gestión de los datos de tránsito provistos por la arquitectura desplegada.

En primera instancia, se ha procedido al despliegue de la solución según el escenario definido anteriormente, esto es, mediante la puesta en marcha de tres servidores semánticos distribuidos e interoperables (uno por cada provincia de la CAV) publicando como LOD la información semántica generada (ver sección 6.2.3).

## 6. Experimentación y resultados

---

Ha quedado demostrado mediante la comparación directa con una alternativa basada en un diseño centralizado (mayoritariamente utilizado en la literatura) que la solución presentada es más eficiente en la resolución de consultas que afecten a una gran cantidad de información y, por ende, conlleven mayores necesidades de cómputo. Este hecho reafirma la idoneidad de un modelo como el propuesto para la provisión de información semántica de transporte en escenarios de actuación reales, donde el volumen de información semántica a gestionar es grande.

Se ha querido evidenciar dicha afirmación, verificando además la capacidad de integración de la arquitectura propuesta, con la modificación del proyecto de código abierto OTP (ver sección 6.3.2) para la generación de un planificador multimodal semántico, corriendo sobre la arquitectura distribuida desplegada y haciendo uso de la información semántica de transporte provista.

Las ventajas identificadas en la utilización de dicha arquitectura como soporte de este tipo de herramientas de consulta (ver sección 6.3.3), limitando al máximo la gestión de la configuración al descargar automáticamente los datos de tránsito, generando el grafo de enrutamiento de forma dinámica y mejorando la precisión y la actualidad de la información proporcionada, permiten demostrar la validez de esta segunda parte de la hipótesis, estableciendo la arquitectura diseñada como una opción destacada para la generación de herramientas avanzadas de consulta de información de transporte.

Como conclusión, destacar cómo mediante la ejecución del conjunto de pruebas funcionales con las que se ha dirigido el proceso de experimentación: el despliegue de la solución distribuida en un escenario de actuación real, la modelización de los datos de tránsito y la generación de un dataset ontológico, y la adaptación del proyecto OTP para la construcción de un planificador multimodal semántico corriendo sobre la arquitectura propuesta y consumiendo la información provista; se pueden dar por validados los requisitos contemplados tanto en los objetivos iniciales de este trabajo como en la propia hipótesis.

Se adjunta a continuación, a modo de síntesis de los resultados obtenidos, la tabla 6.6, que permite ver claramente la relación entre las capacidades teóricas indicadas en la hipótesis, las contribuciones realizadas en pos de su consecución, y las evidencias que han servido para validar su cumplimiento.

## 6.4 Resultados de la experimentación

Hipótesis	Contribuciones	Evidencia
<b>Modelización de los datos y enriquecimiento</b>		
Generación de información multimodal estructurada	Ontología para el transporte multimodal (ver sección 4.2)	Generación del contenido ontológico (ver sección 6.2.2)
Agregación de datos desde diversas fuentes e información relacionada	Vocabularios enlazados e información contextual (ver sección 4.3)	Dataset de la CAV. Características de STP (ver sección 6.3.3)
<b>Explotación y publicación de la información</b>		
Arquitectura para la provisión de información de transporte	Arquitectura distribuida de servidores semánticos (ver sección 5.2)	Despliegue de la arquitectura en la CAV (ver sección 6.2.3)
Habilitando la creación de servicios avanzados de consulta	Desarrollo planificador multimodal semántico (ver sección 6.3)	Despliegue de STP sobre la arquitectura (ver sección 6.3.3)

**Tabla 6.6:** Actividades realizadas conducentes a la validación de la hipótesis





*La ciencia no es sino una perversión de sí misma a menos que tenga como objetivo final la mejora de la humanidad.*

Nikola Tesla

# 7

## Conclusiones

A lo largo de los capítulos previos se han ido detallando de forma exhaustiva las actividades llevadas a cabo durante el desarrollo de esta tesis doctoral. Así, una vez situado el ámbito de los ITS, se abordó la problemática relacionada con la deficiente gestión y publicación de la información multimodal de transporte y sus consecuencias en cuanto a la aparición de herramientas y servicios software avanzados que basan su funcionamiento en la cantidad y la calidad de los datos disponibles. Tras realizar el análisis de la literatura existente se encontraron diversas oportunidades de mejora en este área, atendiendo a las necesidades de integración, interoperabilidad, y apertura de la información, entre otras.

Para dar respuesta a estas limitaciones se desarrollan dos soluciones principales: una ontología para la representación de los datos de transporte multimodal y una arquitectura distribuida para la provisión de información semántica de transporte. Posteriormente se lleva a cabo la experimentación, atestiguando la veracidad de la hipótesis planteada al inicio del trabajo y permitiendo, además, contextualizar los beneficios e inconvenientes derivados de la aplicación de la arquitectura propuesta en un entorno de actuación real, mediante la construcción de una solución de planificación de rutas semántica corriendo sobre dicha arquitectura.

## 7. Conclusiones

---

En este capítulo se ofrece una perspectiva global de las cuestiones que se han tratado en este documento, haciendo hincapié en las soluciones y contribuciones principales y en los resultados de la experimentación. Para finalizar, se introducen las líneas futuras de trabajo, proponiendo una serie de mejoras o ampliaciones en torno a esta tesis que se consideran de interés.

### 7.1 Visión general, motivación y objetivos

Una de las premisas fundamentales sobre las que se constituye el trabajo presentado es la necesidad de contar con información de transporte de calidad. Siendo como se ha mostrado la movilidad ciudadana sostenible una prioridad para las instituciones, y teniendo en cuenta los avances realizados en los últimos años en la creación de servicios software de información al viajero, se ha podido establecer una relación directa entre el valor añadido aportado por dichos servicios y la calidad de la información de transporte disponible.

De esta manera, cada vez adquiere más importancia establecer mecanismos que den soporte a las aplicaciones mediante la correcta gestión y publicación de la información. Tras efectuar el análisis del estado del arte se encontraron diversas oportunidades en este sentido. Se consideró que las soluciones existentes no eran lo suficientemente interoperables, ni fácilmente consultables o reutilizables, convirtiéndose en silos de información donde el acceso a los datos es muy costoso, derivando en información inconexa, desactualizada y aislada.

Este trabajo propone un nuevo enfoque para la gestión y el intercambio de información de transporte. Dicho enfoque se plasma mediante el diseño de una arquitectura software distribuida, permitiendo, gracias a la estructuración de datos provenientes de fuentes heterogéneas, junto con la integración de información relacionada, ofrecer una plataforma para la habilitación de servicios software relacionados con la movilidad. Se han utilizado para tal fin un amplio abanico de tecnologías innovadoras. Ejemplo de ello son el diseño efectuado de un middleware ontológico para modelar e integrar los datos, la publicación de éstos mediante una arquitectura distribuida de servidores semánticos interoperables, o la agregación del contexto de cara a enriquecer la información de transporte provista.

### 7.1.1 Contribuciones principales

El desarrollo del presente trabajo de tesis, alineado con los objetivos específicos inicialmente establecidos, ha arrojado principalmente dos contribuciones destacadas: un modelo ontológico para la representación de los datos de transporte multimodal (MTO) y una arquitectura distribuida, basada en tecnologías de la web semántica, enfocada a la publicación como LOD de la información de transporte disponible.

En cuanto a la ontología MTO, ésta cumple con la función de integrar los datos de tránsito no estructurados disponibles, enlazando vocabularios existentes que enriquezcan la información provista y generando un modelo semántico para la representación y publicación de la información de transporte.

Para su diseño se decide tomar como base la especificación GTFS, reformulando los archivos CSV de dicho formato a entidades y relaciones dentro de la ontología resultante. Esto ha permitido generar un modelo semántico para un dominio estable y con un alcance concreto, en el que la definición de las clases y sus propiedades son conocidas, favoreciendo su comprensión y facilitando su adopción. El proceso de carga de los ficheros GTFS y generación del contenido ontológico se encuentra completamente automatizado gracias al adaptador desarrollado: una aplicación Java encargada de realizar todo el proceso, incluyendo las referencias semánticas a vocabularios externos.

A este respecto, una de las características fundamentales del modelo es el diseño utilizado para almacenar la información geográfica. Se decidió la utilización del estándar GeoSPARQL soportando de este modo la representación y consulta de datos geoespaciales mediante la extensión del lenguaje de consulta SPARQL. Así mismo, también se enlaza la ontología GeoNames, una base de datos geográfica enfocada al almacenamiento de información geopolítica.

Gracias a la incorporación de estos vocabularios, junto con otros como Linked-GeoData, enlazando información colaborativa de interés para el dominio del transporte, es posible la ejecución de cálculos geoespaciales avanzados asociados a consultas basadas en conceptos semánticos y al contexto concreto del usuario, personalizando así la respuesta. Esto, unido al despliegue de un servidor de contenido semántico geoespacial, otorga a MTO funcionalidades avanzadas de búsqueda y filtrado inasumibles por otro tipo de formatos para la representación de información de transporte, dotando de mayor exactitud y relevancia a los datos provistos.

## 7. Conclusiones

---

La manera en la que se representa y se estructura la información de transporte, en este caso mediante la creación de la ontología MTO, supone parte de la innovación realizada en el presente trabajo. Sin embargo, dicho modelo de datos debe tener un soporte que permita la consulta, búsqueda y filtrado de la información disponible. Dicho soporte, en forma de arquitectura semántica distribuida, supone la segunda aportación principal de esta tesis, y su despliegue está permitiendo la publicación efectiva de la información ontológica generada.

En una arquitectura enfocada a la provisión de información interoperable resulta fundamental contar con un protocolo de acceso a datos que garantice la apertura, disponibilidad y adecuada explotación de ésta. En este sentido se establece el uso de la web semántica como infraestructura estable sobre la que construir la arquitectura, haciendo un uso extensivo del conjunto de protocolos disponible (desde HTTP a OWL o SPARQL) y basándose en el paradigma LOD para la publicación de información abierta, integrada y enlazada.

La información semántica generada, accesible desde múltiples dispositivos y provista en diversos formatos, es almacenada en servidores SPARQL distribuidos e interrelacionados. Dichos servidores semánticos cuentan con dos funcionalidades destacadas: la capacidad para resolver consultas geoespaciales y la posibilidad de atender a consultas federadas, mecanismo que garantiza el correcto funcionamiento del sistema bajo condiciones de gran volumen de datos, como las producidas en escenarios de actuación reales en el ámbito del transporte multimodal.

Esto es debido a que la arquitectura diseñada posibilita bajo dicha especificación la compartición efectiva de los datos de transporte. El servidor SPARQL correspondiente, haciendo uso de sus capacidades de agregación, es el encargado de integrarlos junto con otros provenientes de servidores afines remotos, distribuyendo así la carga de procesamiento al mantener varios servidores colaborando entre sí. Cada uno de los servidores SPARQL desplegados gestiona información de transporte propia y es administrado localmente, pero facilita la interoperabilidad gracias a su interconexión con el resto de servidores del sistema, habilitando de este modo un mecanismo para la realización de consultas distribuidas.

Se ha definido una organización de servidores SPARQL jerárquica, permitiendo distribuir la información multimodal en servidores clasificados por área geográfica administrativa. Así, el servidor asociado a la CAV contendrá la información con los *datasets* de las tres provincias vascas: Araba, Bizkaia y Gipuzkoa.

El hecho de poder acceder a un conjunto de datos extenso y complejo (con más de 10 agencias de transporte y 350 rutas únicas para el despliegue realizado), desde un único lugar asociado a la URL pública del servidor SPARQL utilizado y mediante una arquitectura distribuida estable basada en tecnologías web, supone otra de las aportaciones realizadas en el presente trabajo de tesis.

### 7.1.2 Resultados de la experimentación

Dado que las principales aportaciones realizadas se fundamentan en el desarrollo de elementos software en el ámbito de los ITS, se consideró adecuado establecer una serie de acciones encaminadas a verificar funcionalmente el cumplimiento de los objetivos inicialmente establecidos y, en consecuencia, a la validación de la propia hipótesis. Se resumen aquí dichas acciones:

1. La carga de los datos de transporte disponibles publicados por el Gobierno Vasco en formato GTFS y la generación del contenido semántico correspondiente (MTO) mediante el adaptador desarrollado.
2. El despliegue de la arquitectura, con 3 servidores SPARQL distribuidos publicando como LOD la información de transporte multimodal semantizada referente a las 3 provincias de la CAV.
3. La puesta en marcha de un planificador multimodal semántico, extendiendo el proyecto OTP y corriendo sobre la arquitectura desplegada:
  - (a) Consumiendo el modelo ontológico desarrollado
  - (b) Proveyendo información multimodal satisfactoria
  - (c) Proveyendo información adicional relacionada

Inicialmente, se estableció la utilización de un escenario de pruebas fidedigno, usando datos de tránsito provistos por las agencias de transporte de la CAV como base sobre la que generar la nueva información ontológica. Partiendo de dicho escenario y utilizando el adaptador desarrollado, se obtuvo un *dataset* semántico completo de transporte público para la CAV.

## 7. Conclusiones

---

Con objeto de validar las características de integración y agregación de información se procedió al despliegue de la solución según el escenario definido, esto es, mediante la puesta en marcha de tres servidores semánticos distribuidos e interoperables (uno por cada provincia) publicando como LOD la información semántica generada. Además, se realizó una comparativa directa con una alternativa basada en un diseño centralizado, evidenciando la idoneidad del modelo distribuido para la gestión de escenarios de actuación reales con un gran volumen de datos.

Posteriormente, se quiso comprobar la capacidad de integración y la funcionalidad provista por la arquitectura, mediante la modificación del proyecto OTP y la generación de un planificador multimodal semántico, corriendo sobre ésta y haciendo uso del *dataset* generado. Las ventajas identificadas en la utilización de dicha arquitectura como soporte de este tipo de herramientas permitieron posicionar también la solución como una opción óptima para la generación de servicios avanzados de consulta de información. Por otra parte, el análisis funcional de las características aportadas por dicho planificador en comparación con el proyecto OTP ha demostrado que la herramienta STP es capaz de proveer información enriquecida más completa, mejorando la precisión y la actualidad de los datos proporcionados.

Como conclusión, destacar cómo el proceso de experimentación realizado, además de dar cumplimiento al objetivo general de la tesis con el despliegue de la solución propuesta bajo condiciones reales, ha servido para satisfacer la hipótesis enunciada al inicio del trabajo, estableciendo una arquitectura interoperable con la que proveer información de transporte completa, estructurada e integrada, habilitando de igual modo la creación de servicios en torno a ella.

### 7.2 Líneas futuras de trabajo

Se podría establecer que la arquitectura planteada es formalmente independiente, ya que por sí sola es capaz de facilitar la interoperabilidad (mediante la agregación de servidores distribuidos) y el enriquecimiento (mediante la vinculación de MTO con otros vocabularios). Sin embargo, a nivel práctico, la solución depende de componentes anejos que están permitiendo la agregación efectiva de los datos disponibles. En el caso concreto de la implementación realizada, se ofrece un adaptador para el formato GTFS, ampliamente utilizado por las agencias de transporte, con el que automatizar el proceso de semantización de la información de tránsito.

Siguiendo esta misma línea, de cara a extender universalmente el uso de la arquitectura propuesta y en base a la experiencia obtenida en proyectos de investigación a nivel Europeo, se propone específicamente la generación de un adaptador aplicado al estándar Datex II analizado en la sección 2.3.1 del estado del arte.

La UE sigue apostando por la interoperabilidad en lo que respecta a la gestión de la información de transporte, lo que se demuestra con el apoyo realizado bajo el Programa Marco Horizonte 2020<sup>1</sup> a iniciativas de investigación en dicho ámbito, como el proyecto Europeo TIMON<sup>2</sup>, impulsado y coordinado desde la Universidad de Deusto con la colaboración de otras 11 entidades de 8 países. Este proyecto tiene como objetivo incrementar la seguridad y la eficiencia de los sistemas de transporte por carretera mediante el uso de comunicaciones móviles cooperativas y la armonización de datos abiertos de transporte.

Como parte de las actividades de pilotaje del proyecto se ha planificado un despliegue real en la ciudad de Liubliana (Eslovenia). Se abre por tanto una oportunidad para integrar los datos recogidos en dicho despliegue dentro del modelo semántico diseñado en esta tesis, enriqueciendo así la información y posibilitando la creación de servicios específicos para dicho área basados en la explotación de la arquitectura propuesta.

En lo que se refiere específicamente al diseño realizado, se vislumbra también un área de investigación vinculada al enriquecimiento de la información. Durante el desarrollo de la solución se han focalizado los esfuerzos en la modelización, integración y distribución de la información semántica generada. En relación al enriquecimiento, se establecieron también sus bases (agregación de datos y adecuación al contexto) lo que ha permitido generar respuestas más completas y precisas. Se considera sin embargo que existe aún capacidad de mejora en este aspecto. Por ello, se proponen las siguientes opciones para continuar dicha investigación:

- ◇ En cuanto a la agregación de datos desde vocabularios externos relacionados, se contempló la extensión de la ontología MTO enlazando y clasificando los puntos de interés expuestos por el *dataset* de LinkedGeoData, ejemplificando cómo el modelo ontológico propuesto habilita la integración a nivel semántico de información relacionada, enriqueciendo así las respuestas.

---

<sup>1</sup> Programa Horizon 2020, <http://ec.europa.eu/programmes/horizon2020>

<sup>2</sup> Proyecto TIMON, <http://www.timon-project.eu>

## 7. Conclusiones

---

Dicho ejemplo, aún siendo efectivo, no permite valorar suficientemente el nivel de enriquecimiento derivado del uso de un modelo de datos semántico para la representación de los datos de transporte. Se propone una ampliación de dicho concepto, con la consiguiente ampliación de la ontología MTO, enlazando y extendiendo conceptos mediante la participación de un mayor número de vocabularios, permitiendo valorar de esta manera la correlación entre el modelo de datos usado y la completitud de la información provista.

- ◇ Otro de los aspectos relevantes a la hora de proveer información enriquecida es la consideración del contexto del usuario. Se definió para ello una fase de búsqueda y filtrado desde la que recoger los intereses del usuario en forma de parámetros de búsqueda. Dicha propuesta permite ejemplificar cómo la utilización del contexto genera información personalizada más precisa.

Dado el mecanismo utilizado para la captura del contexto, basado principalmente en los parámetros de búsqueda facilitados, se considera que no se ha explotado suficientemente la ventaja derivada de su utilización. Se propone una ampliación de dicho modelo, atendiendo a parámetros como el historial o los datos de uso, por ejemplo. Estos parámetros, recogidos con la autorización del usuario y de manera no intrusiva, permitirían generar un modelo de contexto más preciso, mejorando previsiblemente los resultados provistos.

### 7.3 Consideraciones finales

Los sistemas software, incluidos los vinculados al transporte, se asientan fundamentalmente bajo dos elementos principales: la lógica de negocio, regulando las acciones a realizar; y los datos: definiendo el contenido, la información a manejar. Ambos conceptos son clave, condición *sine qua non* dentro de un software cada vez más sofisticado y a la vez más presente en nuestra vida cotidiana.

Valga esta aseveración para poner de relieve la importancia de los datos en la construcción de software complejo, cuya finalidad última, dentro del ámbito que nos ocupa, es mejorar algo tan importante como la calidad de vida en nuestras ciudades. Objetivo a cumplir mediante la correcta gestión de la movilidad ciudadana, lo que incluye la eficiencia y la seguridad de los sistemas de transporte y también su sostenibilidad, económica y medioambiental.



Así, siguiendo la estrategia apuntada por la UE en el Libro Blanco del transporte [Europea 11], se pretende mejorar y ampliar la capacidad de instituciones públicas, empresas y usuarios de cara a obtener, compartir y proporcionar información de transporte, de manera sencilla y completa.

El estudio llevado a cabo sobre los sistemas que actualmente se utilizan para representar la información de transporte ofreció la oportunidad de mejorar su representación, gestión y provisión, a menudo basada en el uso de bases de datos heterogéneas o formatos cerrados no interoperables. Por otro lado, dicho estudio también permitió evidenciar cómo las nuevas tendencias aplicadas a la gestión de los datos, como la modelización mediante ontologías o el uso de la web semántica como infraestructura sobre la que vincular la información, podían ayudar en la tarea de estructuración, compartición y distribución de los datos de transporte actualmente disponibles.

El trabajo realizado se constituye así como promotor para el desarrollo de aplicaciones y servicios innovadores en un área, la gestión de la movilidad y el transporte, con un gran apoyo social e institucional y un amplio campo de investigación por delante. Los próximos pasos apuntan, inequívocamente, a una concienciación general sobre los beneficios derivados de la apertura y la compartición de la información, lo que a buen seguro se traducirá en importantes avances al servicio de las personas.



# Acrónimos

<b>API</b>	Interfaz de programación de aplicaciones (del inglés <i>Application Programming Interface</i> )
<b>ATIS</b>	Sistemas Avanzados de Información al Viajero (del inglés <i>Advanced Traveller Information Systems</i> )
<b>CAV</b>	Comunidad Autónoma Vasca
<b>GDF</b>	Fichero de Datos Geográficos (del inglés <i>Geographic Data File</i> )
<b>GFM</b>	<i>General Feature Model</i>
<b>GIS</b>	Sistema de Información Geográfica (del inglés <i>Geographic Information System</i> )
<b>GML</b>	Lenguaje de Marcado Geográfico (del inglés <i>Geography Markup Language</i> )
<b>GTFS</b>	Especificación general de fuentes de transporte público (del inglés <i>General Transit Feed Specification</i> )
<b>HTTP</b>	Protocolo de Transferencia de Hipertexto (del inglés <i>Hypertext Transfer Protocol</i> )
<b>ISO</b>	Organización Internacional de Normalización (del inglés <i>International Organization for Standardization</i> )
<b>ITS</b>	Sistemas Inteligentes de Transporte (del inglés <i>Intelligent Transportation Systems</i> )

## ACRÓNIMOS

---

<b>LD</b>	Datos Enlazados (del inglés <i>Linked Data</i> )
<b>LOD</b>	Datos Enlazados Abiertos (del inglés <i>Linked Open Data</i> )
<b>NOR</b>	Recursos no ontológicos (del inglés <i>Non-Ontological Resources</i> )
<b>OGC</b>	<i>Open Geospatial Consortium</i>
<b>OWL</b>	Lenguaje de Ontologías Web (del inglés <i>Ontology Web Language</i> )
<b>RDF</b>	Infraestructura para la Descripción de Recursos (del inglés <i>Resource Description Framework</i> )
<b>REST</b>	Transferencia de Estado Representacional (del inglés <i>Representational State Transfer</i> )
<b>SOAP</b>	Protocolo de comunicación web sin estado (del inglés <i>Simple Object Access Protocol</i> )
<b>SPARQL</b>	Protocolo y Lenguaje de Consulta para RDF (del inglés <i>SPARQL Protocol and RDF Query Language</i> )
<b>SWRL</b>	Lenguaje de Reglas para la Web Semántica (del inglés <i>Semantic Web Rule Language</i> )
<b>TIC</b>	Tecnologías de la Información y la Comunicación
<b>UE</b>	Unión Europea
<b>UML</b>	Lenguaje Unificado de Modelado (del inglés <i>Unified Modeling Language</i> )
<b>URI</b>	Identificador de Recursos Uniforme (del inglés <i>Uniform Resource Identifier</i> )
<b>W3C</b>	<i>World Wide Web Consortium</i>
<b>WFS</b>	<i>Web Feature Service</i>
<b>WGS84</b>	Sistema Geodésico Mundial 1984 (del inglés <i>World Geodesic System 84</i> )

<b>WSDL</b>	Lenguaje para la descripción de servicios web (del inglés <i>Web Services Description Language</i> )
<b>XML</b>	Lenguaje de Marcado Extensible (del inglés <i>eXtensible Markup Language</i> )



# Bibliografía

- [Abowd 97] Gregory D Abowd, Christopher G Atkeson, Jason Hong, Sue Long, Rob Kooper and Mike Pinkerton. *Cyberguide: A mobile context-aware tour guide*. Wireless networks, Vol. 3, No. 5, pp. 421–433, 1997 (pág. 44).
- [Adler 98] Jeffrey L Adler and Victor J Blue. *Toward the design of intelligent traveler information systems*. Transportation Research Part C: Emerging Technologies, Vol. 6, No. 3, pp. 157–172, 1998 (pág. 20).
- [Agarwal 05] Pragya Agarwal. *Ontological considerations in GIScience*. International Journal of Geographical Information Science, Vol. 19, No. 5, pp. 501–536, 2005 (pág. 51, 60, 109).
- [Al-Deek 93] Haitham Al-Deek and Adib Kanafani. *Modeling the benefits of advanced traveler information systems in corridors with incidents*. Transportation Research Part C: Emerging Technologies, Vol. 1, No. 4, pp. 303–324, 1993 (pág. 20).
- [Anand 12] Nilesh Anand, Mengchang Yang, J.H.R. van Duin and Lori Tavasszy. *GenCLOn: An ontology for city logistics*. Expert Systems with Applications, Vol. 39, No. 15, pp. 11944 – 11960, 2012 (pág. 72).
- [Ashburner 00] Michael Ashburner *et al.* *Gene Ontology: tool for the unification of biology*. Nature genetics, Vol. 25, No. 1, pp. 25–29, 2000 (pág. 44).

## BIBLIOGRAFÍA

---

- [Auer 12] Sören Auer *et al.* *Managing the Life-Cycle of Linked Data with the LOD2 Stack*. In *The Semantic Web - ISWC 2012*, pp. 1–16, 2012 (pág. 42).
- [Berners-Lee 01] Tim Berners-Lee, James Hendler, Ora Lassila *et al.* *The semantic web*. *Scientific american*, Vol. 284, No. 5, pp. 28–37, 2001 (pág. 38).
- [Berners-Lee 09a] *Design issues: Linked data*. Disponible en <http://www.w3.org/DesignIssues/LinkedData.html> (consultado en 2015) (pág. 60).
- [Berners-Lee 09b] Tim Berners-Lee, Christian Bizer and Tom Heath. *Linked data-the story so far*. *International Journal on Semantic Web and Information Systems*, Vol. 5, No. 3, pp. 1–22, 2009 (pág. 40).
- [Boucher 09] Serge Boucher and Esteban Zimányi. *An ontology-based geodatabase interoperability platform*. *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications*, pp. 294–315, 2009 (pág. 45, 109).
- [Buján 13] David Buján, David Martín, Ortzi Torices, Diego López-de Ipiña, Carlos Lamsfus, Joseba Abaitua and Aurkene Alzua-Sorzabal. *Context Management Platform for Tourism Applications*. *Sensors*, Vol. 13, No. 7, pp. 8060–8078, 2013 (pág. 44).
- [Cali 13] Andrea Cali, Diego Calvanese, Giuseppe De Giacomo and Maurizio Lenzerini. *Data Integration under Integrity Constraints*. In *Seminal Contributions to Information Systems Engineering*, pp. 335–352. Springer Berlin Heidelberg, 2013 (pág. 25).
- [Campbell 98] John Lyle Campbell, C Carney and Barry H Kantowitz. *Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO)*. National Technical Information Service, 1998 (pág. 3, 25).



- [Caulfield 07] B. Caulfield and M. O'Mahony. *An Examination of the Public Transport Information Requirements of Users*. Intelligent Transportation Systems, IEEE Transactions on, Vol. 8, No. 1, pp. 21–30, March 2007 (pág. 20).
- [Chen 03] Harry Chen, Tim Finin and Anupam Joshi. *An ontology for context-aware pervasive computing environments*. The Knowledge Engineering Review, Vol. 18, No. 03, pp. 197–207, 2003 (pág. 4).
- [DaPaaS 13] DaPaaS. *A data-and-platform-as-a-service approach to efficient open data publication and consumption*. Disponible en <http://project.dapaas.eu/> (consultado en 2014) (pág. 42).
- [Davies 02] John Davies, Frank van Harmelen and Dieter Fensel. *Towards the semantic web: ontology-driven knowledge management*. John Wiley & Sons, Inc., 2002 (pág. 2).
- [De Oliveira 13] KáThia MarçAl De Oliveira, Firas Bacha, Houda Mnasser and Mourad Abed. *Transportation ontology definition and application for the content personalization of user interfaces*. Expert Systems with Applications, Vol. 40, No. 8, pp. 3145–3159, 2013 (pág. 54, 63, 96).
- [Delling 09] Daniel Delling, Peter Sanders, Dominik Schultes and Dorothea Wagner. *Engineering Route Planning Algorithms*. In *Algorithms of Large and Complex Networks*, volume 5515 of *Lecture Notes in Computer Science*, pp. 117–139. Springer Berlin Heidelberg, 2009 (pág. 20).
- [Dey 01] Anind K Dey. *Understanding and using context*. Personal and ubiquitous computing, Vol. 5, No. 1, pp. 4–7, 2001 (pág. 85).
- [Ding 12] Li Ding, Vassilios Peristeras and Michael Hausenblas. *Linked Open Government Data [Guest editors'introduction]*. Intelligent Systems, IEEE, Vol. 27, No. 3, pp. 11–15, 2012 (pág. 40).

## BIBLIOGRAFÍA

---

- [Dingus 97] T. A. Dingus, M. C. Hulse, M. A. Mollenhauer, R. N. Fleischman, D. V. Mcgehee and N. Manakkal. *Effects of age, system experience, and navigation technique on driving with an advanced traveler information system*. Human Factors: The Journal of the Human Factors and Ergonomics Society, Vol. 39, No. 2, pp. 177–199, 1997 (pág. 20).
- [Doan 12] AnHai Doan, Alon Halevy and Zachary Ives. Principles of data integration. Elsevier, 2012 (pág. 34).
- [Dong 08] Hai Dong, F.K. Hussain and E. Chang. *Transport service ontology and its application in the field of semantic search*. In Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on, volume 1, pp. 820–824, Oct 2008 (pág. 44).
- [Dong 13] Xin Luna Dong and Divesh Srivastava. *Big data integration*. In Data Engineering (ICDE), 2013 IEEE 29th International Conference on, pp. 1245–1248. IEEE, 2013 (pág. 34).
- [EDRM 94] *European Digital Road Map (EDRM)*. Disponible en [http://cordis.europa.eu/project/rcn/16990\\_en.html](http://cordis.europa.eu/project/rcn/16990_en.html) (consultado en 2014) (pág. 46).
- [Encuentro 12] *Gasto público en redes de transporte y su evolución*. Disponible en [http://www.fund-encuentro.org/informe\\_espana/descargar.php?id=2012-C5](http://www.fund-encuentro.org/informe_espana/descargar.php?id=2012-C5) (consultado en 2015) (pág. 15).
- [Europea 11] Comisión Europea. Libro blanco del transporte: Hoja de ruta hacia un espacio único europeo de transporte: por una política de transportes competitiva y sostenible. Oficina de publicaciones de la unión europea, 2011 (pág. 1, 18, 135).
- [Fensel 11] Dieter Fensel, Federico Michele Facca, Elena Simperl and Ioan Toma. Semantic web services. Springer, 2011 (pág. 42).

- [Fernández 97] Mariano Fernández, Asunción Gómez and Natalia Juristo. *Met-hontology: from ontological art towards ontological engineering*. American Association for Artificial Intelligence, 1997 (pág. 37).
- [Fernández 06] *Pila de la Web Semántica en español*. Disponible en <http://www.wikier.org/blog/pila-de-la-web-semantica-en-espanol> (consultado en 2014) (pág. 38).
- [Fomento 10] Ministerio de Fomento de España Fomento. Los sistemas inteligentes de transporte. Centro de publicaciones del Ministerio, 2010 (pág. 14).
- [Fomento 13] *El transporte y las infraestructuras en España*. Disponible en <http://www.fomento.gob.es/MFOM.CP.Web/handlers/pdfhandler.ashx?idpub=BTW023> (consultado en 2015) (pág. 16).
- [Fomento 14] *El lenguaje del transporte intermodal. Vocabulario ilustrado*. Disponible en [http://www.fomento.gob.es/NR/rdonlyres/17FBCF00-91E0-4761-A11C-88A16277D8A4/1550/01\\_lenguaje\\_transporte\\_intermodal.pdf](http://www.fomento.gob.es/NR/rdonlyres/17FBCF00-91E0-4761-A11C-88A16277D8A4/1550/01_lenguaje_transporte_intermodal.pdf) (consultado en 2015) (pág. 17).
- [Fomento 15] *Plan de infraestructuras, transporte y vivienda*. Disponible en <http://www.lamoncloa.gob.es/espana/eh15/infraestructuras/Documents/PITVI20122024en2015> (pág. 16).
- [Fonseca 99] Frederico T Fonseca and Max J Egenhofer. *Ontology-driven geographic information systems*. In Proceedings of the 7th ACM international symposium on Advances in geographic information systems, pp. 14–19. ACM, 1999 (pág. 46).
- [Frank 01] Andrew U Frank. *Tiers of ontology and consistency constraints in geographical information systems*. International Journal of Geographical Information Science, Vol. 15, No. 7, pp. 667–678, 2001 (pág. 51).

## BIBLIOGRAFÍA

---

- [Fu 05] Gaihua Fu, Christopher B Jones and Alia I Abdelmoty. *Ontology-based spatial query expansion in information retrieval*. In *On the move to meaningful internet systems 2005: CoopIS, DOA, and ODBASE*, pp. 1466–1482. Springer, 2005 (pág. 46).
- [Gan 06] HongCheng Gan, LiJun Sun, JianYang Chen and WenPing Yuan. *Advanced traveler information system for metropolitan expressways in Shanghai, China*. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1944, No. 1, pp. 35–40, 2006 (pág. 20).
- [Giménez 08] Diego M Giménez, Marcela Vegetti, Horacio P Leone and Gabriela P Henning. *PRoduct ONTOlogy: Defining product-related concepts for logistics planning activities*. Computers in Industry, Vol. 59, No. 2, pp. 231–241, 2008 (pág. 44).
- [Gomez 02] A Gomez and O. Corcho. *Ontology languages for the Semantic Web*. Intelligent Systems, IEEE, Vol. 17, No. 1, pp. 54–60, Jan 2002 (pág. 38).
- [Gruber 93] Thomas R Gruber. *A translation approach to portable ontology specifications*. Knowledge acquisition, Vol. 5, No. 2, pp. 199–220, 1993 (pág. 35).
- [Gruber 95] Thomas R Gruber. *Toward principles for the design of ontologies used for knowledge sharing?* International journal of human-computer studies, Vol. 43, No. 5, pp. 907–928, 1995 (pág. 34, 36, 68).
- [Guangzuo 04] Cui Guangzuo, Chen Fei, Chen Hu and Li Shufang. *OntoEdu: a case study of ontology-based education grid system for e-learning*. In *GCCCE2004 International conference*, Hong Kong. Citeseer, 2004 (pág. 44).
- [Guarino 98] Nicola Guarino. *Formal ontology in information systems: Proceedings of the first international conference (fois'98)*, june 6–8, trento, italy, volume 46. IOS press, 1998 (pág. 34).

- [Gunay 14] Arif Gunay, Ozgun Akcay and Mehmet Orhan Altan. *Building a semantic based public transportation geoportal compliant with the INSPIRE transport network data theme*. Earth Science Informatics, Vol. 7, No. 1, pp. 25–37, 2014 (pág. 55, 63, 96).
- [Hausenblas 15] *5-star Open Data*. Disponible en <http://5stardata.info/> (consultado en 2015) (pág. 40).
- [Heath 11] Tom Heath and Christian Bizer. *Linked data: Evolving the web into a global data space*. Synthesis lectures on the semantic web: theory and technology, Vol. 1, No. 1, pp. 1–136, 2011 (pág. 43).
- [Horridge 06] Matthew Horridge, Nick Drummond, John Goodwin, Alan L Rector, Robert Stevens and Hai Wang. *The Manchester OWL Syntax*. In OWLED, volume 216, 2006 (pág. 73).
- [Houda 10] M. Houda, M. Khemaja, K. Oliveira and M. Abed. *A public transportation ontology to support user travel planning*. In Research Challenges in Information Science (RCIS), pp. 127–136, 2010 (pág. 53, 63).
- [Ihobe 12] *Inventario de Emisiones de Gases de Efecto Invernadero*. Disponible en [https://www.euskadi.eus/contenidos/inventario/inventarios\\_gei/es\\_pub/adjuntos/2011.pdf](https://www.euskadi.eus/contenidos/inventario/inventarios_gei/es_pub/adjuntos/2011.pdf) (consultado en 2015) (pág. 22).
- [Jia 11] Haiyang Jia, Minhong Wang, Weijia Ran, Stephen JH Yang, Jian Liao and Dickson KW Chiu. *Design of a performance-oriented workplace e-learning system using ontology*. Expert Systems with Applications, Vol. 38, No. 4, pp. 3372–3382, 2011 (pág. 44).
- [Kumar 05] Praveen Kumar, Varun Singh and Dhanunjaya Reddy. *Advanced traveler information system for Hyderabad City*. Intelligent Transportation Systems, IEEE Transactions on, Vol. 6, No. 1, pp. 26–37, 2005 (pág. 20).

## BIBLIOGRAFÍA

---

- [LATC 11] *Linked open data around the clock (LATC)*. Disponible en <http://latc-project.eu/> (consultado en 2014) (pág. 41).
- [Lemmens 06] RLG Lemmens. *Semantic interoperability of distributed geo-services*. PhD thesis, [Sl: sn], 2006 (pág. 47).
- [LOD2 10] *Linking Open Data 2 (LOD2)*. Disponible en <http://lod2.eu/> (consultado en 2014) (pág. 42).
- [Lorenz 05] Bernhard Lorenz, Hans Jürgen Ohlbach and Laibing Yang. *Ontology of transportation networks*. In Rewerse, 2005 (pág. 46).
- [Mol 02] Annemarie Mol. *The body multiple: Ontology in medical practice*. Duke University Press, 2002 (pág. 44).
- [Niaraki 09] Abolghasem Sadeghi Niaraki and Kyeheun Kim. *Ontology based personalized route planning system using a multi-criteria decision making approach*. Expert Systems with Applications, Vol. 36, No. 2, pp. 2250–2259, 2009 (pág. 20, 52).
- [Norton 10] Barry Norton, Reto Krummenacher, Adrian Marte and Dieter Fensel. *Dynamic linked data via linked open services*. In Workshop on Linked Data in the Future Internet at the Future Internet Assembly, pp. 1–10, 2010 (pág. 2).
- [Noy 01] Natalya F Noy, Deborah L McGuinness *et al.* *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05, 2001 (pág. 37, 68).
- [Noy 04] Natalya F. Noy. *Semantic Integration: A Survey of Ontology-based Approaches*. SIGMOD Rec., Vol. 33, No. 4, pp. 65–70, December 2004 (pág. 39).
- [OGC 12] *GeoSPARQL - A Geographic Query Language for RDF Data*. Disponible en [https://portal.opengeospatial.org/files/?artifact\\_id=47664](https://portal.opengeospatial.org/files/?artifact_id=47664) (consultado en 2015) (pág. 94).

- [Ozbay 99] Kaan Ozbay and Pushkin Kachroo. Incident management in intelligent transportation systems. Artech House Publishers, 1999 (pág. 14).
- [PlanetData 10] PlanetData. *A European network of excellence on large-scale data management*. Disponible en <http://planet-data.eu/> (consultado en 2014) (pág. 41).
- [Raines 08] Alan Raines and Paul Rowley. *Coordinated traffic management through data exchange*. In Road Transport Information and Control-RTIC 2008 and ITS United Kingdom Members' Conference, IET, pp. 1–4. IET, 2008 (pág. 26).
- [Rosse 03] Cornelius Rosse and José LV Mejino Jr. *A reference ontology for biomedical informatics: the Foundational Model of Anatomy*. Journal of biomedical informatics, Vol. 36, No. 6, pp. 478–500, 2003 (pág. 4).
- [Ruckhaus 05] Edna Ruckhaus. *Lógicas descriptivas y ontologías*. Universidad Simón Bolívar, 2005 (pág. 34).
- [Schilit 94] Bill Schilit, Norman Adams and Roy Want. *Context-aware computing applications*. In Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on, pp. 85–90. IEEE, 1994 (pág. 85).
- [Schofer 93] Joseph L Schofer, Asad Khattak and Frank S Koppelman. *Behavioral issues in the design and evaluation of advanced traveler information systems*. Transportation Research Part C: Emerging Technologies, Vol. 1, No. 2, pp. 107–117, 1993 (pág. 20).
- [Speiser 11] Sebastian Speiser and Andreas Harth. *Integrating linked data and services with linked data services*. In The Semantic Web: Research and Applications, pp. 170–184. Springer, 2011 (pág. 2).

## BIBLIOGRAFÍA

---

- [Studer 98] Rudi Studer, V Richard Benjamins and Dieter Fensel. *Knowledge engineering: principles and methods*. Data & knowledge engineering, Vol. 25, No. 1, pp. 161–197, 1998 (pág. 34).
- [Suárez 09] Mari Carmen Suárez and Asunción Gómez. *NeOn methodology for building ontology networks: a scenario-based methodology*. In Proceedings of the International Conference on Software, Services & Semantic Technologies, 2009 (pág. 37, 69).
- [Szczerba 00] R.J. Szczerba, P. Galkowski, I.S. Glicktein and N. Ternullo. *Robust algorithm for real-time route planning*. Aerospace and Electronic Systems, IEEE Transactions on, Vol. 36, No. 3, pp. 869–878, Jul 2000 (pág. 20).
- [UITP 09] International Association of Public Transport UITP. *Doubling the market share of public transport worldwide by 2025*. In 58th UITP World Congress. UITP, 2009 (pág. 1).
- [UITP 14] International Association of Public Transport UITP. *Local public transport trends in the European Union*. In 60th UITP World Congress. UITP, 2014 (pág. 1).
- [Uschold 95] Michael Uschold and Martin King. Towards a methodology for building ontologies. Citeseer, 1995 (pág. 37).
- [Vaccari 09] Lorenzino Vaccari, Pavel Shvaiko and Maurizio Marchese. *A geo-service semantic integration in spatial data infrastructures*. International Journal of Spatial Data Infrastructures Research, Vol. 4, pp. 24–51, 2009 (pág. 45).
- [W3C 09] W3C Standard Facilitates Data Management and Integration. Disponible en <http://www.w3.org/2009/10/owl2-pr> (consultado en 2014) (pág. 69).
- [W3C 15a] Linking Open Data. Disponible en <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> (consultado en 2015) (pág. 42).



- [W3C 15b] *Guia breve de la web semántica*. Disponible en <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica> (consultado en 2015) (pág. 37).
- [Wang 02] Xiaohang Wang, Jin Song Dong, ChungYau Chin, Sanka Ravi-  
priya Hettiarachchi and Daqing Zhang. *Semantic space: An  
infrastructure for smart spaces*. Computing, Vol. 1, No. 2, pp.  
67–74, 2002 (pág. 4).
- [Wei-feng 08] Lv Wei-feng, Cui Wei and Huang Jian. *Research on a DATEX  
II Based Dynamic Traffic Information Publish Platform*. In In-  
telligent Information Technology Application, 2008. IITA'08.  
Second International Symposium on, volume 3, pp. 412–416.  
IEEE, 2008 (pág. 26).
- [Zhang 10] Chuanrong Zhang, Tian Zhao, Weidong Li and Jeffrey P Os-  
leeb. *Towards logic-based geospatial feature discovery and  
integration using web feature service and geospatial semantic  
web*. International Journal of Geographical Information Scien-  
ce, Vol. 24, No. 6, pp. 903–923, 2010 (pág. 45).
- [Zhang 11] Jianwei Zhang, Feixiong Liao, Theo Arentze and Harry Tim-  
mermansa. *A multimodal transport network model for advan-  
ced traveler information systems*. Procedia Computer Science,  
Vol. 5, pp. 912 – 919, 2011 (pág. 52).
- [Zhao 00] Yilin Zhao. *Mobile phone location determination and its im-  
pact on intelligent transportation systems*. Intelligent Transpor-  
tation Systems, IEEE Transactions on, Vol. 1, No. 1, pp. 55–64,  
2000 (pág. 14).
- [Zhao 08] Tian Zhao, Chuanrong Zhang, Mingzhen Wei and Zhong-Ren  
Peng. *Ontology-Based Geospatial Data Query and Integra-  
tion*. In Geographic Information Science, volume 5266 of *Lec-  
ture Notes in Computer Science*, pp. 370–392. Springer Berlin  
Heidelberg, 2008 (pág. 48, 51, 63, 96).