



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Ingeniería Informática

Trabajo Fin de Grado

**Introducir Soporte al Formato
GeoPackage en Herramientas de
Linked Data Geográfico Desarrolladas
por el Grupo de Ingeniería Ontológica**

Autor: Beñat Agirre Arruabarrena
Tutor(a): Oscar Corcho García

Madrid, Marzo 2021

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Ingeniería Informática

Título: Introducir Soporte al Formato GeoPackage en Herramientas de Linked Data Geográfico Desarrolladas por el Grupo de Ingeniería Ontológica

Marzo 2021

Autor: Beñat Agirre Arruabarrena
Tutor: Oscar Corcho García
Departamento de Inteligencia Artificial
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

«Aquí va el resumen del TFG. Extensión máxima 2 páginas.»

Abstract

«Abstract of the Final Degree Project. Maximum length: 2 pages.»

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	1
1.2. Estado del Arte	1
1.2.1. GIS	1
1.2.1.1. Shapefile	2
1.2.1.2. GeoPackage	3
1.2.2. Datos enlazados	3
1.2.3. Portales de Datos abiertos	3
1.2.4. Map4RDF	3
1.2.5. GeoKettle	3
1.3. Herramientas de desarrollo	3
2. Desarrollo	5
2.1. Apartado 1 de capítulo 2	5
2.1.1. Sección 1 de apartado 1 de capítulo 2	5
2.1.1.1. Sub sección 1	5
2.1.1.2. Sub sección 2	5
2.1.2. Sección 2 de apartado 1 de capítulo 2	5
2.2. Apartado 2 de capítulo 2	5
2.3. Apartado 3 de capítulo 2	5
3. Resultados y conclusiones	7
4. Análisis de impacto	9
Bibliografía	12
Anexos	13
A. Anexos	13
A.1. Glosario de términos	13

Capítulo 1

Introducción

La introducción del TFG debe servir para que los profesores que evalúan el Trabajo puedan comprender el contexto en el que se realiza el mismo, y los objetivos que se plantean.

1.1. Objetivos

El objetivo principal del trabajo es introducir soporte al formato GeoPackage en herramientas de Linked Data Geográfico desarrolladas por el Grupo de Ingeniería Ontológica. En el OEG se ha venido tradicionalmente trabajando con el Instituto Geográfico Nacional para la exportación de algunos de sus datos geográficos a formato Linked Data. Un ejemplo se puede encontrar en la web del Instituto Geográfico Nacional. [7]

Recientemente, el Open Geospatial Consortium ha publicado el formato GeoPackage, que tiene el objetivo de convertirse en un estándar para la representación de datos geográficos. El objetivo de este trabajo es el de dar soporte GeoPackage para las herramientas normalmente utilizadas para este tipo de tareas.

- Dar soporte GeoPackage a la herramienta Map4RDF.
- Dar soporte GeoPackage a la herramienta GeoKettle y su plugin para transformar a RDF.
- Realizar un procesado completo de todos los datos del IGN para generar este tipo de formato.

1.2. Estado del Arte

1.2.1. GIS

Los sistemas de información geográfica son herramientas que permiten almacenar y analizar datos geoespaciales. Los sistemas digitales actuales permiten realizar consultas interactivas, añadir entradas a las bases de datos y visualizarlos de manera intuitiva. La información geográfica se puede aplicar a todo

tipo de áreas, entre las que se encuentran la ingeniería, transporte, telecomunicación, economía, sociología... Debido a la gran importancia tanto en el sector público como el privado[5], los estándares abiertos cobran importancia por estar disponibles al público, no tener que pagar licencias y ser consensuados por organizaciones de estándares internacionales. Entre ellas se encuentra el Open Geospatial Consortium(OEG) que se creó en 1994 y agrupa a 521 (en marzo de 2021) miembros de organizaciones públicas y privadas.[6] El OGC trabaja junto con las principales organizaciones de estándares de su ámbito (ISO/TC 211, W3C, IETF...) [4]

Existen diversos formatos de fichero GIS, divididos en **raster** y **vector**. La diferencia es equivalente a la que existe entre imágenes con resolución limitada por el número de píxeles (raster) y las imágenes vectoriales formadas por puntos, líneas y polígonos; con resoluciones infinitas. Cada tipo de formato tiene sus ventajas y desventajas y la elección dependerá del caso de uso. Existen varios formatos vectoriales pero para este trabajo sólo se considerarán el formato *shapefile* y el *GeoPackage* para cumplir los objetivos.

1.2.1.1. Shapefile

El formato ESRI Shapefile (SHP) es un formato de archivo de datos espaciales vectorial desarrollado por la compañía ESRI a principios de la década de 1990. A pesar de ser propietario, la especificación es abierta, y se considera un estándar de facto. Debido a su popularidad, goza de gran compatibilidad con sig. Gracias al uso de un fichero índice, se obtiene una velocidad de lectura alta, y su eficiencia de tamaño produce archivos relativamente pequeños.

Sin embargo, tiene varias desventajas, algunas derivadas del uso del estándar dBase [11] [14]:

1. No tiene definición de sistema de referencia de coordenadas ¹, se puede usar uno pero no es parte estándar de la especificación.
2. Se reparte en múltiples ficheros: es incómodo y lleva a errores al compartirlos.
3. Los nombres de atributos están limitados a 10 caracteres ASCII
4. El número máximo de campos de atributo es 255.
5. Solo admite float, integer, date y text con un máximo de 254 caracteres.
6. No se puede especificar conjunto de caracteres de la BBDD.
7. El tamaño está limitado a 4GB.
8. No admite valores NULL
9. No hay forma de describir las relaciones topológicas en el formato.

¹Un sistema de coordenadas es un sistema de referencia que se utiliza para representar la ubicación de entidades geográficas, imágenes y observaciones (como las localizaciones GPS) dentro de un marco geográfico común. Los sistemas de coordenadas permiten a los datasets geográficos utilizar ubicaciones comunes para la integración de datasets. [15]

Introducción

10. Solamente puede almacenar una geometría por archivo.
11. Utiliza una estructura de datos de tabla plana, sin jerarquías, relaciones ni estructura en árbol.
12. El soporte 3D es muy limitado.

1.2.1.2. GeoPackage

GeoPackage es un formato GIS implementado en SQLite publicado por el OEG en 2014. [12]

El formato geotools tiene las siguientes ventajas [14]:

- Es abierto, no propietario, basado en estándares, independiente de plataformas, portable y compacto.
- Gracias a SQLite puede almacenar datos grandes (hasta 140TB)[17] y los atributos de las geometrías pueden contener nombres muy largos.
- Dispone de índices espaciales basados en R-trees [16] que incrementan la velocidad de búsquedas espaciales y su visualización en los SIG de escritorio.
- Todo el contenido se almacena en un único archivo .gpkg que puede almacenar multitud de tipos de geometrías
- Soporta el uso directo, para acceder a los datos de GeoPackage de forma «nativa» sin traducciones de formato intermedio.
- GeoPackage es soportado por GDAL[18], la librería de conversión de datos utilizada por multitud de programas GIS (incluido GeoKettle), y los principales programas GIS.

1.2.2. Datos enlazados

Explicar qué son los datos enlazados y en específico RDF

1.2.3. Portales de Datos abiertos

Poner ejemplo del IGN.

1.2.4. Map4RDF

1.2.5. GeoKettle

Explicar qué es GeoKettle y su plugin.

1.3. Herramientas de desarrollo

Docker

La imagen de docker de GeoKettle

Capítulo 2

Desarrollo

Capítulo dedicado a describir el desarrollo del Trabajo realizado. De acuerdo con el tutor, este capítulo puede tener distintas estructuras, e incluso pueden existir varios capítulos.

Todos los capítulos deben empezar en una página nueva.

Los apartados dentro de los capítulos se numeran de forma jerárquica, pero siempre deben estar alineados al margen izquierdo. Ejemplo:

2.1. Apartado 1 de capítulo 2

2.1.1. Sección 1 de apartado 1 de capítulo 2

2.1.1.1. Sub sección 1

2.1.1.2. Sub sección 2

2.1.2. Sección 2 de apartado 1 de capítulo 2

2.2. Apartado 2 de capítulo 2

2.3. Apartado 3 de capítulo 2

Capítulo 3

Resultados y conclusiones

Resumen de resultados obtenidos en el TFG. Y conclusiones personales del estudiante sobre el trabajo realizado.

Capítulo 4

Análisis de impacto

En este capítulo se realizará un análisis del impacto potencial de los resultados obtenidos durante la realización del TFG, en los diferentes contextos para los que se aplique:

- Personal
- Empresarial
- Social
- Económico
- Medioambiental
- Cultural

En dicho análisis se destacarán los beneficios esperados, así como también los posibles efectos adversos.

Se recomienda analizar también el potencial impacto respecto a los Objetivos de Desarrollo Sostenible (ODS), de la Agenda 2030, que sean relevantes para el trabajo realizado (ver enlace)

Además, se harán notar aquellas decisiones tomadas a lo largo del trabajo que tienen como base la consideración del impacto.

Bibliografía

- [1] Publicaciones utilizadas en el estudio y desarrollo del trabajo. Hay que utilizar un sistema internacional para referencias bibliográficas, de acuerdo con las indicaciones del tutor. Por ejemplo, el **sistema de IEEE**.
- [2] A. de León, V. Saquicela, LM. Vilches-Blázquez, B. Villazón-Terrazas, F. Priyatna and Oscar Corcho, “Geographical Linked Data: a Spanish Use Case”, in *Proceedings of the In I-SEMANTICS '10 6th International Conference on Semantic Systems*
- [3] Espinoza-Arias, P., García-Delgado, M., Corcho, O. et al. “A sustainable process and toolbox for geographical linked data generation and publication: a case study with BTN100.”, in *Open geospatial data, softw. stand. 4, 2 (2019)*.
- [4] OGC, “OGC’s Role in the Spatial Standards World”, in *An Open GIS Consortium (OGC) White Paper*
- [5] OGC, ISO, TC/ 211, IHO, “A Guide to the Role of Standards in Geospatial Information Management”, in *Fifth Session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). Held from 3-7 August 2015 at the United Nations Headquarters in New York*.
- [6] Miembros del OGC, “<https://www.ogc.org/ogc/members>”,
- [7] Web de datos del Instituto Geográfico Nacional, <http://datos.ign.es/>
- [8] Leon, Alexander de; Wisniewski, Filip; Villazón-Terrazas, Boris y Corcho, Oscar “ Map4rdf - Faceted Browser for Geospatial Datasets”, in *PMOD workshop, 2012*
- [9] Web de Map4rdf “<https://oeg-upm.github.io/map4rdf/>”,
- [10] “ESRI Shapefile Technical Description”, *An ESRI White Paper—July 1998*
- [11] ESRI, “Geoprocessing considerations for shapefile output”, *Arcgis Desktop 9.3 Help*
- [12] OGC® GeoPackage Encoding Standard 1.3
“<http://www.geopackage.org/spec/>”,
- [13] Tim Berners-Lee “Linked Data Design Issues”, in *W3C 2009*

- [14] Aurelio Morales “Di no al Shapefile y sí al GeoPackage”, in *mappingis.com* 2018
- [15] ESRI ArcMap 10.8 Docs, “¿Qué son las proyecciones cartográficas?”, in <https://desktop.arcgis.com/es/arcmap/latest/map/projections/what-are-map-projections.htm>
- [16] Antonin Guttman “1984. R-trees: a dynamic index structure for spatial searching”, in *Proceedings of the 1984 ACM SIGMOD*
- [17] “Limits In SQLite”, in <https://www.sqlite.org/limits.html>
- [18] “GDAL”, in <https://gdal.org/>
- [19] ab “cd”, in *ef*

Apéndice A

Anexos

A.1. Glosario de términos

- *OGC*: Open Geospatial Consortium
- *GIS*: Geographic Information System
- *ESRI*: Environmental Systems Research Institute
- *IGN*: Instituto Geográfico Nacional
- *ab*: cd
- *GDAL*: Geospatial Data Abstraction Library