

[1017]

KONUMSAL VERİLERİN BAĞLANTILI VERİ OLARAK YAYINLANMASI

Gülten KARA¹, Çetin CÖMERT²

¹Yrd. Doç. Dr., Karadeniz Teknik Üniversitesi, Harita Mühendisliği Bölümü, 61080, Trabzon, gispir@ktu.edu.tr

²Prof. Dr., Karadeniz Teknik Üniversitesi, Harita Mühendisliği Bölümü, 61080, Trabzon, ccmert@ktu.edu.tr

ÖZET

Semantik Web uygulamaları için gerekli konumsal verilerin büyük bir çoğunluğu veri tabanlarında depolanmaktadır. Diğer yandan ontolojiler "Semantik Web" in gerçekleştirilmesi için en önemli role sahiptir. Veri tabanları ve ontolojiler arasındaki ilişkilerin belirlenmesi problemi Semantik Web alanında en popüler araştırma konularından biridir. Ontoloji-veri tabanı eşleştirme olarak adlandırılan bu konu, farklı araştırma konularını da gündeme getirmektedir. Bu bağlamda, mevcut veri tabanlarından ontoloji oluşturulması, mevcut ontolojilerin veri tabanlarıyla ilişkilendirilmesi, veri kaynaklarının Bağlantılı Veri olarak yayınlanması, vb., konularda çalışmalar gerçekleştirilmiştir. Özellikle OpenStreetMap, Geonames, Wikipedia gibi projelerin temel altlık teşkil edecek konumsal veri sağlaması, bu veri kaynaklarının Semantik Web uygulamaları için kullanılmasını zorunlu hale getirmiştir. Konumsal arenada, LinkedGeoData, GeoLinkedData, OS Open Data, vb., proje ve girişimlerle konumsal veriler Bağlantılı Veri olarak yayınlanmıştır. Konumsal Bağlantılı Veri olarak adlandırılan bu veriler, Bağlantılı Veri teknolojileri kullanılarak farklı veri kaynakları üzerinde semantik sorgulama imkânı sağlayacaktır. Bu noktadan hareketle, konumsal veri altyapılarında, konumsal verilerin Bağlantılı Veri olarak yayınlanmasında gerekli teknolojik altyapının kurulması için gereksinimlerin belirlenmesi gerekir.

Konumsal veriyi Bağlantılı Veri olarak yayınlamak için en önemli gereksinim şüphesiz ki dönüşüm araçlarıdır. Geometry2RDF, TripleGeo ve GeoTriples gibi açık kaynak kodlu araçlar mevcuttur. Bu çalışmada yazılım araçları kullanılarak KGM-Yol verisi Bağlantılı Veri olarak yayınlanmıştır. Bunun yanı sıra yayınlanan Bağlantılı Veriler karşılaştırılarak araçların birbirlerine göre üstünlükleri irdelenmiştir.

Anahtar Sözcükler: Bağlantılı veri, Konumsal veri, Ontoloji, Semantik Web.

ABSTRACT

PUBLISHING GEOSPATIAL DATA AS LINKED DATA

A majority of the necessary spatial data for Semantic Web applications are stored in the database. Ontologies, on the other hand, are the most important role for performing "Semantic Web". The problem of determination the relationships between ontologies and databases is one of the most popular research topics in the field of the Semantic Web. This subject named as Ontology-database mapping brings to the agenda different research topics. In this context, It have been performed studies such as the creation of an ontology from an existing database, the discovery of mappings between database and existing ontology and the publishing of spatial data sources as Linked Data. Especially, that projects such as OpenStreetMap, Geonames, Wikipedia, etc. provide spatial data used as based map, has make compulsory the use of these data sources for the Semantic Web applications. In geospatial arena, geospatial data were published as linked data with projects and initiatives such as LinkedGeoData, GeoLinkedData, OS OpenData, etc. These data called as Geospatial Linked Data will enable semantic queries on the different data sources using Linked Data Technologies. From this point forward, for the publishing spatial data as Linked Data should be determined the requirements for the establishment necessary technological infrastructure in the spatial data infrastructure.

The most important requirement for the publishing the spatial data as Linked Data, doubted, is the transformation tools. It is available open source tools such as Geometry2RDF, TripleGeo, GeoTriples. In this study, General Directorate of Highways-Road Data was published as Linked Data to use these tools. As well as, published Linked Data were compared and examined the advantage relative to each other of tools.

Keywords: Linked data, Geospatial Data, Ontology, Semantic Web.

1.GİRİŞ

Gerek konumsal veri tabanlarında gerekse web üzerinde konumsal verinin her geçen gün artmasıyla birlikte Semantik Web uygulamalarında konumsal verilerin kullanılmasına yönelik çalışmalar da hız kazanmıştır. Web üzerinde veya konumsal veri tabanlarında tutulan veriler farklı dosya formatlarındadır. OGC (Open GeoSpatial Consortium) tarafından yayınlanan standartlar veya INSPIRE Direktifi kapsamında gerçekleştirilen konumsal veri altyapısını kurmak için gerçekleştirilen girişimler, konumsal verinin birlikte işlerliğine önemli katkı yapmıştır. Bunun yanı sıra kitle kaynaklı konumsal veriler (crowdsourced geodata) giderek yaygınlaşmaktadır. Bu bağlamda

güncel konumsal veri sunan OpenStreetMap¹, GeoNames², Wikipedia gibi önemli projeler gerçekleştirilmiştir. Kitle kaynaklı girişimlerden bir diğeri, verileri diğeri web kaynakları ile ilişkilendirmek amacıyla Wikipedia dan yapısal bilginin çıkarılmasıyla oluşturulan DBpedia³ dir.

Kitle kaynaklı konumsal verilerin gerçekleştirilen girişimlerle web üzerinde sunulması Açık Konumsal Veri anlayışının yaygınlaşmasına öncülük etmiştir. Buna paralel olarak konumsal gösterim ve Bağlantılı Veri paradigmasına göre farklı veri kaynaklarının semantik tanımları kullanılarak çıkarsama gerçekleştirilmesi Semantik Web uygulamaları için son derece önemlidir. Konumsal verilerin Semantik Web uygulamalarında kullanılması için web üzerinde veya konumsal veri tabanlarında tutulan verilerin ontoloji dillerinde tanımlanması gerekir. RDF ve OWL gibi ontoloji dilleri ve SPARQL gibi sorgulama dilleri sayesinde ilişkisel verileri RDF ye dönüştürmek için çok sayıda proje girişim ve akademik çalışma yapılmıştır.

W3C RDB2RDF Çalışma Grubu herhangi bir ortak model kullanmaksızın ilişkisel veri tabanındaki tablolardan otomatik olarak RDF üretmek için Doğrudan Eşleştirmeyi (Direct Mapping) ve kelime hazinesi kavramlarını veri tabanı şemasına ilişkilendirmek için R2RML (RDB to RDF Mapping Language) önerir. R2RML W3C Tavsiyesi ilişkisel tabloların RDF ye eşleştirilmesini, RDF veri modelinin görüntülenmesini ve sorgulanmasını sağlamak için bir RDF notasyonu tanımlar (W3C, 2012). İlişkisel veri tabanlarını RDF ye dönüştürmek için çok sayıda yazılım aracı mevcuttur. Bunlardan bazıları, Triplify (Auer vd., 2009), D2R Server⁴, Virtuoso RDFizer Middleware⁵, OpenRefine⁶ RDF genişletmesi, vb. dir. Bu yazılım araçlarının yanı sıra farklı dosya formatlarındaki verileri RDF ye dönüştürmek için çok sayıda yazılım geliştirilmiştir⁷. Ancak belirtilen yazılım araçlarının hiçbirisinde konumsal veri tutma, sorgulama veya dönüştürme desteği yoktur.

Konumsal veri gerçek dünya varlıklarının gösterimini ve onlar arasındaki topolojik ve konumsal ilişkilerin açıkça tanımlanması yönüyle konumsal olmayan verilerden farklılaşır. Konumsal verilerin birlikte işlerliğini sağlamak için “Bağlantılı Açık Veri” olarak yayınlanmasına yönelik proje, girişim ve akademik çalışmalar mevcuttur. Konumsal verilerin geometri tanımları için ortak bir standart oluşturulması amacıyla OGC tarafından GeoSPARQL⁸ standardı (OGC, 2012) önerilmiştir. GeoSPARQL, konumsal veri gösterimi ve sorgulama için OGC tarafından standart olarak kabul edilmiştir. SPARQL standardı konumsal veri tanımlama ve topolojik ilişkilerle genişletilmiştir. OGC GeoSPARQL standardı, GML (Geography Markup Language) ve WKT (Well Known Text) gibi RDF de geometri ve detayları tanımlamak için belirli bir ontoloji önerir. GeoSPARQL Ontolojisi⁹ konumsal verileri tanımlamak için kullanılabilir durumdadır.

Web üzerinde Açık Bağlantılı Verileri toplamak için bir veri yönetimi platformu olan Datahub¹⁰ da Ordnance Survey (OS), National Geographical Institute of Spain, U.S. Geological Survey ve daha bir çok önemli konumsal veri sağlayıcıları tarafından yayınlanan Konumsal Bağlantılı Veriler yer almaktadır. Bunların yanı sıra OpenStreetMap, GeoNames, gibi önemli girişimlerde bu hareketin öncülerindedir. Diğeri bir girişim, Geo.LinkedData.es, İspanya konumsal verisini Semantik Web ortamında sunmak için gerçekleştirilmiştir. Ayrıca LinkedGeoData¹¹ olarak adlandırılan girişimin amacı ise Semantik Web i OpenStreetMap projesi ile elde edilen konumsal verilerle zenginleştirmektir. Yani, Semantik Web e konumsal bir boyut eklemeyi amaçlar. OpenStreetMap verisini, RDF verisi olarak sunmak için SPARQL Endpoint desteği ve konumsal semantik verinin ilişkilendirilmesi ve eşleştirmelerin gerçekleştirilmesi için yazılım desteği ile birlikte, RDF ye eşleştirmek için esnek bir platform sunar. OS Bağlantılı Veri¹² girişimi ile İngiltere’nin Ulusal Harita Kurumu (Ordnance Survey) tarafından üretilen konumsal verilerin Bağlantılı Veri olarak sunulması amaçlanmıştır. Buna paralel olarak, GeoLinkedData(.es)¹³ İspanya’da konumsal veri ile Semantik Web i zenginleştirmek için açık bir girişimdir. Girişim kapsamında konumsal verileri RDF ye dönüştürmek için Geometry2RDF yazılım aracı geliştirilmiştir. Konumsal verilerin en temel kaynağını teşkil eden kadastral verileri dikkate alan Saavedra vd., (2014) LADM (ISO 19152) kullanarak kadastral alanda Bağlantılı Veri ilkelerine uygun olarak kadastral verilerin LOD Bulutunda ilgili veri kaynakları ile referanslandırılması ve Bağlantılı Veri olarak yayınlanması üzerine bir çalışma

¹ <https://www.openstreetmap.org/#map=5/51.500/-0.100>

² <http://www.geonames.org>

³ <http://wiki.dbpedia.org>

⁴ <http://d2rq.org/d2r-server>

⁵ <http://docs.openlinksw.com/virtuoso/virtuososponger.html>

⁶ <http://openrefine.org>

⁷ <https://www.w3.org/wiki/ConverterToRdf>

⁸ <http://www.opengeospatial.org/standards/geosparql>

⁹ http://schemas.opengis.net/geosparql/1.0/geosparql_vocab_all.rdf

¹⁰ <http://datahub.io/es/dataset?tags=lod>

¹¹ <http://linkedgeodata.org/About>

¹² <http://data.ordnancesurvey.co.uk>

¹³ <http://geo.linkeddata.es>

yapmıştır. Bütün bu gelecek vadeden girişimlere rağmen Bağlantılı Veri olarak yayınlanan konumsal veri, KVA ve Açık Veri platformları ile kıyaslandığında çok sınırlı düzeydedir. Çünkü geleneksel konumsal veri formatlarından RDF ye dönüşüm için sınırlı sayıda yazılım aracı mevcuttur. Ayrıca semantik referanslandırma için LOD (Linking Open Data) Bulutundaki konumsal veri kaynaklarının yeterli sayıya ulaştırılması ve konumsal verilerin Bağlantılı Veriye dönüşümü için daha fazla yazılım aracının geliştirilmesi gerekir.

Bu çalışmada, konumsal veriyi RDF ye dönüştürmek için mevcut yazılım araçları incelenecektir. Uygulama senaryosu kapsamında mevcut yazılım araçları ile konumsal veriler RDF ye dönüştürülecek ve sonuçlar irdelenecektir. Ayrıca bu çalışma kapsamında mevcut yazılımların özellikleri uygulama senaryosu sonuçları bağlamında karşılaştırılarak, konumsal verilerin RDF ye dönüşümünde kullanılacak yazılım aracının belirlenmesinde kullanıcılara önemli katkı sağlaması amaçlanmaktadır.

1.2.Bağlantılı Konumsal Veri

Bağlantılı Veri genel olarak veri kaynaklarındaki veriler arasında semantik ilişkilerin kurulması ve web üzerinde bu verilerin kullanılması anlamına gelir (Bizer vd., 2009; Heath ve Bizer, 2011). Diğer bir deyişle, Bağlantılı Veri, veri kaynaklarındaki veriler üzerinde semantik sorgulamaların gerçekleştirilmesi için bu verilerin Bağlantılı Veri ilkelerine göre yapıları veri olarak yayınlama yöntemidir. Bağlantılı Veri, veri kaynakları arasında birlikte işliğin sağlanması ve veriler üzerinde semantik sorgulamaların gerçekleştirilmesi amacıyla taşır. Bağlantılı Veri teknolojileri ise veri setleri ile ilişkilendirilen verilerin semantik olarak işlenmesini, çıkarılmasını, birleştirilmesini ve görselleştirilmesini sağlar. Bu bağlamda Bağlantılı Veri İlkelerinin (Berners-Lee, 2006) yaygınlaşması ve benimsenmesi için Linking Open Data¹⁴ Projesi kapsamında Bağlantılı Açık Veri Bulutu (LOD Cloud) oluşturulması amaçlanmıştır. Projenin temel amacı, mevcut veri setlerini (GeoNames, Dbpedia, Foaf, MusicBrainz, vb.) tanımlayarak veri kaynaklarındaki verilerin Bağlantılı Veri ilkelerine göre dönüşümünün sağlanması ve web üzerinde yayınlanmasıdır. LOD Cloud, 2014 yılı itibariyle farklı alanlarda 1014 veri seti içermektedir.

Konumsal veriler vektör veri tabanlarında tematik öznitelikleri ve geometrik bilgileri ile depolanmaktadır. Dolayısıyla Semantik Web uygulamalarında gerekli sorgulama ve analizler için tematik özniteliklerin yanı sıra konumsal özniteliklerin de RDF ye dönüştürülmesi gerekir. Konumsal verinin Bağlantılı Açık Veri (Linked Open Data) olarak yayınlanması için belirli standartlara bağlı kalınarak Semantik Web Teknolojileri ile gerçekleştirilir. Bağlantılı Verinin elde edilmesi için gerekli metodoloji ve teknolojilerin belirlenmesi konusunda akademik çalışmalar yapılmıştır. Konumsal verilerin Bağlantılı Açık Veri ye dönüşümü Bağlantılı Veri İlkeleri nin uygulanmasını gerektirir. Konumsal Bağlantılı Açık Veri elde edilmesi ile semantik olarak konumsal veri değişiminin gerçekleştirilmesi sağlanacaktır. Ana amaç, farklı konumsal veri kaynakları arasında semantik birlikte işliğin artırmak ve Bağlantılı Konumsal Veriler üzerinden sorguların gerçekleştirilmesi ve çıkarsama sonuçlarının web üzerinde sunulmasıdır.

Konumsal verilerin Bağlantılı Veri olarak yayınlanması için konumsal veri tabanlarındaki verilerin ontolojilerdeki ilgili sınıf ve özniteliklerle ilişkilendirilmesi gerekir. Ontoloji Mimarisinde Veri Ontolojisi olarak adlandırılan düzeyde Veri tabanı şemasından elde edilen ontoloji ile veri kaynağındaki tablo ve öznitelikler ilişkilendirilmesi gerekir. Bu nedenle veri ontolojisine ihtiyaç vardır. Uygulama senaryosu kapsamında veri ontolojisini oluşturmak için Protege 5.0 yazılımı kullanılmıştır.

Konumsal veri tabanlarındaki verilerin RDF dosyaları üzerinde sorgulamaların gerçekleştirilmesi için veri tabanlarında depolanması gerekir. Konumsal verilerden elde edilen RDF üçlülerini depolamak için Triple Store olarak adlandırılan veri tabanlarına ihtiyaç vardır. Konumsal RDF verilerini depolamak için etkin olarak kullanılan Triple Store lardan bazıları, AllegroGraph¹⁵, Oracle Spatial and Graph¹⁶, GraphDB¹⁷, Parliament¹⁸, Strabon¹⁹, uSeekM²⁰, Virtuoso RDF Triple Store²¹, vb. dir.

¹⁴ <http://linkeddata.org>

¹⁵ <http://franz.com/agraph/allegrograph/>

¹⁶ <http://www.oracle.com/technetwork/database-options/spatialandgraph/overview/spatialandgraph-1707409.html>

¹⁷ <http://ontotext.com/products/graphdb/>

¹⁸ <http://parliament.semwebcentral.org>

¹⁹ <http://www.strabon.di.uoa.gr>

²⁰ <https://www.w3.org/2001/sw/wiki/USeekM>

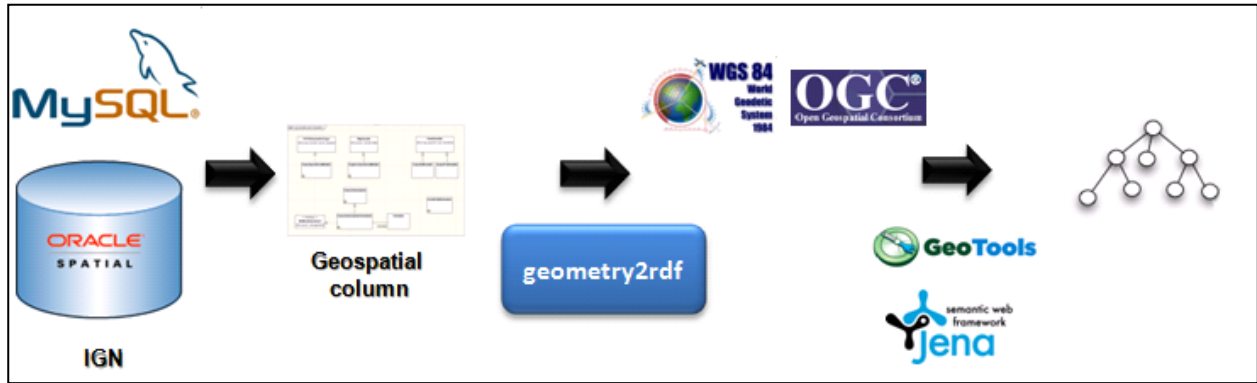
²¹ <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VOSRDF>

2.DÖNÜŞÜM ARAÇLARI

Konumsal veri tabanlarından RDF dosyalarını elde etmek için Geometry2RDF, TripleGeo, GeoTriples, shp2GeoSPARQL, GeomRDF gibi çeşitli yazılım araçları geliştirilmiştir. Bu bölümde konumsal veriyi RDF ye dönüştürmek için geliştirilen yazılım araçlarının mimarisi, özellikleri, olumlu ve olumsuz yönleri incelenmiştir. Konumsal veri tabanlarında tutulan verilerin uygulama senaryosu kapsamında belirtilen yazılım araçlarıyla ESRI shapefile formatından RDF dosyasına dönüşümü gerçekleştirilmiştir. Sonuçlar irdelenerek uygulama kapsamına göre konumsal verinin RDF dosyasına dönüşümünde uygun yazılım aracının seçilmesi esnasında kriterlerin belirlenmesine yardımcı olacaktır. Mevcut yazılım araçlarının özellikleri incelenerek Karayolu Yol verisi RDF ye dönüştürülmüştür. Dönüşüm sonrası üretilen RDF dosyaları karşılaştırılarak yazılımların özellikleri irdelenmiştir ve sonuçlar karşılaştırılmıştır.

2.1.Geometry2RDF

Geometry2RDF²² GML veya WKT olarak mevcut olan konumsal bilgiden RDF dosyalarını oluşturmak için Geo.LinkedData.es girişimi kapsamında oluşturulan Java tabanlı bir yazılımdır. Konumsal veriyi RDF verisine dönüştürmek için geliştirilen ilk yazılımdır. GML ve WKT formatındaki dosyalar GeoTools ile işlenir. Yazılım girdi olarak ESRI shapefile, konumsal VTYS (Oracle, PostgreSQL, MySQL, vb.) lerdeki dosyayı alır, GeoTools kullanarak GML formatına dönüştürülür ve Jena kullanılarak NeoGeo Kelime Hazinesine (NeoGeo Vocabulary) uygun RDF dosyası üretilir. Yazılımın güncel versiyonu Oracle konumsal veri tabanları ile çalışmaktadır. Varsayılan Koordinat Referans Sistemi WGS 84 tür. Şekil 1 de gösterilen Geometry2RDF mimarisi kendi içinde bir platform veya bir kütüphane olarak çalışmak için yeterince modülerdir.



Şekil 1. Geometry2RDF Mimarisi²³

Uygulama kapsamında Geometry2RDF yazılım aracı kullanılarak Karayolu verisi (shapefile), RDF formatına dönüştürülmüştür. Dönüşüm sonrasında oluşan dosya incelendiğinde;

Girdi dosyası olarak konumsal veri formatları ESRI Shapefile, GML ve konumsal VTYS kullanılabilir.

Dönüşüm sonrası oluşturulan dosya yalnızca RDF/XML formatını desteklemektedir. Yani farklı RDF notasyonları ile çıktı dosyası elde edilemez.

Tematik öznitelikler (konumsal olmayan öznitelikler) dönüşüm kapsamında değildir. Yazılım yalnızca geometri bilgilerinin dönüşümünü gerçekleştirdiği için tematik öznitelikler dönüşüm sonrası oluşan dosyaya aktarılmamaktadır.

Dönüşüm sonrasında oluşan RDF dosyası yalnızca NeoGeo Kelime Hazinesi ile uyumludur. Yani GeoSPARQL Kelime Hazinesi ni desteklememektedir.

Farklı koordinat sistemleri arasında dönüşüm gerçekleştirilebilir.

2.1.shp2GeoSPARQL

shp2GeoSPARQL²⁴, ISO19152 (Land Administration Domain Model) ve GeoSPARQL kullanarak kadastr alanında ESRI shapefile formatındaki dosyaları RDF ye dönüştürmek amacıyla geliştirilmiştir ve Geometry2RDF yazılımının bir genişletmesidir. shp2GeoSPARQL kullanılarak RDF dönüşümü sonrasında elde edilen geometriler GeoSPARQL standardı ile uyumludur. Shp2GeoSPARQL, Veri setlerindeki geometrik bilgiyi RDF ye dönüştürmek için Geometry2RDF ye dayanarak geliştirilen bu yazılımın eksikliklerini gidermek amacıyla

²² <https://github.com/boricles/geometry2rdf/tree/master/Geometry2RDF>

²³ <http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/technologies/151-geometry2rdf/>

²⁴ <https://github.com/jasaavedra/shp2geosparql>

geliştirilen bir yazılım aracıdır. En önemli farkı GeoSPARQL standardını desteklemesidir.

Shp2GeoSPARQL yazılımı ile gerçekleştirilen konumsal veri-RDF dönüşümü sonrasında oluşan dosya incelendiğinde;

Dönüşüm sonrası oluşan RDF verisi GeoSPARQL Kelime Hazinesi ile uyumludur.

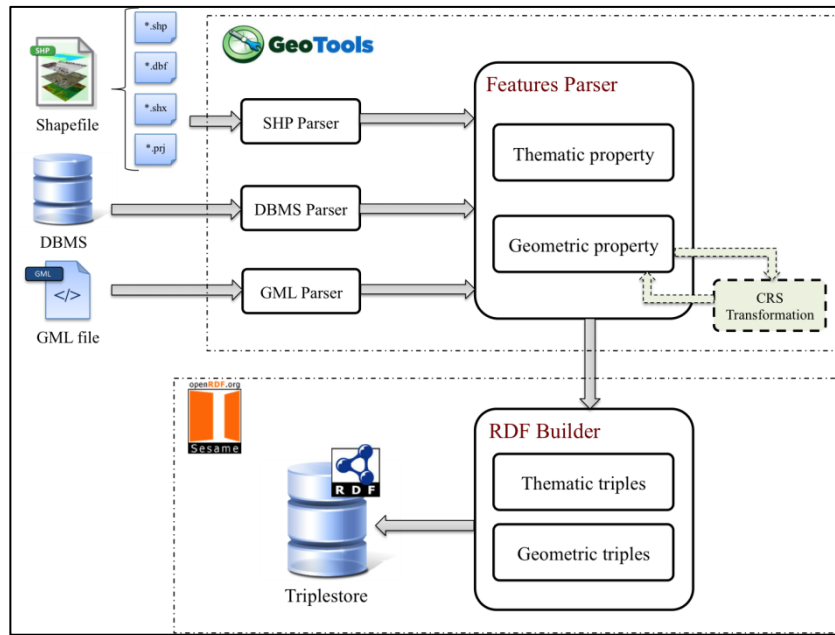
Yazılım konumsal verilerin dönüşümünü desteklerken tematik verilerin dönüşümünü gerçekleştirmemektedir.

Girdi konumsal veri formatı yalnızca ESRI Shapefile dir.

Çıktı dosya formatı yalnızca RDF dir.

2.3. GeomRDF

GeomRDF25 geleneksel CBS formatlarındaki konumsal veri setlerini RDF ye dönüştüren bir yazılım aracıdır. GeoSPARQL ve NeoGeo Kelime Hazinesi ile uyumludur. Veriler herhangi bir koordinat sistemine dönüştürülebilir. ESRI Shapefile Konumsal VTYS ve GML gibi formatların yanında CSV gibi koordinat bilgilerini içeren metin dosyaları da girdi dosyası olarak kullanılabilir. GeomRDF DataLift26 platformunda bir modül olarak geliştirilmiştir.



Şekil 2. GeomRDF Mimarisi (Hamdi vd., 2105).

GeomRDF kullanılarak konumsal verilerin RDF ye dönüşüm işleminden sonra oluşan dosya incelendiğinde;

GeoSPARQL ve NeoGeo Kelime Hazinesi ile uyumludur.

Konumsal ve bütün tematik özniteliklerin dönüşümü gerçekleştirilmektedir.

Koordinat sistemleri arasında dönüşüm gerçekleştirilebilir.

Girdi dosya formatı olarak ESRI Shapefile, konumsal VTYS ve GML desteklenmektedir.

Farklı koordinat sistemleri arasında dönüşüm mümkündür.

Çıktı dosyası RDF/XML, Turtle, N-Triple formatlarını destekler.

2.4. TripleGeo

TripleGeo²⁷ IMIS (Institute for the Management of Information Systems Center) tarafından EU/FP7 GeoKnow Projesi²⁸ kapsamında geliştirilen bir yazılım aracıdır. Yazılım, Geometry2RDF den genişletilmiş olup bir ETL (Extract-Transform-Load) aracı olarak görev yapar. TripleGeo konumsal verileri RDF ye dönüştürülmesi sırasında girdi olarak konumsal tabloların yanı sıra ESRI Shapefile dosyalarını alır. GeoSPARQL standardını desteklediği için GML ve Simple Feature da yer alan bütün geometri tiplerini destekler.

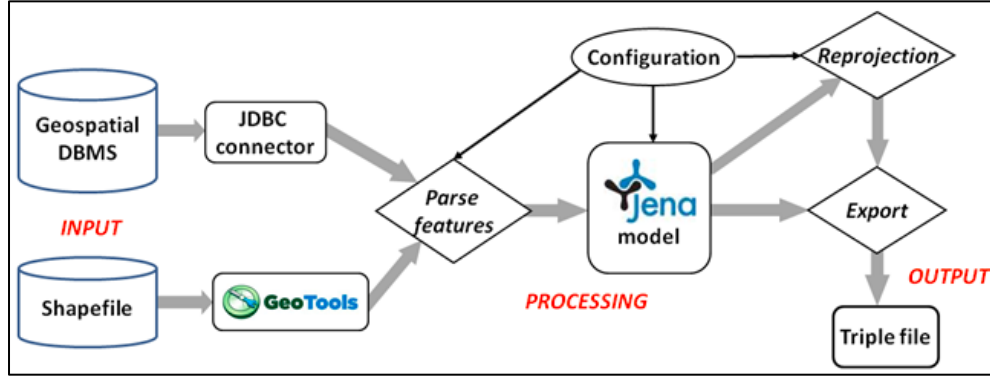
²⁵ <https://github.com/fhamdi/GeomRDF>

²⁶ <http://datalift.org>

²⁷ <https://github.com/GeoKnow/TripleGeo>

²⁸ <http://geoknow.eu/Welcome.html>

Yazılım açık kaynak kodlu olup konumsal veri tabanlarındaki verilerden RDF üçlülerini (triples) elde etmek amacıyla geliştirilmiştir. TripleGeo kendisi gibi açık kaynak kodlu yazılım olan Geometry2RDF ye dayanır. TripleGeo, yazılımının Geometry2RDF ye üstünlüğü GeoSPARQL standardı ile uyumlu olması ve konumsal olmayan özniteliklerin RDF ye dönüştürülmesidir. Girdi dosyası olarak konumsal detayları depolamak için geniş ölçüde kullanılan ESRI Shape dosyalarından, GML ve KML formatında depolanan coğrafi verilerden, INSPIRE Ek-1 de yer alan 7 veri teması için GML formatındaki veri setlerinden, Oracle Spatial, POSTGIS (PostgreSQL için konumsal modül), MySQL, IBM DB2 (with Spatial extender) gibi konumsal olarak etkin veri tabanı yönetim sistemleri ile depolanan verilerden geometri bilgilerine doğrudan erişebilir.



Şekil 2. TripleGeo Mimarisi (Patroumpas vd., 2014).

Şekil 2 de verilen TripleGeo mimarisinin çalışma prensibi aşağıda verilmiştir:

Geometrik detaylara erişmek için kaynak veriye *Bağlantılar (Connectors)* gereklidir. Konumsal verilerin ilişkisel veri tabanlarında olması durumunda JDBC sürücülerini ile ESRI shapefile dosyalarında ise GeoTools kütüphanesi gerekli erişime imkân sağlar.

Configuration dosyası birçok aşamayı kontrol eden birçok özelliği listeler. Girdi dosyasına nasıl erişileceği, hangi veri formatında olduğu, hangi geometrik gösterimin kullanılacağı, koordinat sistemi dönüşümünün yapıp yapılmayacağı, vb. gibi özellikleri içerir. Aynı kontroller çıktı dosyası içinde geçerlidir.

Kullanıcı tarafından belirtilen özelliklere göre her bir kayıt için geometri gösterimleri uygun gösterimlere dönüştürülür. Detaylara ait (tematik öznitelikler de veri tiplerine uygun olarak RDF dosyasına aktarılır.

Jena Modeli, üretilen RDF üçlülerini içeren bütün bilgileri hafızada tutmak için kullanılır.

İsteğe bağlı olarak konumsal koordinat sistemleri arasında dönüşüm gerçekleştirilir.

Üretilen üçlülerin çıktı dosyasına aktarılması *Jena* tarafından gerçekleştirilir (Patroumpas vd., 2014).

Konumsal verinin bir tabloda (veri tabanı) olması durumunda, bir GML/KML dosyasında veya bir shapefile dosyasında bulunmalıdır. Konumsal temsille ilgili olarak GeoSPARQL, WGS84 RDF GeoPosition Vocabulary, ve Virtuoso RDF Vocabulary standartlarında üçlüler export edilebilir. Üçlüler GeoSPARQL standardı, WGS84 Vocabulary ve Virtuoso Vocabulary e göre export edilebilir. Sonuçlar belirtilen dosyaya aktarılır veya depolamak için "Triple Store" a import edilebilir (Patroumpas vd., 2014).

TripleGeo yazılımı kullanılarak Karayolu Verisinin RDF ye dönüştürülmesi sonrasında;

Girdi dosyası olarak ESRI shapefile, GML, KML ve konumsal VTYS ler kullanılabilir.

Çıktı formatı olarak RDF/XML (varsayılan), RDF/XML-ABBREV, N-TRIPLES, N3, TURTLE (TTL) formatları desteklenmektedir.

Konumsal özniteliklerin dönüşümü gerçekleştirilirken sınırlı sayıda tematik öznitelik dönüşümü desteklenmektedir. Üretilen RDF dosyası GeoSPARQL standardı ile uyumludur.

2.5. GeoTriples

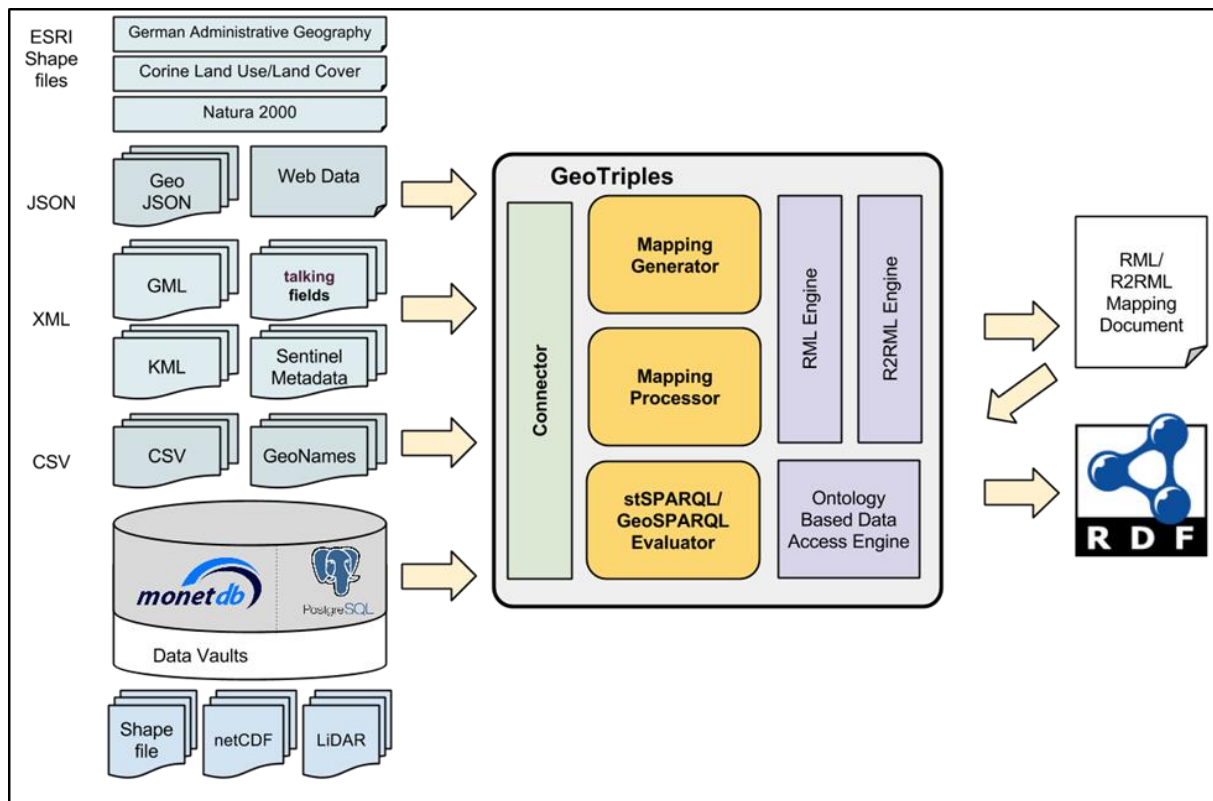
GeoTriples²⁹ konumsal verileri Bağlantılı Açık Konumsal Veri olarak yayınlamak için geliştirilmiştir. GeoTriples, belirli bir kelime hazinesine bağlı olmayan, GeoSPARQL ve stSPARQL gibi kelime hazinelerini kullanarak konumsal bilginin RDF olarak yayınlanmasını sağlayan yarı otomatik bir yazılım aracıdır. GeoTriples girdi veri kaynağının şemasına göre otomatik olarak R2RML/RML eşleştirme dosyasını üreten *Mapping Generator*, bir R2RML/RML eşleştirmesini işleyen ve bir RDF çizgesi üreten *R2RML/RML processor* ve ilişkisel veri tabanı üzerinde stSPARQL/GeoSPARQL sorgularını değerlendiren *Ontology Based Data Access (OBDA) modülü* olmak

²⁹ <https://github.com/LinkedEOData/GeoTriples>

üzere üç modül içerir.

İlk işlem adımı ilişkisel veri tabanlarından Bağlantılı Veri nin elde edilmesini içerir. İlişkisel veri tabanlarının Bağlantılı Veri ye dönüşümü için RDB to RDF eşleştirme dili (RDB to RDF Mapping Language-R2RML) W3C Standardıdır. Dönüşüm sırasında, Subject birincil anahtara sahip tablodaki sütunun adıdır, Bir tablodaki her bir sütunun adı RDF predicate dir. Bir tablonun her bir hücre değeri bir object tir. R2RML eşleştirme işleminin tamamlanmasından sonra GeoSPARQL kelime hazinesine uygun olarak RDF dosyası oluşturulur.

Şekil 3 te GeoTriples mimarisinde görüldüğü gibi girdi olarak ESRI Shapefile gibi konumsal veri formatları alınır ve RDF formatına dönüştürülür. GeoTriples, *Mapping Generator* ve *R2RML processor* olmak üzere iki ana bileşenden oluşur. *Mapping Generator*, veri kaynağını alır ve onu bir RDF çizgesine dönüştüren R2RML eşleştirmesini otomatik olarak oluşturur. Oluşturulan eşleştirme, konumsal verinin özelliklerini dikkate almak ve GeoSPARQL Kelime Hazinesi ile uyumlu bir RDF çizgesi üretmek için ihtiyaç duyulan bütün dönüşümleri sağlamak için subject-object-predicate eşleştirmeleri ile zenginleştirilir. Yazılım, R2RML eşleştirme dosyasının farklı kelime hazineleri kullanılacak şekilde kullanıcı gereksinimlerine izin vermektedir. Sonrasında üretilen RDF dosyası Strabon gibi Triple Store lar kullanılarak depolanabilir.



Şekil 3. GeoTriples Mimarisi (Kyzikaros vd., 2014).

GeoTriples yazılımı ile Karayolu shapefile dosyası RDF formatına dönüştürülmüştür. Dönüşüm sonucunda oluşturulan RDF dosyası incelendiğinde;

Girdi dosya formatı olarak Shapefile (ESRI shapefile), KML, PostgreSQL, MonetDB gibi konumsal VTYS desteklenir.

R2RML dosyasını otomatik oluşturur kullanıcı müdahalesine imkan verir.

GeoSPARQL ve stSPARQL standardı desteklenir.

Veri tabanı tabloları mevcut bir ontoloji ile ilişkilendirilebilir.

Kullanıcıların veri kaynağına bağlanıp ontoloji yükleyebilecekleri ve dönüşümü gerçekleştirebilecekleri bir GUI desteği mevcuttur.

R2RML eşleştirme dosyası ve RDF çizgesi otomatik olarak oluşturulur.

Çıktı dosya formatları RDF/XML, RDF/XML-ABBREV, N3, N-Triple, Turtle destekler.

Koordinat dönüşümüne imkan verir.

2.5.Mevcut Yazılım Araçlarının Karşılaştırılması

Tim Berners-Lee tarafından belirtildiği gibi Semantik Web belirli kurallar ve standartlar aracılığıyla farklı

kullanıcılar ve platformlar arasında verinin paylaşılmasını ve yeniden kullanımına izin verir. Semantik kavramının anlamı aslında belirli bir nesne veya varlığın anlamı için ortak bir tanım anlamına gelir. Semantik Web vizyonunda bilgisayarlar verinin gerçek anlamını yada kastedilen anlamını anlama yeteneğine sahiptir. Buna paralel olarak konumsal verilerin bilgisayarlar tarafından işlenerek analiz edilmesi ve mevcut verilerden çıkarsama işleminin gerçekleştirilmesi ve sorgu sonuçlarının gösterilmesi için Semantik Web teknolojileri kullanılarak gerekli tanımların yapılması gerekir. Konumsal veri kaynaklarındaki semantik tanımların oluşturulmasında Bağlantılı Veri Yaklaşımı, veri kaynaklarının ortak bir modele bağlı kalmaksızın avantajlı bir yaklaşımdır. Konumsal verilerin Bağlantılı Açık Veri olarak yayınlanması sürecinde aşağıdaki adımların gerçekleştirilmesi gerekir.

Konumsal verilerin RDF ye dönüşümü

RDF verisinin tanımlanacağı ontolojinin belirlenmesi veya yeni baştan oluşturulması

RDF verilerinin ontoloji ile ilişkilendirilmesi

RDF verisinin Bağlantılı Veri İlkelerine uygun olarak yayınlanması (endpoint)

RDF verisinin diğer veri setleri ile veya Bağlantılı Veri Bulutunda ilgili veri setleri ile ilişkilendirilmesi (owl:sameAs)

Konumsal sorgulama ve semantik çıkarsama

Sorgu sonuçlarının görselleştirilmesi (web harita)

Konumsal verilerin RDF ye dönüştürülmesi için geliştirilen yazılım araçlarının bazı sınırlı özellikleri vardır. GeoSPARQL kelime hazinesi ile uyumlu olan yazılım araçlarıyla çalışmak herhangi bir koordinat sisteminde geometrilerin tanımlanmasına izin vermesi nedeniyle tercih edilmelidir. NeoGeo kelime hazinesi ile tutarlı yazılım araçları ile elde edilen geometri koordinatları WGS84 koordinatlarıyla sınırlı kalmasına rağmen bu Kelime hazinesine uygun RDF verileri SPARQL ile sorgulanabilmektedir.

Çizelge 1. Yazılım araçlarının karşılaştırılması.

Yazılım Araçları	Tematik Veri Desteği	Doğrudan Eşleştirme	R2RML	RML	Otomatik Eşleştirme	GeoSPARQL Uyumu	RDBMS	ESRIShapefile dosyası
Geometry2RDF	-	X	-	-	X	-	X	X
shp2GeoSPARQL	-	-	-	-	X	X	-	X
GeomRDF	X	X	-	-	X	X	X	X
TripleGeo	(X)	X	-	-	-	X	X	X
GeoTriples	X	-	X	X	X	X	X	X

Çizelge 1 deki karşılaştırma sonuçlarına göre yalnızca konumsal verilerin RDF ye dönüşümünü içeren uygulamalarda geliştirilen yazılım araçlarının hemen hepsi kullanılabilir. Uygulama kapsamında gerçekleştirilen dönüşüm sonuçlarına göre hem konumsal hem tematik verilerin RDF ye dönüşüm ihtiyacı olan uygulamalar için GeomRDF ve GeoTriples yazılımlarının kullanılması uygun görülmektedir.

3.SONUÇLAR

Son yıllarda web üzerinde Bağlantılı Veri ve konumsal bağlantılı veriler üzerine gerçekleştirilen proje, girişim ve akademik çalışmaların sayısında büyük bir artış vardır. Aynı zamanda semantik arama, sorgulama ve arama konusunda geliştirilen standartlar da önemli gelişmelere adım atılmıştır. Konumsal verilerin Bağlantılı Veri olarak yayınlanmasındaki temel amaç verinin yeniden kullanılabilirliğini sağlamak ve veri kaynakları arasında birlikte işlerliği sağlamaktır.

Konumsal Bağlantılı Veri son yıllarda giderek yaygınlaşmakta ve ulusal konumsal veri sağlayıcıları tarafından yayınlanan Konumsal Bağlantılı veriler her geçen gün daha da artmaktadır. Konumsal Semantik Web uygulamalarına hız kazandıracak olan “Konumsal Bağlantılı Veri Yaklaşımı” henüz gelişim aşamasındadır ve bu yüzden veri dönüşümünü gerçekleştirmek için yazılım bulunamaması veya yeterli özellikleri desteklememesi en önemli eksikliklerin başında gelir. Diğer bir eksiklik ise mevcut yazılım araçlarının geometrik veri tiplerinin temsilinde kendine özgü kelime hazineleri kullanmalarıdır. Bu eksiklik son yıllarda geliştirilen yazılım araçları ile giderilse de Bağlantılı Konumsal Veri nin üretilmesinden yayınlanmasına kadar gerçekleştirilen tüm işlemlerin tek bir platformda gerçekleştirilmesi tüm bu eksiklikleri geride bırakacaktır. Dönüşümün gerçekleştirilmesi sırasında ise yazılımlar kullanıcı deneyimi gerektirir. Bu ise yazılım araçlarının kullanımını ve anlaşılmasını zorlaştırır. Bu nedenle yazılım araçlarına uygulama programlama ara yüzü desteği getirilmelidir. Ayrıca Konumsal verileri Bağlantılı Veri olarak yayınlamak için LOD Bulutu yeterli sayıda alanı içermemektedir ve konumsal verilerin referanslandırılması için gerekli alanları tanımlayan veri setleri henüz tamamlanmamıştır. Konumsal Bağlantılı verilerin görselleştirilmesi ve konumsal analiz desteği diğer bir sorundur.

Konumsal verilerin Bağlantılı Veri olarak yayınlanması güncel çalışma alanlarından birisidir. Ayrıca konumsal alanda gerçekleştirilecek Semantik Web uygulamaları için en temel gereksinimdir. Bu eksikliği gündeme getirmek ve ulusal düzeyde Bağlantılı Verilerin yayınlanması için gereksinimlerin belirlenmesi adına bu çalışma

gerçekleştirilmiştir. İlgili literatürde konumsal veri kaynaklarındaki verilerin RDF ye dönüştürülmesi için uygulama senaryosu dikkate alınarak böyle bir dönüşüm ve karşılaştırma yapılmamıştır. Dönüşüm ve Bağlantılı Verilerin yayınlanmasında gereksinimlerin belirlenmesi ve kurumsal düzeyde eksikliklerin giderilmesi dikkate alındığında çalışmanın önemi ortaya çıkacaktır. Gerçekleştirilen çalışma ile birlikte ulusal ve kurumsal düzeyde gerçekleştirilecek olan semantik web uygulamalarının temel gereksinimlerinin belirlenmesine ışık tutmaktadır. Bağlantılı Veri yayınlamak için kurumsal düzeyde temel olarak kullanılacak bir referanstır.

KAYNAKLAR

Auer, S., Dietzold, S., Lehmann, J., Hellmann, S. and Aumueller, D., 2009, April. Triplify: light-weight linked data publication from relational databases. In *Proceedings of the 18th international conference on World wide web* (pp. 621-630). ACM.

Berners-Lee, T., 2006, Design issues: Linked data, <https://www.w3.org/DesignIssues/LinkedData.html>. (15.08.2016).

Bizer, C., Heath, T. and Berners-Lee, T., 2009, Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pp.205-227.

Hamdi, F., Abadie, N., Bucher, B. and Feliachi, A., 2015, Geomrdf: A geodata converter with a fine-grained structured representation of geometry in the web. *arXiv preprint arXiv:1503.04864*.

Heath, T. and Bizer, C., 2011, Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), pp.1-136.

Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S. and Koubarakis, M., 2014, October. GeoTriples: a tool for publishing geospatial data as RDF graphs using R2RML mappings. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272* (pp. 393-396). CEUR-WS. org.

OGC, 2012, OGC GeoSPARQL - A Geographic Query Language for RDF Data, OGC[®] Implementation Standard, Editors: Matthew Perry and John Herring, Open Geospatial Consortium. <http://www.opengis.net/doc/IS/geosparql/1.0>, (12.07.2016).

Patroumpas, K., Alexakis, M., Giannopoulos, G. and Athanasiou, S., 2014. TripleGeo: an ETL Tool for Transforming Geospatial Data into RDF Triples. In *EDBT/ICDT Workshops* (pp. 275-278).

W3C, 2012, <https://www.w3.org/TR/r2rml/>, R2RML: RDB to RDF Mapping Language, Editors: Souripriya Das-Oracle, Seema Sundara-Oracle, Richard Cyganiak-DERI, National University of Ireland, Galway, W3C Recommendation 27 September 2012.