



## Meteorolojik Verilerin Bağlantılı Veriye Dönüştürülmesi: Karadeniz Bölgesi Örneği

### Transforming Meteorological Data into Linked Data: The Case of the Black Sea Region

Deniztan Ulutaş Karakol<sup>1\*</sup>, Çetin Cömert<sup>1</sup>

<sup>1</sup>Karadeniz Teknik Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 61080, Trabzon/Türkiye.

#### ARAŞTIRMA MAKALESİ

##### \*Sorumlu yazar:

Deniztan Ulutaş Karakol  
deniztanulutas@ktu.edu.tr

doi: 10.48123/rsgis.869992

##### Yayın süreci

Geliş tarihi: 29.01.2021

Kabul tarihi: 08.03.2021

Basım tarihi: 13.03.2021

#### Özet

Semantik Web'in gereği olan Bağlantılı Açık Veri üretimi, bağlantılı veri teknolojilerinin sağladığı avantajlar ile konumsal ve konumsal olmayan veri ile iş yapan bütün kamu ve özel kuruluşlar için önemlidir. Veriye daha kolay erişim, veri paylaşımı, veri kullanılabilirliğinin artması ve tekrarının önlenmesi bu avantajlar arasındadır. Coğrafi Bilgi Sistemleri, risk yönetimi, hava tahminleri ve hava kirliliğinin modellenmesi, su kaynakları, tarım, hava trafiği ve endüstri gibi birçok alanda önemli derecede kullanılan meteorolojik verilerin bağlantılı açık veriye dönüştürülmesi gerekli ve önemlidir. Bu çalışmada meteorolojik verilerin bağlantılı veriye dönüştürülmesi Karadeniz Bölgesi örneği ile gerçekleştirilmiştir. Bölgeye ait meteorolojik istasyon ve gözlem verileri ilgili hava ve meteoroloji kavramlarını içeren ontolojilerle referanslandırılarak RDF verisi üretilmiştir. Üretilen meteorolojik RDF verisi ile bağlantılı açık veri kaynağı olan DBpedia veri seti arasında bağlantılar kurularak bağlantılı meteorolojik veriler elde edilmiştir. Silk bağlantı editörü ile otomatik olarak bulunan bağlantı sonuçları, meteorolojik verilerle DBpedia veri kaynağı arasında hesaplanan benzerlik değerlerinin yüksek olduğunu göstermektedir.

**Anahtar kelimeler:** Bağlantılı veri, Meteorolojik bağlantılı veri, RDF, Semantik Web, Ontoloji

#### Abstract

Linked Open Data creation, which is a requirement of the semantic Web, is significant for all public and private parties related with spatial and non-spatial data with the advantages of linked data technologies. Among these advantages are data being more discoverable, data sharing, increased reusability of data and preventing data duplication. Meteorological data is used significantly in many areas such as Geographical Information Systems, risk management, weather forecasts, modeling of air pollution, water resources, agriculture, air traffic and industry. So it is necessary and significant to transform meteorological data into linked open data. In this study, the transformation of meteorological data into linked data was carried out with the case of the Black Sea region. RDF data was generated by referencing the meteorological station and observation data of the region with ontologies including the relevant weather and meteorology concepts. Linked meteorological data were obtained by establishing links between the generated meteorological RDF data and DBpedia data set, which is an open linked data source. The automatically found link results with the Silk link editor show that the similarity values calculated between the meteorological data and the DBpedia data source are high.

**Keywords:** Linked data, Meteorological linked data, RDF, Semantic Web, Ontology

## 1. Giriş

Gelişen teknoloji ile birlikte Web’de anahtar kelime eşleştirme yöntemleri ile veri bulan geleneksel Web arama motorları, yerini semantik veri araması sağlayan Bağlantılı Veri (Linked Data) yaklaşımına bırakacaktır. Bağlantılı Veri, bilgisayarlar tarafından okunup anlaşılacak şekilde semantik olarak tanımlanmış ve diğer Web kaynakları ile ilişkilendirilmiş veri anlamına gelmektedir. Bağlantılı veri, yapısal verinin Web ortamında yayımlanması ve aralarında ilişki kurulabilmesi için farklı Web kaynakları ile bağlanması ile ilgili yöntem ve kuralları belirtir.

Bağlantılı Veri, farklı veri kaynakları arasında birlikte işlerliği, yeniden kullanılabilirliği ve semantik sorgular yapabilmeyi sağlar. Bağlantılı Veri teknolojileri verilerin semantik işlenmesi, görselleştirilmesi, Web’de yayınlanması ve paylaşılmasını sağlar. Semantik olarak verileri tanımlayabilmek için, bir alan için ortak kavramsallaştırmalar olarak tanımlanan ontolojilere ya da harici Web kaynaklarına, bağlantılar kurulur. Kaynak veri sözcükleri ile global olarak yayınlanmış kaynakların kavramları arasında bağlantı kurmak, farklı veri kaynaklarını ve veri kümelerini birleştirir, verileri daha anlaşılır, keşfedilebilir ve kullanılabilir hale getirir, verilerin birlikte işlerliğini ve entegrasyonunu geliştirir, otomatik anlam çıkarsaması sağlar ve veri tekrarını önler (Ulutaş Karakol vd., 2018).

Bağlantılı Veri ilkelerinin yaygınlaştırılması ve benimsenmesi için Bağlantılı Açık Veri Topluluğu Projesinin ürünü olan Bağlantılı Açık Veri Bulutu<sup>1</sup> (LOD Cloud) oluşturulmuştur (Berners-Lee, 2021). Bağlantılı Açık Veri Bulutu, bağlantılı veri formatında yayınlanmış veri setlerini içeren Alanlar arası, Coğrafya, Ülke, Yaşam Bilimleri, Dilbilim, Medya, Yayınlar, Sosyal Ağ ve Kullanıcı olarak dokuz alt buluttan oluşmaktadır. Mayıs 2020 itibarıyla 16201 bağlantı ile 1269 veri seti içermektedir (The LOD cloud, 2021). Projenin ana amacı Bağlantılı veri prensiplerine göre mevcut veri kaynaklarından bağlantılı veri oluşturmak ve Web’de yayınlamaktır (GeoNames, DBpedia, FOAF, MusicBrainz, vb.) (Ulutaş Karakol vd., 2018).

Semantik Web sadece veriyi Web’de paylaşmak değil, bağlantı kurmaktır. Böylece bir kişi veya bilgisayar/makine Web’deki verileri keşfedebilir. Bağlantılı Veri sayesinde, verinin bir kısmına erişim sağlandığında diğer ilgili veriler de bulunabilir (Berners-Lee, 2021).

Bağlantılı Veri prensipleri Berners-Lee (2021) tarafından dört ilke şeklinde sıralanmaktadır:

- Varlıkların isimleri olarak URI’leri kullanın.
- İnsanların o isimleri arayabilmeleri için HTTP URI’leri kullanın.
- Bir URI arandığında, standartları (RDF, SPARQL) kullanarak yararlı bilgi sağlayın.
- Başka URI’lere bağlantılar ekleyin ki insanlar daha fazla bilgi keşfedebilsinler.

Literatürde Bağlantılı Veri oluşturma konusunda akademik çalışmalar ve projeler (Ulutaş Karakol vd., 2018; Kara vd., 2018; Kara vd., 2020; Bischof vd., 2018; Margan vd., 2018; Qiu vd., 2017; Iwaniak vd., 2017; Szekely vd., 2013; Debruyne vd., 2016; Hietanen vd., 2016; Consolia vd., 2017; Ateazing vd., 2011) mevcuttur.

Ulutaş Karakol vd. (2018) konumsal RDF verilerinin mevcut durumunu, standartları, açık kaynak kodlu platformlar ve Web veri kaynaklarını incelemiştir. Yol verisi ile DBpedia Web veri kaynağı arasında bağlantı kurmuştur. Kara vd. (2018) konumsal verilerin bağlantılı veri olarak yayınlanması ve harita üzerinde gösterilmesi için mevcut durum değerlendirmesi yaparak bir yöntem önermiştir. Önerilen yöntemde göre konumsal verileri bağlantılı veri olarak Web’de yayınlamış ve harita üzerinde göstermiştir. Kara vd. (2020) “Trabzon İdari Sınırlar” veri setini ontolojilerle ilişkilendirerek Web üzerinde Bağlantılı Veri olarak yayınlamış ve konumsal sorgular gerçekleştirmiştir. Bischof vd. (2018) şehir düzeyindeki istatistiksel verilerin toplanması, birleştirilmesi, zenginleştirilmesi ve elde edilen veri kümesinin yeniden kullanılabilir şekilde Bağlantılı Veri olarak yayınlanmasını sağlayan Açık Şehir Veri Hattı oluşturmuştur. Hem istatistiksel regresyon yöntemlerine hem de arka plan bilgi tabanı üzerinde ontolojik çıkarsamalara dayanan sarmalayıcılar (Wrappers), veri tarayıcısı, ontoloji tabanlı entegrasyon platformu, eksik değer tahmin modülü gibi bileşenler geliştirmiştir. Margan vd. (2018) şehirlerde duyarlı ve korunmasız alanlar için hava kalitesi endeksinin hesaplanması amacıyla hava kirliliği sensör verileri için Konumsal Bağlantılı Veri (Geo Linked Data) oluşturmuş ve konumsal istatistiksel analiz gerçekleştirmiştir. Çalışmada hava kirliliği bilgilerini zenginleştirmek ve zararlı parçacık kirliliği seviyelerine sahip alanları belirlemek için DBpedia veri kaynağı kullanılmıştır. Qiu vd. (2017) çevresel modelleri ve afetle ilgili verileri birbirine bağlayan bir afet azaltma sürecinde iş akışı oluşturulmasında sel afeti yönetim sistemi otonomisini ve veri alımını sağlayan ontoloji tabanlı bir yaklaşım önermiştir. Modellerin ve verilerin birbirine bağlanması, sistemin doğru zamanda heterojen veri kaynaklarından gelen doğru modeller için doğru veri kümelerini bulmasını sağlamıştır. Iwaniak vd. (2017) RDF veri modelini Bağlantılı Veri olarak kullanarak kurumlardan veya veri tabanlarından veri ve meta verileri yayınlamak için bir sistem tasarlamış ve uygulamıştır. Szekely vd. (2013) Smithsonian Amerikan Sanat Müzesi verilerini Bağlantılı Açık Veri ile eşleştiren bir çalışma yapmıştır. Çalışmada herhangi bir müzenin veya kurumun kendi verisini kendi kendine Bağlantılı Veriyle eşleştirmesini sağlayan yazılımlar geliştirilmiştir. Aynı zamanda başka kaynaklara otomatik olarak oluşturulmuş linkleri müze personelinin kolayca görüntüleyip hataları düzeltebileceği bir doğrulama yazılımı geliştirilmiştir.

<sup>1</sup> <https://lod-cloud.net>

Debruyne vd. (2016) İrlanda'nın idari sınırlarını farklı genelleştirme düzeylerinde oluşturarak Bağlantılı Veri olarak yayınlamıştır. Hietanen vd. (2016) WFS'in (Web Feature Service) sunduğu coğrafi isimler veri setini Bağlantılı Veriye çeviren bir servis geliştirmiştir. Çalışmada WFS'in çıktısı anlık olarak RDF formatına dönüştürülmektedir. Consolia vd. (2017) Catania belediyesi için Bağlantılı Açık Verinin toplanması, zenginleştirilmesi ve yayınlanmasını sağlayan bir sistem geliştirmiş ve bu sistemi semantik birlikte işlerliği sağlayan yöntemler, ontoloji tasarım modeli ve araçlar/yazılımlar ile birlikte tanıtmıştır. Ateazing vd. (2011) çalışmasında AEMET (Agencia Estatal de Meteorología) (İspanyol Meteoroloji Ofisi) tarafından yayınlanan meteorolojik verileri Bağlantılı Veriye dönüştürmüştür. Ayrıca otomatik veri bulma ve güncelleme yapmış ve verilerin konumlarını dikkate alarak ürettiği Bağlantılı Veriyi görselleştirmiştir.

Ülkemiz literatüründe meteorolojik verileri Bağlantılı Veriye dönüştürme konusunda bir çalışma yoktur. Bu çalışmada meteorolojik verilerin bilgisayarların/yazılımların anlayacağı biçimde anlamlandırılması, herkesin erişebileceği şekilde paylaşımına açılıp farklı veri setleri ile entegre edilerek zenginleştirilmesi amaçlanmıştır. Bu amaçla bu çalışmada Karadeniz Bölgesine ait meteorolojik verilerin semantik olarak tanımlanması ve Bağlantılı Veriye dönüştürülmesi gerçekleştirilmiştir.

## 2. Materyal ve Metot

Çalışmada meteorolojik verilerin Bağlantılı Veri olarak yayınlanması için dört işlem adımı gerçekleştirilmiştir (Şekil 1). Öncelikle çalışma alanına ait meteorolojik veriler temin edilmiş ve bu veriler kullanılacak yazılımın veri formatına dönüştürülmüştür. İkinci aşamada meteorolojik verilerin semantik tanımlarının oluşturulması için veri setinde bulunan kavramlarla aynı ya da benzer kavramları içeren ontolojiler belirlenmiştir. Üçüncü aşamada ontolojilerle referanslandırılan verilerden RDF verisi üretilmiştir. Son aşamada oluşturulan RDF verisi ile farklı veri setleri arasında bağlantılar kurulmuştur. Sıradaki bölümlerde bu işlem adımları detaylı bir şekilde açıklanacaktır.



Şekil 1. Meteorolojik verilerin bağlantılı veri olarak yayınlanması için iş akışı

### 2.1 Çalışma Alanının Belirlenmesi ve Kullanılan Veriler

Çalışma alanı Karadeniz bölgesi sınırları içindeki illeri kapsamaktadır. Çalışmada kullanılan meteorolojik veriler Meteoroloji Genel Müdürlüğü'nün izni ile kullanılmıştır. Meteorolojik verilerin içeriği Karadeniz Bölgesindeki illerin istasyon numarası, istasyon adı, bölgesi, ilçesi, enlem-boylam koordinatları, rakımı, günlük buhar basıncı, günlük güneşlenme süresi, günlük maksimum aktüel basınç, günlük maksimum buhar basıncı, günlük maksimum deniz suyu sıcaklığı, günlük maksimum nispi nem, günlük maksimum rüzgâr yönü ve hızı, günlük maksimum sıcaklık, günlük minimum aktüel basınç, günlük minimum buhar basıncı, günlük minimum deniz suyu sıcaklığı, günlük minimum nispi nem, günlük minimum sıcaklık, günlük ortalama aktüel basınç, günlük ortalama deniz suyu sıcaklığı, günlük ortalama nispi nem, günlük ortalama sıcaklık, günlük ortalama rüzgâr yönü ve hızı, günlük ortalama sıcaklık ve günlük toplam yağış miktarıdır.

## 2.2. Meteorolojik Verilerle İlgili Ontolojilerin Belirlenmesi

Veri setinde yer alan kavramların ve özniteliklerin semantik olarak tanımlanabilmesi için ontolojilere ihtiyaç vardır. Bu çalışmada yeni bir ontoloji oluşturmak yerine ontolojilerin tekrar kullanılabilirliğini sağlamak için Web’de mevcut ontolojiler araştırılmış ve meteorolojik verileri temsil eden anahtar kelimeler ve veri tanımlarını içeren Sweet (BioPortal, 2021), Geonames (GeoNames Ontology, 2021), DBpedia (DBpedia, 2021), WGS84 Geo Positioning (Geo, 2021) ve W3C zaman ontolojisi (Time, 2021) uygun bulunarak kullanılmıştır.

### 2.2.1 Sweet Ontolojisi

Sweet (The Semantic Web for Earth and Environmental Terminology) 200 ontolojiden oluşturulmuş 6000’den fazla kavram içeren olgun, temel bir ontolojidir. Sweet ontolojisinin içerdiği üst düzey kavramlar arasında Temsil (matematik, uzay, bilim, zaman, veri), Bölge (Okyanus, Kara Yüzeyi, Karasal Hidrosfer, Atmosfer vb.), Fenomenler (makro ölçekli ekolojik ve fiziksel), Süreçler (mikro ölçekli fiziksel, biyolojik, kimyasal ve matematiksel), İnsan Faaliyetleri (Karar, Ticaret, Yargı, Çevre, Araştırma) yer alır (BioPortal, 2021). Çalışma kapsamında Sweet Ontolojisi ‘Meteoroloji istasyonu’, ‘hava tahmini’ ve ‘hava’ kavramlarının referanslandırılmasında kullanılmıştır.

### 2.2.2 DBpedia Ontolojisi

DBpedia ontolojisi, DBpedia veri setinde kullanılan sınıfları ve özellikleri tanımlar. DBpedia Mappings Wiki’de (DBpedia Mappings Wiki, 2021) manuel olarak oluşturulan spesifikasyonlardan üretilmiştir (DBpedia, 2021). Çalışma kapsamında ‘istasyon numarası’ (station number), ‘bölge’ (province), ilçe (district) ve ‘maksimum sıcaklık’ (maximum temperature) kavramlarının referanslandırılması için kullanılmıştır.

### 2.2.3 WGS84 Geo Positioning Ontolojisi

WGS84 Geo Positioning Ontolojisi, WGS84 jeodezik referans datumunda enlem, boylam ve yükseklik bilgilerini temsil eden bir RDF ontolojisidir (Geo, 2021). Çalışma kapsamında meteorolojik veri setindeki ‘enlem’, ‘boylam’ ve ‘yükseklik’ kavramlarını referanslandırmak için kullanılmıştır.

### 2.2.4 GeoNames Ontolojisi

GeoNames Ontolojisi, GeoNames (GeoNames, 2021) veri tabanında tanımlı coğrafi detaylar için açıklamalar içerir. İçeriğindeki ana kavramlar coğrafi kod, GeoNames detayı, coğrafi detay, harita, RDF verisi ve Wikipedia makalesi’dir. (GeoNames Ontology, 2021). Çalışma kapsamında ‘İstasyon ismi’ (station name) kavramının referanslandırılmasında kullanılmıştır.

### 2.2.5 W3C Zaman Ontolojisi

W3C Zaman Ontolojisi, Dünyadaki kaynakların zamansal (tarih, saat) özelliklerini tanımlamak için oluşturulmuş bir ontolojidir. İçerdiği ana kavramlar tarih-zaman tanımı, tarih-zaman aralığı, haftanın günü, süre, süre tanımı, genelleştirilmiş tarih-saat tanımı, genelleştirilmiş süre tanımı, anlık zaman, zaman aralığı, yılın ayı, uygun aralık, zamansal süre, zamansal varlık, zamansal konum, zamansal birim, zaman konumu, saat dilimi ve Zamansal Referans Sistemi’dir (Time Ontology in OWL, 2021). Çalışma kapsamında yıl (year), ay (month), gün (day) kavramlarının referanslandırılmasında kullanılmıştır.

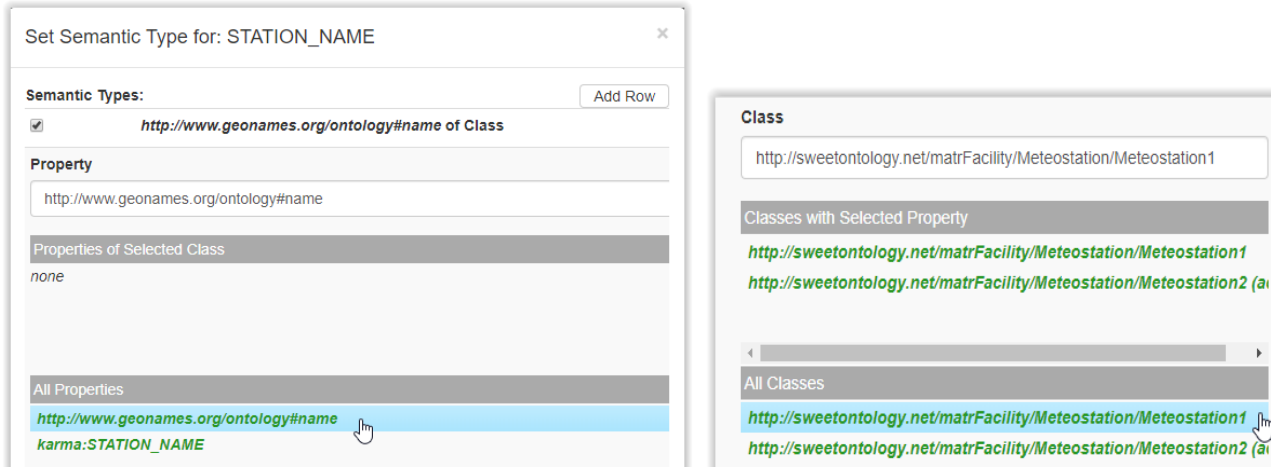
## 2.3 Verilerin RDF’ye Dönüştürülmesi

Literatürde konumsal veri tabanlarının RDF’ye dönüştürülmesi için GeomRDF (Hamdi vd., 2015), TripleGeo (Patroumpas vd., 2014), GeoTriples (Kyzirakos vd., 2014), shp2GeoSPARQL (Saavedra vd., 2014), OpenRefine (OpenRefine, 2021) ve Karma (Karma, 2021) gibi çeşitli yazılım araçları kullanılmaktadır. Bu çalışmada ilgili meteorolojik verileri RDF formatına dönüştürebilmek için Karma yazılımı kullanılmıştır.

Çalışmada Karma’nın tercih edilmesinin nedeni birçok farklı formattaki veriyi RDF’ye dönüştürebilmesi ve RDF verisini oluştururken kullanıcıya farklı ontolojileri kullanabilme olanağı sunmasıdır. Karma, kullanıcıların verilerini standart sözlüklerle/veri kaynaklarıyla eşleştirebilmeleri için farklı ontolojilerin birleştirilmesini sağlar. Kullanıcıların veri tabanları, elektronik tablolar, sınırlandırılmış metin dosyaları, XML, JSON, KML ve Web API’leri dahil olmak üzere çeşitli veri kaynaklarından verileri hızlı ve kolay bir şekilde entegre etmelerini sağlar.

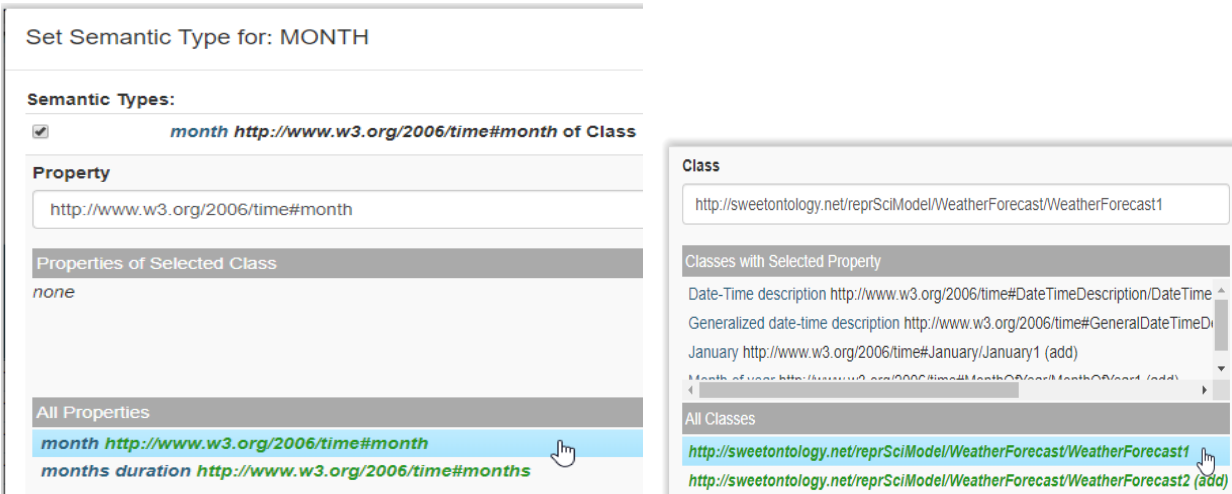
Karma, verilerin ontoloji sınıflarına eşlenmesini öğrenerek bu sınıfları birbirine bağlayan bir model önermek için ontolojiyi kullanır. Bilgileri entegre edebilmek için ontolojileri, alan ve kısıtlamaları kullanır (Karma, 2021).

Çalışmada mevcut meteorolojik istasyon verileri ile günlük maksimum sıcaklık verileri TXT formatından XLSX formatına dönüştürülerek KARMA'nın içine aktarılmıştır. Daha sonra içe aktarılan verilerin modellemesi yapılmıştır. Karma'da veri modelleme, verinin diğer kaynaklardaki verilerle entegre edilebilmesi ve farklı formatlarda yayınlanabilmesi için ontolojiler ile referanslandırılmasıdır. Veri modellemenin ilk aşaması semantik tiplerin belirlenmesidir. Semantik tip, veri setindeki tabloların sütunlarında yer alan öznitelikler ile ontoloji sınıfları ve özellikleri arasındaki ilişkiyi tanımlar. Bu aşamada istasyon veri setindeki İstasyon numarası (station number), İstasyon ismi (station name), Bölge (Province), İlçe (District), Enlem (Latitude), Boylam (Longitude), Yükseklik (Altitude) özniteliklerinin semantik tipleri şekil 2, 3, 4, 5 ve 6 'da gösterildiği gibi belirlenmiştir.



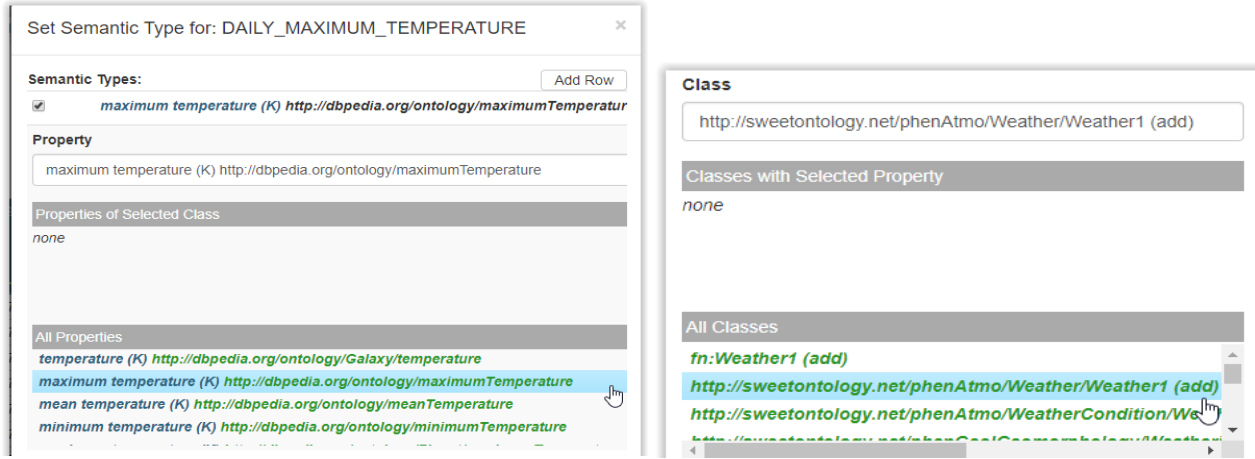
Şekil 2. İstasyon ismi özniteliğinin semantik tipinin belirlenmesi

Şekil 2'de görüldüğü üzere 'İstasyon İsmi' özniteliğinin semantik tipi, Geonames ontolojisindeki "name" özelliği ve Sweet ontolojisindeki 'Meteostation' sınıfı ile referanslandırılması ile belirlenmiştir.



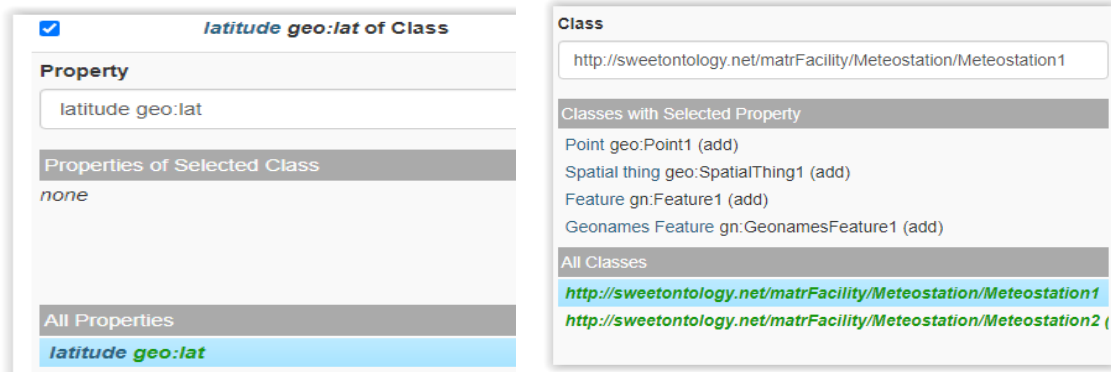
Şekil 3. Ay özniteliğinin semantik tipinin belirlenmesi

Şekil 3'de görüldüğü üzere 'Ay' özniteliğinin semantik tipi W3C Time ontolojisindeki "month" özelliği ve Sweet ontolojisindeki 'Weather Forecast' sınıfı ile referanslandırılması ile belirlenmiştir.



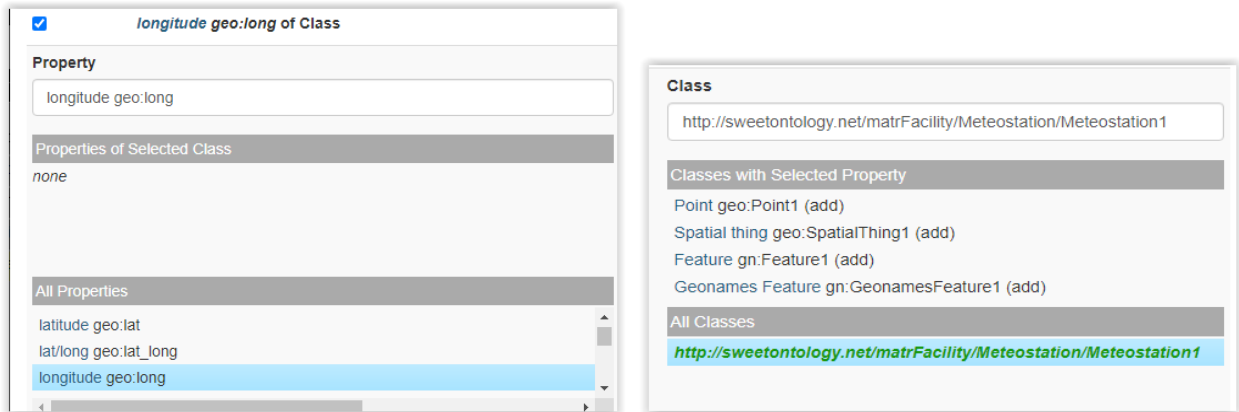
Şekil 4. Günlük Maksimum Sıcaklık özniteliğinin semantik tipinin belirlenmesi

Şekil 4’de görüldüğü üzere ‘Günlük Maksimum Sıcaklık’ özniteliğinin semantik tipi, DBpedia ontolojisindeki “Maximum Temperature” özelliği ve Sweet ontolojisindeki ‘Weather’ sınıfı ile referanslandırılması ile belirlenmiştir.



Şekil 5. Enlem özniteliğinin semantik tipinin belirlenmesi

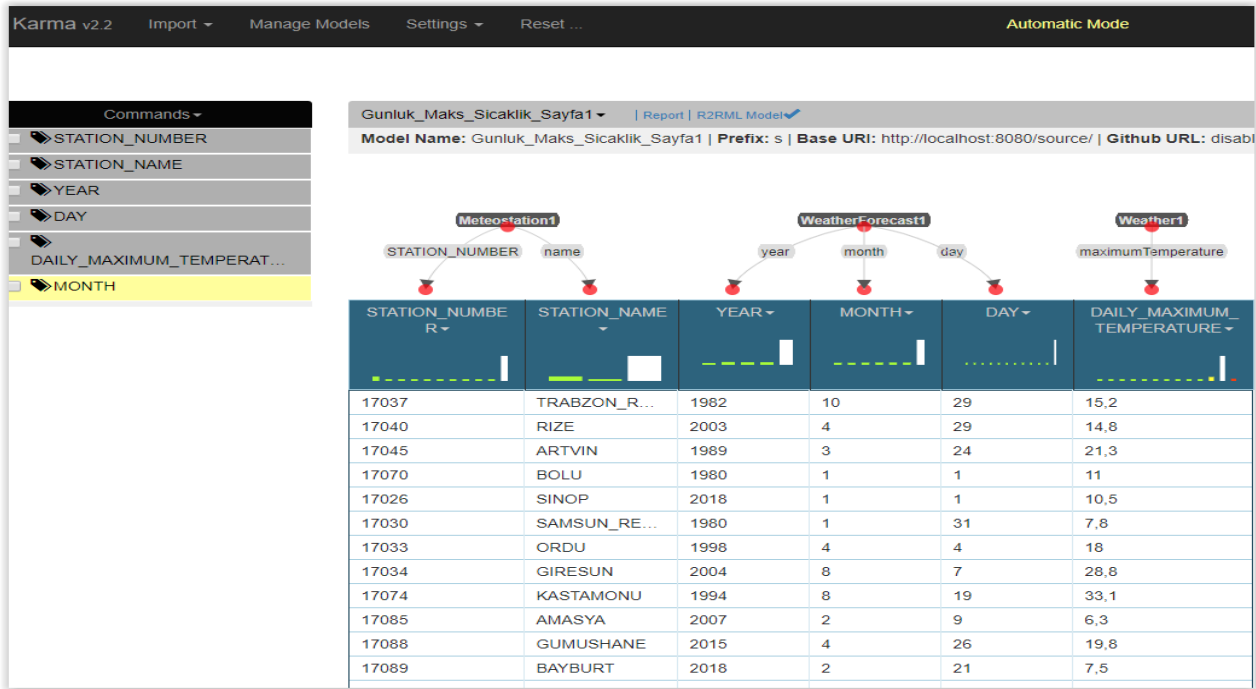
Şekil 5’de görüldüğü üzere ‘Enlem’ özniteliğinin semantik tipi, WGS84 Geo Positioning ontolojisindeki “latitude” özelliği ve Sweet ontolojisindeki ‘Meteostation’ sınıfı ile referanslandırılması ile belirlenmiştir.



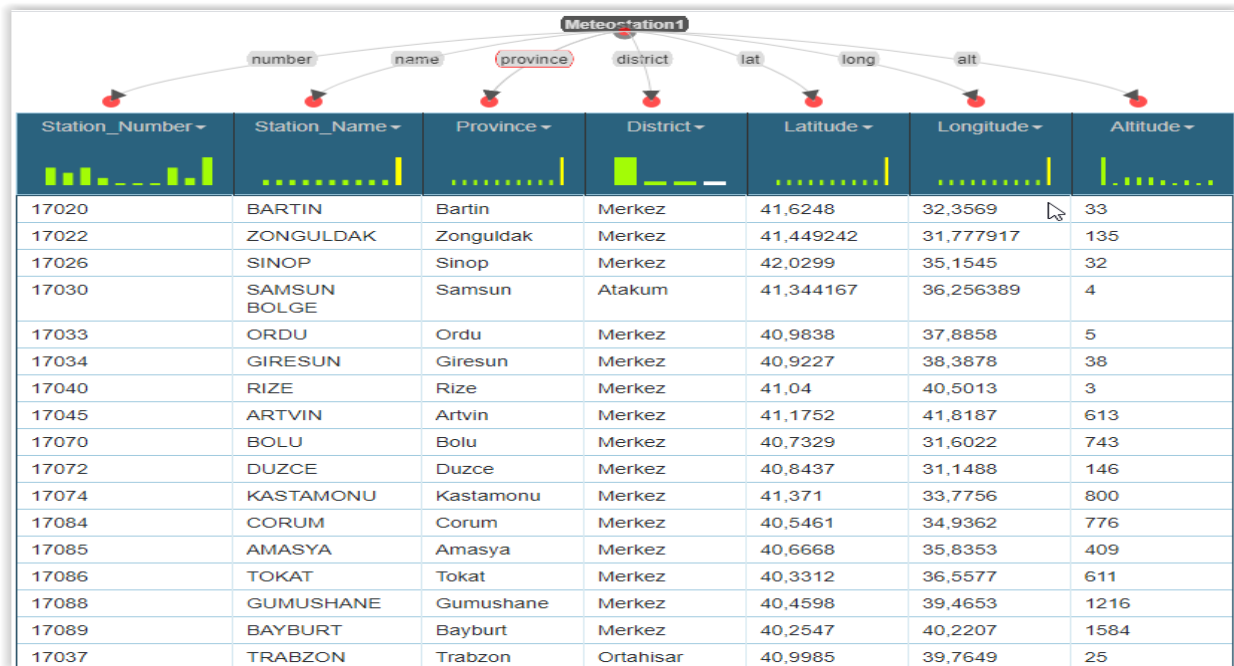
Şekil 6. Boylam özniteliğinin semantik tipinin belirlenmesi

Şekil 6’da görüldüğü üzere ‘Boylam’ özniteliğinin semantik tipi, WGS84 Geo Positioning ontolojisindeki “Longitude” özelliği ve Sweet ontolojisindeki ‘Meteostation’ sınıfı ile referanslandırılması ile belirlenmiştir. Veri setindeki diğer özniteliklerin semantik tipleri de benzer şekilde belirlenmiştir.

Veri modellemenin ikinci aşaması sınıflar arasındaki ilişkilerin belirlenmesidir. Şekil 7 ve 8’de oluşturulan model sınıfı, özellikleri ve ilişkileri ile gösterilmektedir.



Şekil 7. Gözlem verilerinin modeli



Şekil 8. İstasyon verilerinin modeli

Son aşamada Karma’da oluşturulan istasyon ve gözlem veri modelleri RDF formatında yayınlanmıştır. Oluşturulan RDF verisinin bir kısmı Şekil 9’da ‘Artvin’, ‘Giresun’ ve ‘Rize’ istasyon noktaları için gösterilmektedir.

```
[ rdf:type <http://sweetontology.net/matrFacility/Meteostation> ;
  <http://dbpedia.org/ontology/district> "Merkez" ;
  <http://dbpedia.org/ontology/number> "17034" ;
  <http://dbpedia.org/ontology/province> "Giresun" ;
  <http://purl.org/dc/terms/created> "Giresun" ;
  <http://www.geonames.org/ontology#name> "GİRESUN" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#alt> "38" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#lat> "40,9227" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#long> "38,3878"
] .

[ rdf:type <http://sweetontology.net/matrFacility/Meteostation> ;
  <http://dbpedia.org/ontology/district> "Merkez" ;
  <http://dbpedia.org/ontology/number> "17040" ;
  <http://dbpedia.org/ontology/province> "Rize" ;
  <http://purl.org/dc/terms/created> "Rize" ;
  <http://www.geonames.org/ontology#name> "RIZE" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#alt> "3" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#lat> "41,04" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#long> "40,5013"
] .

[ rdf:type <http://sweetontology.net/matrFacility/Meteostation> ;
  <http://dbpedia.org/ontology/district> "Merkez" ;
  <http://dbpedia.org/ontology/number> "17045" ;
  <http://dbpedia.org/ontology/province> "Artvin" ;
  <http://purl.org/dc/terms/created> "Artvin" ;
  <http://www.geonames.org/ontology#name> "ARTVIN" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#alt> "613" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#lat> "41,1752" ;
  <http://www.w3.org/2003/01/geo/wgs84_pos#long> "41,8187"
] .
```

Şekil 9. Artvin, Giresun ve Rize istasyon noktaları için oluşturulan RDF (.ttl) verisi

## 2.4 Verilerin Farklı Veri Setleri ile Bağlanması

Literatürde verinin harici Web kaynaklarına bağlanması konusunda Koho vd. (2018), Zhua vd. (2017), Wetz vd. (2012), Schabus ve Scholz (2017), Iwaniak vd. (2017), Qiu vd. (2017), Bischof vd. (2018), Margan vd. (2018) gibi akademik çalışmalar ve projeler mevcuttur.

Koho vd. (2018) tarihi verilerin harici ilgili veri setlerine bağlanmasını gerçekleştirmiştir. Finlandiya'daki ikinci Dünya savaşı mahkumları ile ilgili verileri Bağlantılı Veriye dönüştürerek WarSampo veri setine entegre etmiştir. Zhua vd. (2017) veri keşfinde konumsal verilerin özelliklerine ve rollerine göre konumsal alandaki Bağlantılı Veriler için çok boyutlu ve nicel bir bağlantı yaklaşımı önermiştir. Çin Ulusal Yer Sistemi Bilimsel Veri Paylaşım Ağında (NSTI-GEO) veri içi bağlantılar oluşturmuş ve NSTI-GEO'da Çin Meteorolojik Veri Ağı ve Ulusal Nüfus ve Sağlık Bilimsel Veri Paylaşım Platformu ile veri bağlantıları oluşturmuştur. Wetz vd. (2012) yerel bir eşanlamlılar sözlüğünde (thesaurus) yer alan bilgileri genişletmek ve zenginleştirmek için LOD (Bağlantılı Açık Veri) varlıklarını eş anlamlılar sözlüğüne bağlamıştır. Schabus ve Scholz (2017) bir üretim organizasyonu içinde farklı iş birimlerinde ve heterojen veri kaynaklarında yer alan veri setlerini, iç mekan alanını ve üretim süreçlerini açıklayan bir ontolojiye dayalı entegre etmek için Bağlantılı Veri yaklaşımı önermiştir. Qiu vd. (2017) çevresel modellerle afet verileri arasında bağlantı kuran ontoloji tabanlı bir yöntem önermiştir. Bischof vd. (2018) istatistiksel açık şehir verisini Bağlantılı Veri ile zenginleştirmiş ve eksik değerleri hesaplamıştır. Margan vd. (2018) hava kirliliği bilgilerini zenginleştirmek ve zararlı partikül kirliliği seviyelerine sahip alanları belirlemek için DBpedia'yi kullanmıştır.

Literatürde farklı veri kaynakları arasında bağlantı kurmak için "owl:sameAs" ve "owl:equivalentClass" OWL ilişkileri kullanılmaktadır. Farklı veri kaynakları arasında bağlantı kuran Silk, Limes, RIMOM, idMash ve ObjectCoref gibi yazılımlar mevcuttur. Çalışma kapsamında veri kaynakları arasında otomatik bağlantı oluşturmak için Silk bağlantı editörü kullanılmıştır. Silk, farklı Bağlantılı Veri kaynaklarına ait veriler arasındaki ilişkileri bulan bir bağlantı bulma çatısıdır. Öznitelik değerlerini veya varlık kümelerini karakter (string), sayısal, veri ve URI gibi benzerlik ölçütleri ve küme karşılaştırma yöntemlerinin yanı sıra bir kavram hiyerarşisinde iki kavram arasında semantik mesafeyi hesaplayan bir taksonomik eşleştirici ile karşılaştırır. Her metrik, '0' veya '1' arasındaki benzerlik değerine göre değerlendirilir, daha yüksek değerler daha büyük bir benzerliği belirtir (Volz vd., 2009).

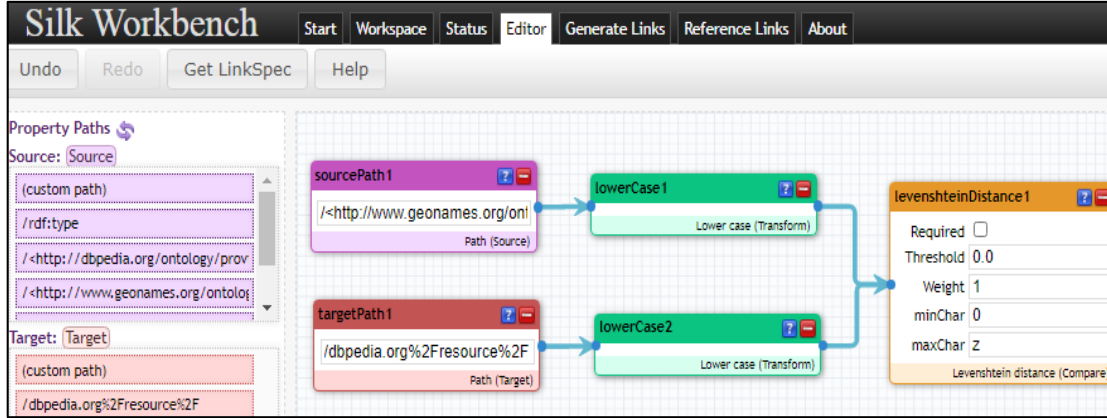
Çalışma kapsamında meteorolojik veriler ile DBpedia Web veri kaynağı arasında bağlantı kurulmuştur. Silk bağlantı bulma editöründe meteorolojik istasyon isimleri ve DBpedia il veri seti arasında kurulacak bağlantı kuralları şekil 10' daki gibi tanımlanmıştır. Karakter tabanlı Levenshtein Distance benzerlik algoritması ile bağlantılar kurulmuştur.



Levenshtein Distance karşılaştırma operatörü bir sözcüğü diğerine dönüştürmek için tek bir karakterin eklenmesi, silinmesi veya değiştirilmesi gibi düzenleme işlemlerini kapsayan minimum düzenleme sayısını hesaplayarak benzerlikleri bulur. Bulunan bağlantı sonuçlarının bir kısmı şekil 11’de gösterilmektedir. Şekilde görüldüğü üzere bağlantı sonuçları %100 doğrulukla bulunmuştur.

Silk’de meteorolojik istasyon koordinatları ve DBpedia il koordinatları veri seti arasında kurulacak bağlantı kuralları şekil 12’deki gibi tanımlanmıştır. Sayı tabanlı Numeric Similarity benzerlik algoritması ile bağlantılar kurulmuştur. Numeric Similarity karşılaştırma operatörü, veri setlerindeki rakamsal benzerlikleri hesaplayarak bağlantıları bulur. Bulunan bağlantı sonuçlarının bir kısmı şekil 13’de gösterilmektedir. Şekilde görüldüğü üzere bağlantı sonuçları çoğunlukla %90 üzeri doğrulukla bulunmuştur. Bunun sebebi kaynak veri setindeki illerin istasyon koordinatlarının DBpedia meteorolojik istasyon koordinatları yerine il koordinatları ile karşılaştırılmasıdır.

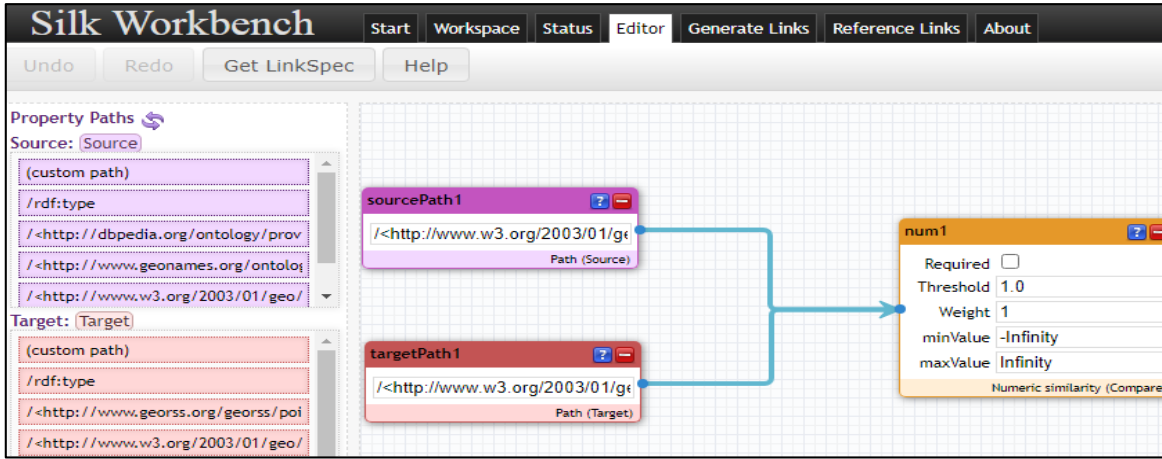
Uygulama kapsamında kaynak veri seti ile hedef veri seti arasındaki bağlantılar “owl: same as” ilişkisi ile kurulmuştur.



Şekil 10. İstasyon isimleri ve DBpedia illeri arasında bağlantı kurallarının tanımlanması

Source	Target	Score
http://www.co-ode.org/ontologies/ont.owl#Tokat		97
Comparison: levenshteinDistance (levenshteinDistance1)		100,0%
Transform: lowerCase (lowerCase1)	tokat	
Input: <http://www.geonames.org/ontology#name> (sourcePath1)	tokat	
Transform: lowerCase (lowerCase2)	tokat	
Input: dbpedia.org%2Fresource%2F (targetPath1)	Tokat	
http://www.co-ode.org/ontologies/ont.owl#Bolu		81
Comparison: levenshteinDistance (levenshteinDistance1)		100,0%
Transform: lowerCase (lowerCase1)	bolu	
Input: <http://www.geonames.org/ontology#name> (sourcePath1)	bolu	
Transform: lowerCase (lowerCase2)	bolu	
Input: dbpedia.org%2Fresource%2F (targetPath1)	Bolu	
http://www.co-ode.org/ontologies/ont.owl#Artvin		80
Comparison: levenshteinDistance (levenshteinDistance1)		100,0%
Transform: lowerCase (lowerCase1)	artvin	
Input: <http://www.geonames.org/ontology#name> (sourcePath1)	artvin	
Transform: lowerCase (lowerCase2)	artvin	
Input: dbpedia.org%2Fresource%2F (targetPath1)	Artvin	
http://www.co-ode.org/ontologies/ont.owl#Amasya		79
Comparison: levenshteinDistance (levenshteinDistance1)		100,0%
Transform: lowerCase (lowerCase1)	amasya	
Input: <http://www.geonames.org/ontology#name> (sourcePath1)	Amasya	
Transform: lowerCase (lowerCase2)	amasya	
Input: dbpedia.org%2Fresource%2F (targetPath1)	Amasya	

Şekil 11. İstasyon isimleri ve DBpedia illeri arasında kurulan bağlantı sonuçları



Şekil 12. İstasyon koordinatları ve DBpedia il koordinatları arasında bağlantı kurallarının tanımlanması

Source: Source	Target: Target	Score
http://www.co-ode.org/ontologies/ont.owl#Trabzon_Bolge	http://tr.dbpedia.org/resource/Trabzon_Havalimani	99,7%
Comparison.num (num1) 99,7% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 40.9985 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 40.995834		
http://www.co-ode.org/ontologies/ont.owl#Trabzon_Bolge	http://tr.dbpedia.org/resource/Karadeniz_Teknik_Üniversitesi	99,4%
Comparison.num (num1) 99,4% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 40.9985 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 40.992428		
http://www.co-ode.org/ontologies/ont.owl#Sinop	http://tr.dbpedia.org/resource/Sinop_Havalimani	98,6%
Comparison.num (num1) 98,6% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 42.0299 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 42.015835		
http://www.co-ode.org/ontologies/ont.owl#Amasya	http://tr.dbpedia.org/resource/Amasya	98,3%
Comparison.num (num1) 98,3% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 40.6668 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 40.65		
http://www.co-ode.org/ontologies/ont.owl#Samsun_Bolge	http://tr.dbpedia.org/resource/Samsun_Saat_Kulesi	95,1%
Comparison.num (num1) 95,1% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 41.344166 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 41.294724		
http://www.co-ode.org/ontologies/ont.owl#Kastamonu	http://tr.dbpedia.org/resource/Kastamonu_Üniversitesi	94,7%
Comparison.num (num1) 94,7% Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (sourcePath1) 41.371 Input: <http://www.w3.org/2003/01/geo/wgs84_pos#lat> (targetPath1) 41.424168		

Şekil 13. İstasyon koordinatları ve DBpedia il koordinatları arasında kurulan bağlantı sonuçları.

### 3. Sonuç ve Öneriler

Bu çalışmada meteorolojik verilerin bilgisayarların/yazılımların anlayacağı biçimde anlamlandırılması, herkesin erişebileceği şekilde paylaşılma açılıp farklı veri setleri ile entegre edilerek zenginleştirilmesi amaçlanmıştır. Bu amaçla çalışmada Karadeniz Bölgesine ait meteorolojik verilerin bir kısmının semantik olarak tanımlanması ve Bağlantılı Veriye dönüştürülmesi gerçekleştirilmiştir. Dönüşüm işleminde Web'de mevcut Sweet, Geonames, DBpedia WGS84 Geo Positioning ve W3C zaman ontolojileri kullanılmıştır.

Meteorolojik istasyon ve gözlem verileri ilgili ontolojilerle referanslandırılarak Bağlantılı Veri için gerekli olan RDF verisi üretilmiştir. Üretilen meteorolojik RDF verisi ile açık Bağlantılı Veri kaynağı olan DBpedia veri seti arasındaki bağlantılar Silk bağlantı editörü'nde otomatik olarak bulunmuştur. Böylece meteorolojik veriler global DBpedia Web kaynağı ile zenginleştirilmiştir. Bulunan bağlantı sonuçları bu iki ayrı veri seti arasındaki benzerliğin yüksek olduğunu göstermektedir.

Çalışmada kullanılan veri setindeki kavramların Türkçe karakterler içermesi bağlantı kurma aşamasında problem oluşturmuştur. Problemi gidermek için veri seti Türkçe karakterlerden ayıklandıktan sonra kullanılmıştır. Ayrıca, Karma ile üretilen RDF veri modelinde istasyon ve gözlem veri setindeki özniteliklerin RDF 'data property' özelliklerine dönüştürülmesi gerekirken, bir kısmının RDF 'annotation property' özelliklerine dönüştüğü, bir kısmının ise hiçbir RDF özelliğine dönüşmediği görülmüştür. Bu nedenle çalışmada üretilen RDF verisi Protege ontoloji editörü ile düzenlenip eksikleri tamamlandıktan sonra kullanılmıştır. Türkçe karakterleri tanıyan bağlantı kurma yazılımlarının geliştirilmesi gereklidir.

Gelecek çalışmalarımızda meteorolojik verilerin tamamının Bağlantılı Açık Veriye dönüştürülmesi, yayınlanması, harita üzerinde görselleştirilmesi ve Geonames, GeoWordnet gibi farklı Web kaynakları ile bağlanması gerçekleştirilecektir. Ayrıca, aynı verilerle farklı bağlantı kurma araçları kullanılarak yazılımların bağlantı kurma performansı test edilecektir.

## Teşekkür

Çalışma kapsamında kullanılan meteorolojik veriler Meteoroloji Genel Müdürlüğü'nün izni ile kullanılmıştır. Meteorolojik verileri sağlayan Meteoroloji Genel Müdürlüğü'ne teşekkür ederiz.

## Kaynaklar

- Atemezing, G., Corcho, O., Garijo, D., Mora, J., Villalón, P., M., Rozas, P., Suero, V.D., & Terrazas, V.B. (2011). Transforming meteorological data into linked data. *Semantic Web journal*, 1(2011), 1-5.
- BioPortal. (2021, Ocak 21). Semantic Web for Earth and Environment Technology Ontology. Retrieved from <https://bioportal.bioontology.org/ontologies/SWEET>.
- Bischof, S., Harth, A., Kämpgen, B., Polleres, A. & Schneider, P. (2018). Enriching integrated statistical open city data by combining equational knowledge and missing value imputation, *Web Semantics: Science, Services and Agents on the World Wide Web*, 48, 22-47, <https://doi.org/10.1016/j.Websem.2017.09.003>.
- Consolia, S., Presuttia, V., Recuperoa, R.D., Nuzzolesea, G.A., Peronia, S., Mongioli, M., Gangemi, A. (2017). Producing Linked Data for Smart Cities: The Case of Catania, *Big Data Research*, 7, 1-15.
- DBpedia. (2021, Ocak 21). Retrieved from <https://dbpedia.org/ontology/>.
- DBpedia Mappings Wiki (2021, Ocak 21). Retrieved from [http://mappings.dbpedia.org/index.php/Main\\_Page](http://mappings.dbpedia.org/index.php/Main_Page).
- Debruyne, C., Clinton, É., McNeerney, L., Nautiyal, A. & O'Sullivan, D. (2016). Serving Ireland's Geospatial Information as Linked Data, *International Semantic Web Conference 2016*, October 17-21, Kobe, Japan.
- Geo. (2021, Ocak 21). Retrieved from <https://www.w3.org/2003/01/geo/>.
- GeoNames. (2021, Ocak 21). Retrieved from <https://www.geonames.org/>.
- GeoNames Ontology. (2021, Ocak 21). Retrieved from <http://www.geonames.org/ontology>.
- Hamdi, F., Abadie, N., Bucher, B., & Feliachi, A. (2015). GeomRDF: A geodata converter with a fine-grained structured representation of geometry in the Web. *arXiv preprint*, arXiv:1503.04864.
- Hietanen, E., Lehto, L., Latvala, P. (2016). Providing Geographic Datasets As Linked Data In SDI, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B2, 2016 XXIII ISPRS Congress, 12-19 July 2016, Prague, Czech Republic.
- Iwaniak, A., Leszczuk, M., Strzelecki, M., Harvey, F., & Kaczmarek, I. (2017). A novel approach for publishing linked open geodata from national registries with the use of semantically annotated context dependent web pages. *ISPRS International Journal of Geo-Information*, 6(8), 252. doi:10.3390/ijgi6080252.
- Kara, G., Yılmaz, C., Ulutaş Karakol, D., Akyazı, İ. & Cömert, Ç. (2018). Bağlantılı Verilerin Yayınlanması ve Görselleştirilmesi, *VII. Uzaktan Algılama-CBS Sempozyumu (Uzal-CBS 2018)*, 18-21 Eylül 2018, Eskişehir.
- Kara, G., Akyazı, İ., Cömert, Ç. (2020). Konumsal Verilerin Bağlantılı Veri Olarak Yayınlanması: Trabzon Örneği. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, 36(2), 228-237.
- Karma. (2021, Ocak 21). A Data Integration Tool. Retrieved from <https://usc-isi-i2.github.io/karma/>.
- Koho, M. K., Heino, E., Ikkala, E., Hyvönen, E. A., Nikkilä, R., Miettinen, K., & Suominen, P. (2018). Integrating prisoners of war dataset into the warsampo linked data infrastructure. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*. CEUR Workshop Proceedings. 241-249.

- Kyzirakos, K., Vlachopoulos, I., Savva, D., Manegold, S. & Koubarakis, M. (2014). GeoTriples: a Tool for Publishing Geospatial Data as RDF Graphs Using R2RML Mappings, *International Semantic Web Conference (Posters & Demos)* (pp. 393-396).
- Margan, B., Hakimpour, F., & Saber, M. (2018, April). Linked data geo-statistical analysis of air pollution in urban areas. In 2018 4th International Conference on Web Research (ICWR) (pp. 86-91). IEEE.
- OpenRefine. (2021, Ocak 21). Retrieved from <https://openrefine.org/>.
- Patroumpas, K., Alexakis, M., Giannopoulos, G., & Athanasiou, S. (2014, March). TripleGeo: an ETL Tool for Transforming Geospatial Data into RDF Triples. In *Edbt/Icdt Workshops* (pp. 275-278).
- Qiu, L., Du, Z., Zhu, Q. & Fan, Y. (2017). An integrated flood management system based on linking environmental models and disaster-related data, *Environmental Modelling & Software*, 91, 111-126.
- Saavedra, J., Vilches-Blázquez, L.M. & Boada, A. (2014). Cadastral data integration through Linked Data, *Proceedings of the AGILE 2014 International Conference on Geographic Information Science*, Castellón.
- Schabus, S., & Scholz, J. (2017). Spatially-Linked Manufacturing Data to Support Data Analysis, *GI\_Forum 2017*, 1, 126 – 140, DOI: 10.1553/giscience2017\_01\_s126.
- Szekely, P., Knoblock, A., C., Yang, F., Zhu, X., Fink, E., E., Allen, R. & Goodlander, G. (2013), Connecting the Smithsonian American Art Museum to the Linked Data Cloud, *P. Cimiano et al. (Eds.): ESWC 2013*, LNCS 7882, pp. 593-607, Springer-Verlag, Berlin Heidelberg.
- The LOD cloud. (2021, Ocak 21). Retrieved from <https://lod-cloud.net/#about>.
- Berners-Lee, Tim. (2021, Ocak 21). Linked Data, Retrieved from, <https://www.w3.org/DesignIssues/LinkedData.html>.
- Time. (2021, Ocak 21). Retrieved from <https://raw.githubusercontent.com/w3c/sdw/gh-pages/time/rdf/time.ttl>
- Time Ontology in OWL (2021, Ocak 21). Retrieved from <https://www.w3.org/TR/owl-time/>
- Ulutaş Karakol, D., Kara, G., Yılmaz, C. & Cömert, Ç. (2018). Semantic Linking Spatial RDF Data to the Web Data Sources. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-4, 2018 ISPRS TC IV Mid-term Symposium “3D Spatial Information Science-The Engine of Change”, 1-5 October 2018, Delft, The Netherlands.
- Volz, J., Bizer, C., Gaedke, M., & Kobilarov, G. (2009, January). Silk-a link discovery framework for the web of data. In *Ldow. LDOW 2009*, April 20, Madrid, Spain.
- Wetz, P., Stern, H., Jakobitsch, J. & Pammer, V. (2012). Matching Linked Open Data Entities to Local Thesaurus Concepts, *Proceedings of the I-SEMANTICS 2012 Posters & Demonstrations Track*, pp. 6-11.
- Zhua, Y., Zhu, A., Song, J., Yang, J., Feng, M., Sun, K., Zhang, J., Hou, Z. & Zhao, H. (2017). Multidimensional and quantitative interlinking approach for Linked Geospatial Data, *International Journal Of Digital Earth*, 10(9), pp. 923-943, <http://dx.doi.org/10.1080/17538947.2016.1266041>.