

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(DOKTORA TEZİ)

**AÇIK BAĞLI VERİ SİSTEMLERİNDE
KÖKEN BAZLI ERİŞİM GERÇEKLEŞTİRİMİ**

Fatih TEKBACAK

Tez Danışmanı: Prof. Dr. Oğuz DİKENELLİ

Bilgisayar Mühendisliği Anabilim Dalı

Sunuş Tarihi: 23.Aralık.2015

Bornova-İZMİR

2015

Fatih TEKBACAK tarafından **DOKTORA TEZİ** olarak sunulan "**AÇIK BAĞLI VERİ SİSTEMLERİNDE KÖKEN BAZLI ERİŞİM GERÇEKLEŞTİRİMİ**" başlıklı bu çalışma E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliği ile E.Ü. Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi'nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve **23.12.2015** tarihinde yapılan tez savunma sınavında aday oybirliği/oyçokluğu ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı: Prof. Dr. Oğuz DİKENELLİ

.....

Raportör Üye: Yrd. Doç. Dr. Selma TEKİR

.....

Üye: Yrd. Doç. Dr. Tuğkan TUĞLULAR

.....

Üye: Yrd. Doç. Dr. Şebnem BORA

.....

Üye: Doç. Dr. Murat Osman ÜNALIR

.....

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ETİK KURALLARA UYGUNLUK BEYANI

E.Ü. Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Doktora Tezi olarak sunduğum "**AÇIK BAĞLI VERİ SİSTEMLERİNDE KÖKEN BAZLI ERİŞİM GERÇEKLEŞTİRİMİ**" başlıklı bu tezin kendi çalışmam olduğunu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

23.12.2015

.....

Fatih TEKBACAK

ÖZET**AÇIK BAĞLI VERİ SİSTEMLERİNDE
KÖKEN BAZLI ERİŞİM GERÇEKLEŞTİRİMİ**

TEKBACAK, Fatih

Doktora Tezi, Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Oğuz DİKENELLİ

Aralık 2015, 82 sayfa

Köken, vebdeki verinin güven ve kalite değerlendirilmesi sürecine yön veren bir kavram olarak ortaya atılmıştır. Bağlı veriyi üreten veya tüketen uygulamalar, köken üstverisi yardımı ile güvene dair tanımlamalar sunmaktadır. Tezde köken üstverisine sahip verisetleri üzerinde yapılan sorgulardan elde edilen filtrelenmiş veri üzerinde durulmaktadır. Güvenin artırılması için köken üstverisi kullanılarak sorgulanacak verisetlerinden en uygun olanlarının filtrelenmesi ve dağıtık sorgu motoruna aktarılması üzerinde çalışılmıştır.

Tezin ilk katkısı, köken farkında bir bağlı veri sorgulama altyapısı ortaya koymaktır. Köken üstverisinin gösterimi ve bağlı verinin organizasyon tarafından sorgulanması sırasında işleyecek süreç altyapı içerisinde tanımlanmaktadır.

Tezin ikinci katkısı, mevcut verisetleri arasından köken üstverisine göre en uygun olanları filtrelemektir. Çalışma içinde filtreleme yapılırken verisetleri hakkındaki üstveri kullanılmaktadır. Organizasyonel tercihler, veriseti üstverisi ve geliştirilen ontolojilerden yararlanılan bir mimari sunulmaktadır.

Tezin üçüncü katkısı, bağlı veri literatürüne İç Veri ve Dış Veri kavramlarını katmasıdır. Bağlı veriye organizasyonel perspektiften bakıldığında, vebde erişilebilecek anlamsal olan veya olmayan tüm veri ile bu verilerin üstverileri Dış Veri olarak adlandırılmıştır. İç Veri ise organizasyon verisi, organizasyonel tercihler ve organizasyonel veri kaynağı üstmodel ambarı olarak Dış Veri’de erişilecek verisetlerinin üstverisini içerir.

Tezin dördüncü katkısı, Dış Veri’ye ait olan veriseti üstverisine eklenebilecek köken üstverisini ifade etmektir. Verisetinin ham hali, anlamsal hali ve üstverisi ile ilgili köken tanımları netleştirilmiştir.

Tezin beşinci katkısı, İç Veri'yi oluşturan organizasyonel kavramlar ile kökeni beraberce gözönüne alan bir ontoloji oluşturmaktır. Ontoloji, çalışma alanından bağımsız olarak kökeni sorgulamak için terimler içermektedir. Ayrıca organizasyonların birbirleri hakkında fikir sahibi olmasına olanak sağlamaktadır.

Tezin altıncı katkısı, organizasyonların köken tercihlerini gösteren bir ontoloji oluşturmaktır. Köken boyutları ve ilgili köken tercihleri çeşitlenmekle beraber farklı durumları kapsayan örnek tercih boyutları ele alınmıştır. Bu doğrultuda organizasyonel tercihlerin nasıl oluşturulabileceği örneklendirilmiştir.

Tezin son katkısı, İç Veri, Dış Veri ve oluşturulan ontolojilerden yararlanarak köken kuralları meydana getirmektir. Mevcut verisetleri üzerinde kuralların yürütülmesi ile sorgu motorunun kullanacağı veriseti listesinin son halinin elde edilmesi sağlanmaktadır.

Anahtar Sözcükler: Köken, Bağlı Veri, Güven, Bağlı Verisetleri Sözlüğü, Köken Üstverisi, Köken Kuralları.

ABSTRACT**PROVENANCE-AWARE QUERY EXECUTION
IN OPEN LINKED DATA SYSTEMS**

TEKBACAK, Fatih

PhD. in Computer Engineering

Supervisor: Prof. Dr. Oğuz DİKENELLİ

December 2015, 82 pages

Provenance is suggested as a concept which directs the trustworthiness and quality assurance process of data on the Web. The applications publishing and consuming linked data represent definitions about trust with the help of provenance metadata. In the thesis, filtered data obtained from queries concerning metadata of datasets is elaborated. To enhance trustworthiness, it is worked on filtering most suitable datasets with provenance metadata to be queried and transferring these datasets to federated query engine.

The first contribution of this thesis is to introduce a provenance-aware linked data querying infrastructure. In the infrastructure, the representation of provenance metadata and the process during querying linked data by the organization is described.

The second contribution of this thesis is to filter most appropriate datasets between available datasets towards provenance metadata. In the work, metadata about datasets are used for filtering. An architecture benefited from metadata and required vocabularies is represented by considering organizational preferences and dataset-based requirements.

The third contribution of this thesis is to append Interior Data and Exterior Data concepts to the linked data literature. When observing linked data from organizational perspective, whole accessible raw or semantic data on the web and the metadata of those data is called as Exterior Data. Interior Data is the aggregation of organization data, organizational preferences and metadata of accessible datasets on the web which is referred to organizational data source metamodel warehouse.

The fourth contribution of this thesis is to express provenance metadata for

extending dataset metadata which belongs to Exterior Data. Provenance definitions related to raw dataset, semantic dataset and dataset metadata are clarified.

The fifth contribution of this thesis is to build an ontology that considers organizational and provenance concepts constructing Interior Data. The ontology includes terms for querying provenance independent of the domain. Additionally, it helps organizations to have ideas about each other.

The sixth contribution is to build another ontology that represents provenance preferences of the organizations. While there are different provenance dimensions and related different provenance preferences, sample preference dimensions involving different conditions are handled. Accordingly, method of creating organizational preferences are exemplified.

The last contribution of this thesis is to generate rules by the help of Internal Data, External Data and built ontologies. Final list of datasets used by query engine are obtained by enforcing rules on available datasets.

Keywords: Provenance, Linked Data, Trust, VoID (Vocabulary of Interlinked Datasets), Provenance Metadata, Provenance Rules.

TEŞEKKÜR

Doktora sürecim boyunca bana sabırla en büyük desteği sağlayan annem Rabia Ayten Örmeci'ye ve eşim Buket Ülgen Aydın'a sonsuz teşekkürlerimi sunuyorum.

Çalışmalarımızı beraber yürüttüğümüz Emrah İnan'a, takıldığımız noktalarda çözümler üretmemize katkıda bulunan Tayfun Gökmen Halaç, Ziya Akar ve Burak Yönyül'e, tez yazım süreci boyunca yardımlarını esirgemeyen Kasım Sinan Yıldırım ve Birol Çiloğlugil'e, yol gösterici tavırlarıyla danışmanım Prof. Dr. Oğuz Dikenelli ve üniversitedeki eğitim hayatım boyunca verdiği destekle Yrd. Doç. Dr. Tuğkan Tuğlular'a teşekkür ederim.

Son olarak, 2211-Yurt İçi Doktora Burs Programı kapsamında beni destekleyen TÜBİTAK'a teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vii
ABSTRACT	viii
TEŞEKKÜR	xi
ŞEKİLLER DİZİNİ.....	xix
ÇİZELGELER DİZİNİ.....	xxi
SİMGELER VE KISALTMALAR DİZİNİ	xxiii
1 GİRİŞ	1
2 ÖNCEKİ ÇALIŞMALAR	7
2.1 Bağlı Veri.....	7
2.1.1 Void	9
2.1.2 Adlı çizge	10
2.1.3 Bağlı veri yayımlama.....	11
2.1.4 Bağlı veri tüketme	11
2.2 Köken.....	14
2.3 Veri Uzayı.....	19
3 KÖKEN FARKINDA ORGANİZASYONLAR İÇİN OLUŞTURULAN ONTOLOJİLER.....	24
3.1 Örnek Senaryo.....	24

İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
3.2 Köken Farkında Organizasyonlar için Oluşturulan Ontolojiler	26
3.2.1 Dış veri bakış açısı	28
3.2.2 İç veri bakış açısı	31
3.3 Kullanılan Senaryo ve Ontolojiler	38
3.4 Adlı Çizgeler	39
4 BAĞLI VERİ ALTYAPISINI GELİŞTİREN KÖKEN FARKINDA FİLTRELEME	43
4.1 Köken Farkında Filtreleme Süreci	43
4.2 Köken Farkında Bağlı Veri Sorgulama Yazılım Mimarisi	47
4.3 İş Akışı	50
5 DURUM ÇALIŞMASI	53
5.1 Durum Çalışmasında Köken İhtiyacının Ele Alınması	53
5.2 Güven Senaryo Adımları	54
5.2.1 Dış veri tabanlı operasyonlar	54
5.2.2 İç veri tabanlı operasyonlar	56
5.2.3 Adlı çizgelerin sorgulanması	57
5.3 Güven Ağı Senaryo Adımları	61
5.3.1 Dış veri tabanlı operasyonlar	61

İÇİNDEKİLER (Devam)

	<u>Sayfa</u>
5.3.2 İç veri tabanlı operasyonlar	62
5.3.3 Adli çizgelerin sorgulanması	62
5.3.4 Sertifika sahibi kişilerle ilgili verisetlerinin doğrulanması	65
6 TARTIŞMA VE ÖNERİLER.....	67
7 SONUÇ.....	73
KAYNAKLAR DİZİNİ	74
ÖZGEÇMİŞ	82
EKLER	

ŞEKİLLER DİZİNİ

<u>Şekil</u>	<u>Sayfa</u>
1.1 Bağlı veri uzayı ve organizasyon gösterimi	2
2.1 Örnek RDF bağı	9
2.2 Tez çalışmasına adapte edilmiş veri yayımlama akışı	12
2.3 Tez çalışmasına adapte edilmiş veri tüketim akışı	14
2.4 Veri yayımlama ve tüketimini gözönüne alan veri uzayı katmanları(Bizer, 2011)	21
3.1 Ham veri, anlamsal veri ve VoID üstverisinin örnek gösterimi	25
3.2 VoID üzerinde kökeni ifade etmek için kullanılan senaryo	26
3.3 PAV Ontolojisi ile VoID Dokümanlarının Köken için Genişletilmesi (PEVoID)	28
3.4 Organizasyonel Köken Ontolojisi	32
3.5 Haber Toplama Senaryosu için Organizasyonel Köken Ontolojisi	33
3.6 Organizasyonel Tercih Ontolojisi.....	35
3.7 Farklı organizasyonel köken tercihlerinin gösterimi	37
3.8 VoID Köken Üstverisini İçeren Veriseti Tabanlı Adlı Çizgenin EBNF Dilbilgisi	40
3.9 Ham veya Yayımlanmış Verisetine Ait Köken Adlı Çizgesinin EBNF dilbilgisi.....	40
3.10 Alana Dair Organizasyonel Adlı Çizgenin EBNF dilbilgisi	41
3.11 Tercih Adlı Çizgesinin EBNF dilbilgisi	41

ŞEKİLLER DİZİNİ (Devam)

<u>Şekil</u>	<u>Sayfa</u>
3.12 Köken kuralının EBNF dilbilgisi	42
4.1 Köken Farkında Sorgulama Altyapısı	44
4.2 Köken Farkında Filtreleme Süreci	46
4.3 PROV4FQ Yazılım Mimarisi	48
4.4 PROV4FQ İçerisinde SPARQL CONSTRUCT Sorgusunun İlkendirilmesi	50
4.5 SPARQL CONSTRUCT Sorgusunun İlkendirilmesi	51
5.1 NYT verisetine dair VoID dokümanındaki adlı çizgelerin ifade edilmesi ..	55
5.2 VOID dokümanı içerisindeki güvene dair ifadeler	55
5.3 Organizasyonun güven tanımlamaları içeren alana dair adlı çizgesi	56
5.4 Organizasyonun tercihlerine dair adlı çizge	57
5.5 Adlı çizgeleri sorgulayan örnek kural tanımı	58
5.6 Dış Veri ve İç Veri'yi içeren adlı çizgeler	59
5.7 PROV4FQ Veri Akışı	60
5.8 VoID dokümanı içerisindeki güvene dair ifadeler	62
5.9 Organizasyonel veriye dair adlı çizgeler	63
5.10 Örnek Güven Ağı kural tanımı-1	64
5.11 Örnek Güven Ağı kural tanımı-2	64

ŞEKİLLER DİZİNİ (Devam)

<u>Şekil</u>	<u>Sayfa</u>
5.12 Verisetine dair VOID’de bulunan uçnoktası bilgisi ile sertifika bilgisinin eşleştirildiği test kod parçası	65
6.1 Kaliteyi ve kökeni ele alan çatıların karşılaştırılması(Zaveri et al., 2015) ..	68
6.2 Veriseti Değişimini Ele Alan Bağlı Veri Ortam Altyapısı	71

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
3.1 VoID Dokümanına Eklenen Veri Yapılarına Uygun PAV Köken Özellikleri	30
3.2 Organizasyonel Tercihler Oluşturulurken Dikkate Alınacak Elemanlar	36

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltma	Açılım
LOD	Linked Open Data (Bağlı Açık Veri)
NG4J	Named Graphs API for Jena (Jena için Adlı Çizgeler Kütüphanesi)
PAV	Provenance, Authoring and Versioning Ontology (Köken, Yazarlık ve Versiyonlama Ontolojisi)
PROV-AQ	Provenance Access and Query (Köken Erişim ve Sorgulaması)
PROV4FQ	Provenance for Federated Queries (Dağıtık Sorgular için Köken Servisi)
RDF	Resource Description Framework (Kaynak Tanımlama Çerçevesi)
SPARQL	Simple Protocol and RDF Query Language (Yalın Protokol ve RDF Sorgu Dili)
SWP	Semantic Web Publishing Vocabulary (Anlamsal Veb Yayımlama Sözlüğü)
SWRL	Semantic Web Rule Language (Anlamsal Veb Kural Dili)
TriG	Triple Graph Syntax (Üçlü Çizge Sözdizimi)
URI	Uniform Resource Identifier (İnternet Kaynak Tanımlayıcısı)
VoID	Vocabulary of Interlinked Datasets (Birbirine Bağlı Verisetleri Sözlüğü)
VoIDP	Provenance Extension to VoID (VoID için Köken Eklentisi)
WoDQA	Web of Data Query Analyzer (Veri Vebi Sorgu Analizcisi)

1 GİRİŞ

Günümüze kadar veb, insanların anlaşması için tasarlanan HTML sayfaları ve bu sayfalar arasındaki bağlantılardan oluşmaktaydı. Veb içeriğinin makine tarafından yorumlanmasına olanak sağlayan RDF¹, RDFS² ve OWL³ gibi veri modelleri bilginin sadece insanlar tarafından anlaşılır olmanın ötesinde makineler tarafından da işlenmesini sağlamıştır. Bu dillerden yararlanan *Bağlı Veri*⁴, farklı veri kaynakları arasında anlamsal bağlantılar sağlayarak *Veri Vebi*'ni⁵ oluşturur. *Bağlı Veri*, veri kaynaklarını *bağlı veri uzayı*⁶ içerisinde birleştirir. *Veri uzayı*⁷, heterojen veri kaynakları üzerinde duran güçlü bir kavramsal modeldir ve bir sanallaştırma katmanı sağlar. *Bağlı veri uzayı*nda ise veb standartlarının kullanımı ve yaygın olarak anlaşılan bir veri modeli ile bu katmanda ihtiyaç duyulan uygulamalar çalıştırılır.

Bir organizasyonun kullandığı dağıtık veriyi yönetebilmek için veritabanı alanında veri uzayı (Franklin et al., 2005; Halevy et al., 2006) kavramı halihazırda tanımlanmıştır ve temel servisler belirlenmiştir (Franklin et al., 2005; Heath and Bizer, 2011). Bu tanım kapsamında veri uzayına dair servis ve veriye dair bileşenler net bir şekilde ifade edilmelidir. Bu servislerden başlıcaları; sorgulama yapmak, veri kökenini izlemek ve veri/üstveriye dair değişimleri yönetebilmek olarak sayılabilir (Franklin et al., 2005). Veri uzayı kavramı literatürde bağlı veri için açıklanmakla beraber (Bizer et al., 2009), bu servislerin veri/servis bağlantıları detaylı olarak incelenmemiştir.

Veri uzayı, veri entegrasyon yaklaşımından öte verinin birlikte var olduğu ve erişilebildiği bir sistem olarak ifade edilir (Franklin et al., 2005). Veri uzayı organizasyon ölçeğinde veri yönetimi açısından ele alındığında, organizasyon ile organizasyonun kullanacağı servis ve veri arasındaki ilişkinin netleştirilmesi gerekmektedir. Organizasyonlar tarafından günümüzde karşılaşılan en ciddi sorunlardan biri, aşırı sayıda ve birbirleri ile ilişkili veri kaynaklarına ihtiyaç duymalarıdır. Organizasyonların belirli bir kapsam içerisinde ihtiyaçları olan veri kaynaklarını etkin kullanmaları için veri uzayında ilgili servisler bulunmalıdır. Veri uzayı, bu servisleri sağlarken verinin nasıl bütünleştirileceğinden organizasyondaki kullanıcıları bağımsız tutacak çözümler sunmalıdır. Örneğin, organizasyon

¹*Açılımı:* Resource Description Format

²*Açılımı:* RDF Schema

³*Açılımı:* Web Ontology Language

⁴*İngilizcesi:* Linked Data

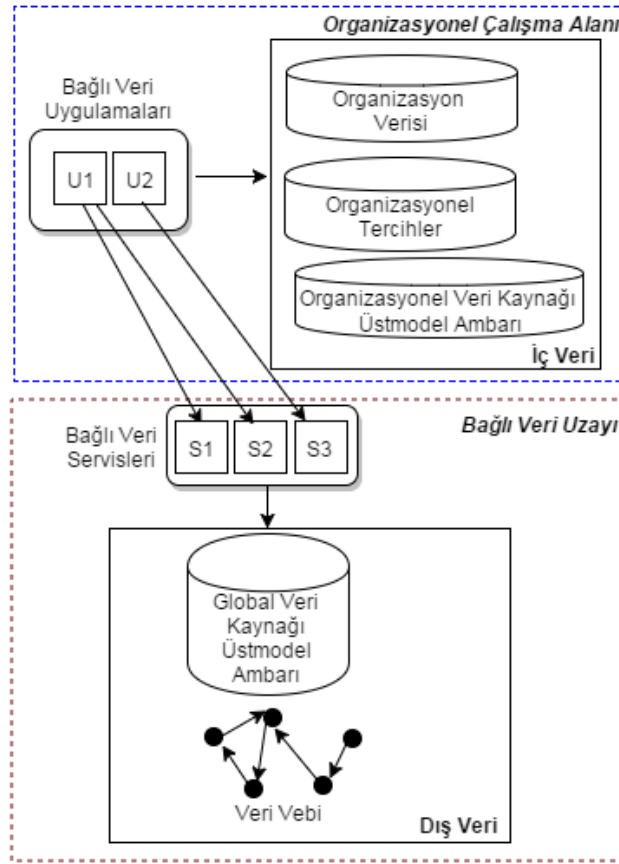
⁵*İngilizcesi:* Web of Data

⁶*İngilizcesi:* Linked dataspace

⁷*İngilizcesi:* Dataspace

verisetleri üzerinde sorgulama yapmak isterken sorgulamanın nasıl yapılıp cevabın nasıl oluşturulacağına dair hizmetlerden veri uzayı sorumludur.

Veri vebinin organizasyon seviyesinde kullanılması için veri uzayı tanımını bağlı veri için genişletmek gerekmektedir. Bağlı veri uzayı için bir kavramsal mimari Şekil 1.1’de gösterilmiştir. Görüldüğü üzere bağlı veri uzayında bağlı veri servisleri ve bağlı veri uygulamaları bulunmaktadır. Bağlı veri uygulamaları bir veya birden fazla bağlı veri servisini kullanır.



Şekil 1.1: Bağlı veri uzayı ve organizasyon gösterimi

Bu tezde öncelikle organizasyonların kendi İç Verilerine ek olarak Veri Vebi'ne erişimleri için bir veri uzayı mimarisi tanımlanmaktadır. Organizasyonların kullandığı veri, İç Veri ve Dış Veri olarak iki farklı şekilde ifade edilmektedir. İç Veri, organizasyonun kendi verisini, organizasyona ait tercihleri ve yerel veri kaynağı üstmodel ambarında tutulan organizasyonun ihtiyaç duyduğu veri kaynaklarına dair veriyi tanımlar. Organizasyonel tercih, Dış Veri'ye erişirken elde edilecek veri kümesine dair talepleri ifade eder. Örneğin organizasyon, haberler ile ilgili sadece Türkiye'den yayımlanmış verisetlerini almayı tercih edebilir. Dış Veri, Veri Vebi'nde bulunan ve veri uzayı yardımı ile erişilen veriyi göstermektedir.

Sorgu alıřtırma, baėlı veri uzayının Veri Vebinde bulunan veriyi tüketeabilmek iin kullandıėı temel servistir. Baė dolařma (Hartig et al., 2009) ve sorgu daėıtılması (Görlitz and Staab, 2011a), sorgu alıřtırma iin var olan iki yaklařımdır. Baė dolařma, baėlı veri izgesinin bir düėümünden bařlar ve düėümler arasındaki iliřkileri izleyerek uygun veriyi keřfeder. Baė dolařma yaklařımının avantajı, güncel veri elemanlarını alıřma zamanında keřfedebilmesidir. Bununla birlikte sonuçlar keřfe bařlanılan düėüme göre deėiřtiėinden uygun düėümlere bařarılı řekilde eriřilemeyebilmektedir. İkinci yaklařım olan sorgu daėıtılması, sorguları alt-sorgulara böler. Bu alt-sorgular iin üstveriye baėlı olarak uygun verisetleri seilir. Sorgu daėıtılması talep edilen sonuca eriřimi garantiye alırken, baė dolařma ok büyük boyuttaki baėlı veri bulutunda doėru sonucu bulmada bařarısız olabilmektedir.

Baėlı veri uzayı kavramını ve servislerini Veri Vebi'ni de iine alacak řekilde geniřletmek gerekmektedir. řekil 1.1'de ifade edilen mimariye uygun řekilde tez kapsamında, köken verisini kullanarak sorgulanacak verisetlerini filtreleyen bir sorgu servisi geliřtirilmiřtir. Bu servis, sorgu daėıtımı yaklařımına göre organizasyonun kullandıėı sorgulama servisi ile beraber kalite deėerlendirme kriterlerinden biri olan kökeni de göz önüne almaktadır. Böylece organizasyonda kullanılan baėlı veri uygulamaları, İ Verideki veri kaynaėı üstmodellerini ve organizasyonel tercihleri kullanarak kalite deėerlendirmesi yaparak Dıř Veri'yi sorgular.

Baėlı veri kapsamında Dıř Veri'nin baėlı veri uzayı servisleri tarafından bulunabilmesi iin veri kaynaklarının üstverisinin tanımlanması gereklidir. Baėlı verisetlerinin keřfi ve kullanımına imkan veren bir sözlük olan VoID (Vocabulary of Interlinked Datasets) (Alexander et al., 2009), veri vebinde bulunan verisetlerinin üstverisini tanımlar ve řekil 1.1'deki veri kaynaėı üstmodeli örneklerini oluřturmak iin kullanılır. VoID, üstveri standartlarını tanımlamak iin daėıtık sorgu motoru geliřtirmede yaygın olarak kullanılmaktadır. WoDQA (Akar et al., 2012) ve SPLENDID (Görlitz and Staab, 2011b) gibi VoID dokümanlarını kullanan daėıtık sorgu motorları, VoID yardımı ile sorgulama yaparak ilgili verisetlerini keřfederler. Veri uzayı bu verisetlerinin seilimine dair servisi organizasyona saėlamaktan sorumludur. Veri uzayında organizasyonu ilgilendiren verisetlerinin üstverisi yerel olarak tutulabilirken bu üstverinin deėiřmesi (varolan verisetlerinin deėiřmesi veya ilgili alana dair yeni verisetlerinin global veri kaynaėı üstmodel ambarına eklenmesi) de ilgili verisetlerine dair yapılabilecek operasyonları etkilemektedir. Dolayısıyla veri uzayı, sahip olduėu organizasyonel veri kaynaėı üstmodel ambarını global ambardaki deėiřiklikleri de takip ederek güncel tutabilmelidir (Erdur et al., 2013). Bu durum organizasyonun ihtiya duyduėu veriye beklentileri karřılayacak řekilde

erişmesine olanak sağlar.

Veri tüketicisi sorgu yaparak veri elde etmeye çalıştığında talep edilen veri birden fazla kaynaktan bulunuyor olabilir. Örneğin, bir gazetede haberi çıkacak bir kişinin fotoğrafı aranıp habere eklenmek istendiğinde erişilebilecek en güncel fotoğraf elde edilip yayımlanmak istenir. Veri tüketicisi (veya organizasyon) de aynı şekilde haber içeriği ve habere dair medyanın (fotoğraf gibi) güncel ve güvenilir olmasını talep eder. Bu durumda veri kaynaklarının organizasyonun tercihlerine göre filtrelenebilir. Tez çalışmasında filtreleme yapıldıktan sonra kalan verisetleri, dağıtık sorgu motorları için alt-sorguların yapılacağı seçilmiş veri kaynakları olarak kullanılmaktadır. Önerilen köken farkında filtreleme servisi için İç Veri kapsamında organizasyon verisi, organizasyonel tercihler ve yerel VoID üstverisi göz önüne alınmaktadır. Dış Veri ise sorgulanabilecek tüm bağlı veri kaynakları (global VoID üstverileri) ve sorgulandıktan sonra bağlı veri bulutundan elde edilen veri örneklerini kapsamaktadır. Köken, Dış Veri'ye üstveri eklenerek genişletilebilecek bir alandır. Literatürde benzeri şekilde VoID kullanan genişletme çalışmaları bulunmaktadır (Böhm et al., 2011; Omitola et al., 2011b; Mountantonakis et al., 2014) fakat bu çalışmalar köken kapsamını detaylı olarak ifade etmek için yeterli değildir. Tez çalışması, Dış Veri'ye eklenen köken üstverisi ile İç Veri'deki organizasyonel tercihlerin anlamsal olarak ifade edilmesi üzerinde durmaktadır. Böylece veri uzayı içerisine eklenebilen köken farkında bir çözüm önerisi sunulmaktadır.

Sorgu dağıtımı için bağlı veri uzayında hangi verisetlerinin sorgulanacağını bulmak amacı ile SPARQL sorgusu, dağıtık sorgu motoru (Akar et al., 2012) tarafından analiz edilir. Tez çalışmasına verilen isim olan PROV4FQ, dağıtık sorgular için köken farkında bir veri kaynağı filtreleme altyapısıdır. İç Veride bulunan organizasyonel veri kaynağı üstmodel ambarındaki VoID dokümanları, PROV4FQ'da filtrelendikten sonra seçilmiş veri kaynakları filtrelenmiş organizasyonel VoID ambarına aktarılır. PROV4FQ, VoID dokümanlarını incelerken izleme servisinin global veri kaynağındaki değişimlerden haberdar olmasından dolayı sürekli güncel veri kaynağı bilgileri üzerinde çalışır. Veri kaynaklarının filtrelenebilirliği için İç Veri'deki tüm elemanlar ele alınarak köken hakkındaki kurallara uygun olmayan veri kaynakları PROV4FQ tarafından elenir. Daha sonra ihtiyaç duyulduğunda, filtrelenmiş veri kaynağı listesi filtrelenmiş organizasyonel VoID ambarından alınarak dağıtık sorgu motorunun sorgu en iyileştirme aşamasına geçirilir. Böylece köken farkında dağıtık sorgulama, dağıtık sorgu motoru üzerinde çalışmaya devam eder.

Köken verisinin sorgu servisi kapsamında kullanılması için İç Veri ve Dış

Veri'ye köken verisinin entegrasyonu net biçimde tanımlanmalıdır. Dış Veri'de bulunan global VoID dokümanları, veri kaynaklarının üstverisini içermektedir. Bahsedilen VoID dokümanları dağıtık olarak bağlamlarına veya yayımlanma stratejilerine göre farklı sunucularda tutulabilir. Verisetlerinin üstverisini içeren bu dokümanlar sadece veri keşfine dair bilgi sunarken verinin kalitesini değerlendirmek için gereken üstveriyi yeteri miktarda içermemektedir. Tez çalışmasında veri kalitesine ait kavramlardan biri olarak ele alınan köken, tarihsel süreçte veriyi yaratan/güncelleyen aktörler ve verinin uğradığı değişiklikleri ifade eder. Köken bilgisinin VoID standardına eklenmesinin kökene dair değerlendirmeye yardımcı olacağı öngörülmektedir. Dolayısıyla tez kapsamında VoID, köken bilgisini de tutacak şekilde genişletilmiştir.

Kalite boyutu, veri tüketicisini ilgilendiren veriseti niteliğini ifade eder (Debattista et al., 2014; Zaveri et al., 2015). Kalite boyutu ilgili çalışmalarda genel olarak tüm kalite kavramlarını ifade etmiştir. Özel olarak Groth et al. (Groth et al., 2012) tarafından bahsedilen durum çalışmaları ile köken için gözönüne alınması gereken noktalar detaylandırılmıştır. Tez çalışmasında ise kökene ait bu noktalar, verisetlerine ait zenginleştirilmiş VoID üstverisi yardımı ile ele alınmıştır. Zenginleştirilmiş VoID üstverisi, veri kaynaklarının sahip olduğu VoID üstverisine köken gibi farklı kalite değerlendirme boyutlarına ait eklemeler yapılması ile elde edilir. Tez çalışmasında bu boyutlardan sadece kökene odaklanılmış olmakla beraber aslında tüm boyutların ve bu boyutlara dair metriklerin (Zaveri et al., 2015) gözönüne alınarak çalıştırılabileceği bir sorgulama altyapısı oluşturulmuştur. Zenginleştirilmiş VoID üstverisi, farklı boyutların kendi bilgisini ifade edebileceği anlamsal bir model üzerine inşa edilmelidir. Bu yüzden VoID içerisinde boyutlar, adlı çizge veri modeli (Carroll et al., 2005) yardımı ile tanımlanmıştır. Bu amaçla, kökenin VoID ile bağlantılarının sorgulanabilir biçimde ifade edilebileceği bir ontoloji üretilmiştir. Bu ontoloji ile VoID'in sahip olduğu köken boyutlarına dair bilgiler eklenmekte veya varolan sözlükler (Vandenbussche, 2014) ile bütünlük içerisinde ilgili boyuta dair bilgi sağlanmaktadır.

İç Veri olarak organizasyonun ifade ettiği tercihler ve Dış Veri kaynaklarına dair üstverinin, sorgulamada kullanılmak amacı ile tutulması gerekir. Organizasyonel tercihler, benzeri verisetlerinden ihtiyaca en uygun olanını seçmek için kullanılır. Organizasyonel tercihler de adlı çizge veri modeli ile ifade edilmektedir. Böylece tercih verisi ile VoID köken üstverisinin eşleştirilebilmesi sağlanmaktadır.

Özet olarak köken farkında veriseti filtreleme, veri kaynaklarının üstverisi ile organizasyon verisi ve organizasyonel tercihlere ihtiyaç duyar. VoID

zenginleştirilerek üretilmiş köken üstverisi ile organizasyon verisi ve organizasyonel tercihlerin eşleştirilmesi, köken farkında dağıtık sorgu motorlarının veri kaynaklarını eleyerek daha güvenilir ve kaliteli veriye ulaşılmasını sağlar.

2 ÖNCEKİ ÇALIŞMALAR

2.1 Bağlı Veri

Veb, sadece insan-makine arasındaki iletişimi hedeflemenin ötesinde makinelerin de iletişime katıldığı ve yardımcı olduğu global bir veri uzayı olarak tasarlanmıştır (Berners-Lee, 1998). Bu vizyonun gerçekleştirilmesinde karşılaşılan temel engellerden biri vebdeki verinin çoğunun sadece insanların tüketimine uygun olmasıdır. Örneğin veritabanından çekilerek vebde kullanıcıya sunulan verinin yapısı bir makinenin bu veriyi anlamlandırması için yeterli değildir. Anlamsal veb yaklaşımı, veriyi göstermek için makinenin işleyeceği biçimde diller geliştirilmesini öngörmüştür. Bağlı veri vebi ise bu diller ile anlamsallaştırılan veriyi (öncesinde sadece insan veya insan/makine tarafından anlaşılır veri) birbiri ile bağlayarak bir bağlı veri bulutu oluşmasını sağlar. Böylece makine tarafından sorgulanabilir bir anlamsal veri çizgesi ortaya çıkacaktır.

URI⁸ ve HTTP⁹, veri vebi için kritik teknoloji olan RDF ile bağlantılıdır. RDF, kavramlara dair URI referansları oluşturmak amacı ile bağlı veri vebinde kullanılan veri modelidir. HTML, vebdeki dokümanlar arasında geçişi sağlamak amacı ile köprüler sunar. RDF ise, çizge-tabanlı bir veri modeli sağlayarak tüm varlıkları birbirine anlamsal olarak bağlayan yapısal bir modeldir. RDF modeli ile veri, özne, yüklem ve nesne üçlülere biçiminde gösterilir. Özne ve yüklem, bir kaynağı tanımlayan URI veya literal (veya karakter katarı) olabilir. Yüklem, özne ve nesnenin nasıl ilişkilendirildiğini belirtir ve yine bir URI tarafından gösterilir.

Bağlı veri vebini oluşturmak amacı ile bağlı veri ilkeleri olarak Tim Berners-Lee tarafından ifade edilen dört adım şu şekilde sıralanmaktadır (Berners-Lee, 2006):

- Varlıkların isimleri olarak URI kullan.
- İnsanların o isimleri arayabilmeleri için HTTP URI kullan.
- Birisi bir URI aradığında, standartları (RDF, SPARQL) kullanarak yararlı bilgi sağla.
- Başka URI'lere bağlar ekle ki daha fazla veri keşfedebilsinler.

⁸Açılımı: Uniform Resource Identifier

⁹Açılımı: HyperText Transfer Protocol

Bağlı veri ilkeleri, veb altyapısı kullanarak verinin yayımlanması ve bağlanması için gerekli maddeleri tanımlarken veb dair mimari ve yapılardan yararlanmaktadır. Bu ilkelerin yanı sıra 2007 yılında ortaya çıkan Linking Open Data Project (Heath, 2007), bağlı veri vebi yaratmanın öncü çalışması olarak görülebilir. Bu proje, ontolojiler ve çıkarsama üzerinde durmak yerine anlamsal yapıları URI ve paylaşılan sözlükler yardımı ile yayımlamayı hedeflemiştir. Anlamsal olmayan Wikipedia verisi gibi veri içeren kümelerin Dbpedia gibi anlamsal şekilde yayımlanması örnek olarak verilebilir. Bu doğrultuda gelinecek nokta ise anlamsal olmayan tüm verisetlerinin yayımlanması ve birbirleri ile bağlanması, böylece veri vebinin tümü ile anlamsallaştırılmasıdır.

İnsanlar günlük hayatlarında farklı organizasyonlar ile ilişkiler içerisinde bulunur. Her organizasyon kendi ile ilgili farklı alanlarda kişisel verileri tutar. Fakat her organizasyonun kişiye ait tüm verileri tutması gibi bir durum söz konusu değildir. Organizasyon tüm veriye ihtiyaç duymaz veya ihtiyacı olduğunda bile çok fazla kişisel bilgi tutulduğundan ölçeklenebilir olmaz. Farklı organizasyonlarda parça parça bulunan bilginin vebde adreslenebilir olması, verilerin bütünleştirilmesi ihtiyacı doğduğunda önem teşkil etmektedir. Böylece farklı veri formatlarındaki veriyi işlemek veya farklı sistemlerin sağladığı uygulamalara erişmek yerine tek bir veri tipi üzerinde çalışılması işleri kolaylaştıracaktır. Örnekleme açısından Türkiye’de araç alım satımında kullanılan Araç ve Sürücü Bilgi Sistemi (ASBİS)’ni ele alalım. Bu sistem ile aracı alacak ve satacak kişiler notere gittiğinde noterin aracın satışa müsaitliği tespit edilir. Tespit edilme esnasında araç sürücüsünün trafik cezası, aracın trafik muayenesi, motorlu taşıtlar vergisinin düzenli ödenmiş olması gibi bilgileri kontrol edilir. Bu bilgiler devletin farklı kuruluşlarının sağladığı birçok bağımsız sisteme ait veb servisleri ile takip edilmektedir. Bu sistemlerin herhangi birisinin çalışmaması noterin işini yapmasını engellemektedir. Halbuki servislerden bağımsız olarak sistemin bağlı veri yardımı ile ifade edilmiş veri yapılarından meydana getirildiğini düşünelim. Bu durumda alıcı ve satıcının kimlik numarasının dağıtık sorgulama ile farklı veri ambarlarına gönderilip sonucun toplanması işi çok kolaylaştırır. Bu durumda bağımsız sistemlerin çalışması ile ilgilenilmez. Sadece sorgudan gelen verinin bütünleştirilmesi ile noterin ihtiyaç duyduğu veri kümesine erişilebilir. Bahsedilen senaryo, veri vebinde bilginin ortak bir veri tipi yani RDF dili ile temsil edilmesi ve ilgili düğümlerin birbirleri ile bağlanması sonucu gerçekleştirilebilir.

RDF bağları, RDF üçlülere biçiminde oluşturulur (Bizer et al., 2007). Şekil 2.1’deki üçlü ele alındığında RDF bağı, Dbpedia (Bizer and Auer, 2014) verisetinde bulunan Sean Penn’in LinkedMDB (Consens and Teisanu, 2014) verisetindeki ilgili yönetmen ile aynı kişi olduğunu göstermektedir. Özne referanstan ayrıldığında

Özne: <http://dbpedia.org/resource/Sean_Penn>
 Yükleme: <<http://www.w3.org/2002/07/owl#sameAs>>
 Nesne: <<http://data.linkedmdb.org/resource/director/8696>>

Şekil 2.1: Örnek RDF bağı

Dbpedia'da bulunan Sean Penn ile ilgili detaylı tüm verilere erişilirken aynı durum nesne için de LinkedMDB veriseti için geçerlidir. Bu durumda iki farklı uygulamadan veriye erişilip eşleme yapmak yerine verinin RDF bağları ile oluşturulduğu bir çizge yapısı kullanılmakta ve bu modele HTTP üzerinden erişilmektedir.

Ayrıca bağıli veri vebinin çizge halindeki yapısını ifade edebilmek için RDF ile üçlü dışında dörütlü (adlı çizge) şeklinde gösterimler de kullanılmaktadır. RDF verisi hakkında üstveri gösterimine ihtiyaç duyulan uygulamalar için üçlü modelinin yeterli olmadığı ve dolayısıyla RDF veri modelinde değişikliklerin yapılabileceğı ifade edilerek adlı çizgeler ortaya çıkmıştır (Bizer, 2007). Adlı çizge, en basit tanımı ile üçlülerden oluşan çizgeleri isimlendiren anlamsal modellerdir. Ayrıca RDF veri modeli, ihtiyaçlara göre üstveri ile zenginleştirilmiş elemanlar oluşturarak farklı sistemler tarafından etkili şekilde kullanılabilir.

RDFS ve OWL, varlıkları tanımlamak ve aralarındaki ilişkiyi ifade etmek için sözlükler yaratmaya temel oluşturur. Sözlükler, sınıflar ve özelliklerin kümesidir. Sözlüklerin kendisi RDF ile oluşturulurken RDFS ve OWL'dan gelen terimleri kullanır. Böylece ilgi alanlarını modellerken daha kapsamlı bir gösterim yapılabilmesini sağlar. Bu doğrultuda sözlükler arasındaki bağlantılar kullanılarak herkes, veri vebinde verisini yayımlayıp diğer verisetleri ile bağlama yeteneğine sahip olmaktadır.

2.1.1 Void

RDF veri kaynaklarındaki verisetlerinin belirtimi için üstveriye ihtiyaç duyulmaktadır. VoID (Alexander et al., 2009), bağıli verisetlerini tanımlayarak bu üstveriyi sağlayan RDF tabanlı bir şemadır. VoID dokümanları, verisetlerinin etkili şekilde keşfi ve kullanımı, araştırma verisinin gösterimi, etkili veriseti seçimi ve sorgu en iyilemesi amacıyla oluşturulur (Alexander et al., 2009). Bunun yanında VoID, ihtiyaçlar doğrultusunda verisetlerinin keşfine olanak sağlamak amacı ile genişletilebilir olarak sunulmuş ve buna uygun VoID dokümanı geliştirme araçları da üretilmiştir (Böhm et al., 2011). Böylece veri tüketicisine sağlanacak veriye ek veri dahil edilerek yayımlama sağlanır (Omitola et al., 2011b; Böhm et al., 2011;

Gray, 2014; Mountantonakis et al., 2014). Bu yaklaşım, zenginleştirilmiş VoID dokümanları elde etmede kullanılabilecek bir yöntem olarak tez çalışmasındaki yaklaşımda kullanılmıştır.

Verisetlerinin keşfinde ne tip verinin bulunduğu, veri örnekleri, verinin güncelliği, kimin yayımladığı, SPARQL uçnoktasının nerede bulunduğu, verinin indirilme bilgisi, ne kadar büyüklükte bir veri olduğu ve lisanslaması gibi unsurlar (Cyganiak, 2010) VoID dokümanlarının oluşturulması sırasında gözönüne alınmaktadır. Tez kapsamında VoID dokümanları, veriseti köken bilgisini tutacak şekilde genişletilmiştir. Bu şekilde verisini yayımlayan bir organizasyonun aynı zamanda kökeni içeren bir üstveriyi de yayımlayabildiği kabul edilmiştir.

2.1.2 Adlı çizge

RDF verisi hakkında üstveri gösterimine ihtiyaç duyulan uygulamalar için üçlü modelinin yeterli olmadığı ve dolayısıyla RDF veri modelinde değişikliklerin yapılabileceği ifade edilmiştir (Bizer, 2007). Bu bakış açısı ile ortaya çıkan adlı çizgeler, en basit tanımıyla üçlülerden oluşan çizgeleri isimlendiren anlamsal modeller olarak tanımlanmaktadır. Adlı çizge veri modeli;

- Üstverinin gösterimi için kullanılır. Varolan RDF modeli ile birlikte üstverinin daha etkili şekilde ifade edilmesine olanak sunar.

- RDF verisinin global olarak aynı şekilde tanımlanması için bir mekanizma sağlamaktadır. Dolayısıyla farklı veri sağlayıcıları aynı RDF verisi üzerinde üstveriyi ifade edebilir.

- Varolan RDF verisi ile geri uyumluluk göstermektedir. Böylelikle mevcut uygulamalar ve yeni oluşturulacak RDF önerileri ile uygun şekilde çalışabilmektedir.

- Üstveri değişimine olanak vermektedir. Adlı çizge veri modeli, verinin üstverisi ile birlikte yayımlanması ve farklı sistem veya organizasyonlar arasında değişimine uygun bir sözdizimi sağlamaktadır.

Adlı çizgeler kullanılırken, RDF üçlülerine erişime olanak sağlayacak mekanizmalarda RDF çizgelerinin kullanılması esastır. Bir adlı çizge, URI referansı şeklinde çizge adı ve RDF çizgesinden meydana gelmektedir.

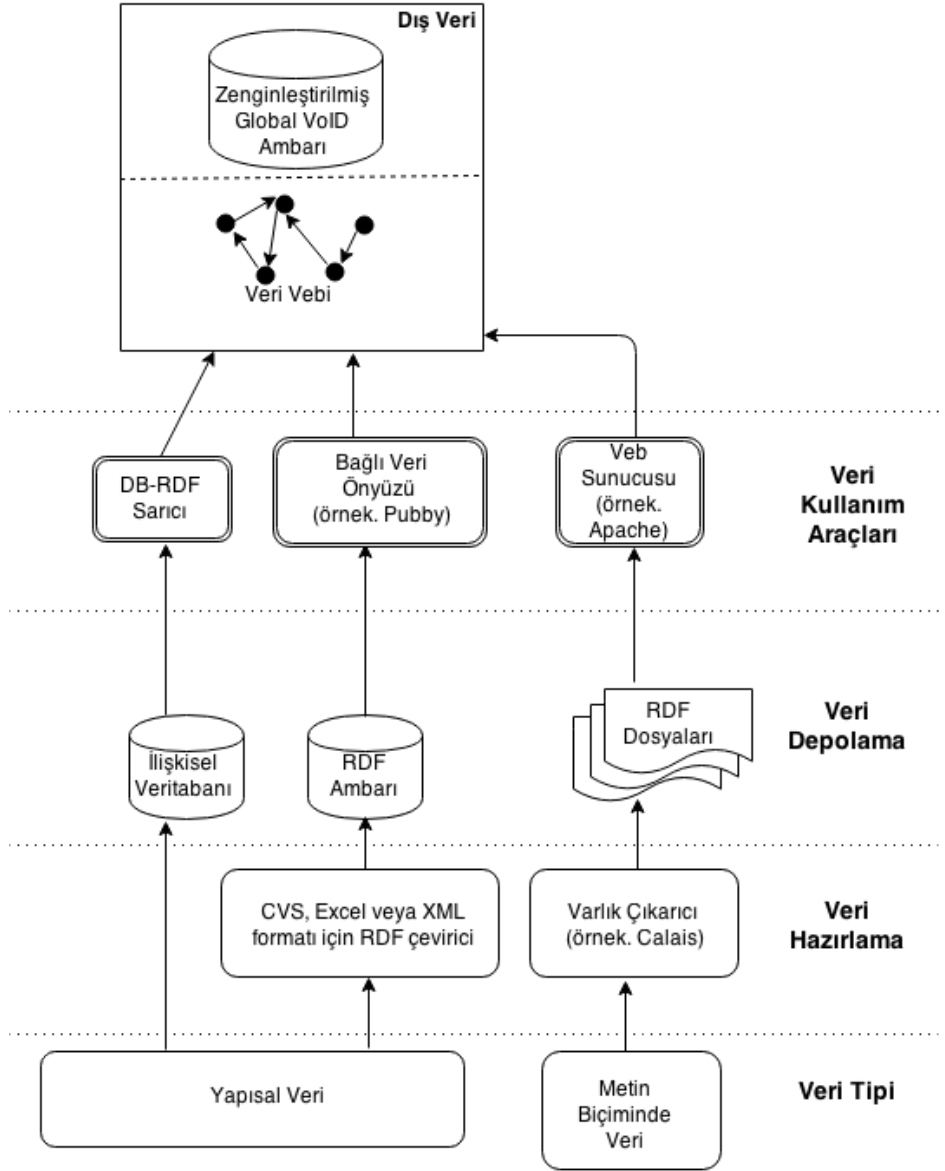
2.1.3 Bağlı veri yayımlama

Bağlı veri, Tim Berners-Lee (Berners-Lee, 2006)'nin ifade ettiği ilkeler ile yayımlanır. Bahsedilen ilkeler, varolan veri yönetim sistemleri ve bu sistemleri kullanan iş uygulamalarından tamamen vazgeçmek yerine teknik bir katman yardımı ile Veri Vebi'ne bu sistemleri bağlamayı öngörmektedir (Heath and Bizer, 2011). Bu ilkelerin yanı sıra Veri Vebi'ne varolan veri kaynaklarını bağlamaya yardımcı olmak için bağlı veri yayımlama desenlerinden de yararlanılmalıdır. Bağlı veriye dönüştürülecek verinin yapısına bağlı olarak üç temel desen tanımlanmıştır. Desenler; Şekil 2.2'deki gibi sorgulanabilir yapısal veriye bağlı veriye dönüştürmek (ilişkisel veritabanlarında bulunan veriye RDF sarıcıları yardımı ile RDF çizgelerine dönüştürmek), statik yapısal veriye (CSV, Excel veya XML dokümanları) bağlı veriye dönüştürmek ve metin dokümanlarını bağlı veriye dönüştürmek şeklinde ifade edilir (Heath and Bizer, 2011).

2.1.4 Bağlı veri tüketme

Bağlı veri, RDF gibi standart bir veri modeli üzerinde inşa edildiğinden dolayı farklı kaynaklardan elde edilen bağlı verinin tümleştirilebilmektedir. RDFS, OWL gibi üstmodellere bağlı olarak tanımlı RDF, SPARQL gibi erişim için ortak olarak kullanılan bir dil ile sorgulanmaktadır. Böylece Şekil 2.2'deki gibi asıl kaynağı farklı formatlarda olsa da veri tüketimi için yayımlanan tüm veriye ortak bir veri ve sorgu modeli ile erişilmektedir.

Bağlı veri uygulamaları iki kategoride sınıflandırılabilir: genel uygulamalar ve çalışma alanına dair uygulamalar (Heath and Bizer, 2011). Genel uygulamalara örnek olarak bağlı veri tarayıcıları (Berners-Lee et al., 2006; Bizer and Gauss, 2007) ve bağlı veri arama motorları (Cheng and Qu, 2009; Tummarello et al., 2010; Hogan et al., 2011) verilebilir. Farklı çalışma alanlarındaki ihtiyaçlara dair uygulamalar ise günümüzde geliştirilmeye devam etmektedir. Örneğin, DbpediaMobile (Becker and Bizer, 2009), turistlere bir şehirde gezerken yardım etme amacı ile geliştirilmiştir. Mobil cihazlar üzerinde çalışan uygulama, yayımlanan veriye Dbpedia ile bağlayarak telefondaki anlık GPS pozisyonuna göre kullanıcıların veri kaynakları arasında gezebilmesini sağlar. Farklı organizasyonların yayımladığı veriye örnek olarak verilebilecek olan NCBO Resource Index (Jonquet et al., 2011), yüzlerce biyomedikal kaynağı keşfederek kullanıcılara istediği veriyi sunar. Böylece tüm kullanıcılar Veri Vebi'nde yayımlanmış veriden yararlanır.



Şekil 2.2: Tez çalışmasına adapte edilmiş veri yayımlama akışı

Şekil 2.3’de çalışma alanına dair uygulamalar, organizasyonel perspektife uygun olarak ifade edilmiştir ve Şekil 1.1’e dayanarak detaylandırılmıştır. Buna göre organizasyonun ihtiyaçları dahilinde kullanmak istediği bağlı veri uygulamaları bağlı veri servislerinden yararlanmaktadır. Bağlı veri servisi, bağlı veri yaşam döngüsü (Auer et al., 2013) içerisindeki her aşamada yer alabilecek bağlı veri uzayı servisini ifade eder. Bağlı veri servisleri, sorgulama (Akar et al., 2012), değişim izleme (Erdur et al., 2013) veya kalite değerlendirmesi (köken takibi) (Zaveri et al., 2015) gibi bağlı veri ile ilgili farklı ihtiyaçları beraberce çalışarak karşılar. Bu servisler, organizasyona özgü veriyi ifade eden İç Veri (köken bazlı veriseti filtreleme) veya Veri Vebi’ni ifade eden Dış Veri’den (veri izleme) yararlanabilir. Uygulama katmanındaki uygulamalar ise bağlı veri uygulamaları ve yönetimsel

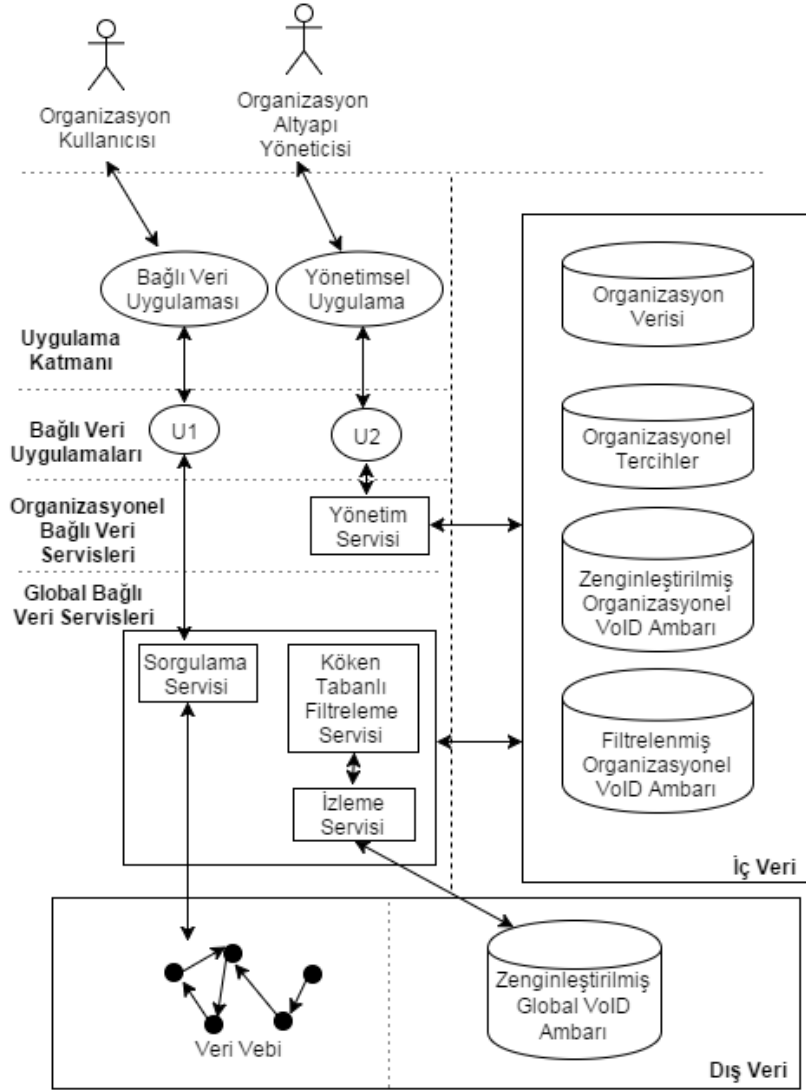
uygulamalar olmak üzere ikiye ayrılmaktadır. Bağlı veri uygulamaları, Veri Vebi'ne erişebilmek için farklı servisleri kullanır. Yönetimsel uygulamalar ise, organizasyon verisi, tercihler ve yayımlanacak organizasyona ait verisetlerine ilişkin operasyonlar ile ilgilenir. Bu tip uygulamalar Veri Vebi'ni kullanmak için değil yardımcı olmak için organizasyon tarafından geliştirilip kullanılan uygulamalardır ve İç Veri'de güncellemelere sebep olur.

Heath ve Bizer (Heath and Bizer, 2011) tarafından sunulan veri tüketim yaklaşımına göre Veri Vebi'ne erişim, veri entegrasyonu ve veri depolama mimarisi, Veri Vebi'ne erişim, sözlük eşleme, kimlik çözümleme, köken takibi, veri kalite değerlendirmesi ve veriyi uygulama bağlamına göre kullanma modüllerinden oluşmaktadır. Bu modüllerden bazıları Şekil 2.3'de ifade edilmekle beraber tez çalışmasının ana fikrini köken takibi oluşturmaktadır. Aslında köken takibi bir çeşit veri kalite değerlendirme boyutu (Zaveri et al., 2015) olmakla beraber Veri Vebi'ne erişim için önem arz ettiğinden Heath ve Bizer (Heath and Bizer, 2011) tarafından farklı bir modül olarak ele alınması ve incelenmesi gerektiği düşünülmüştür. Dolayısıyla tez çalışmasında da bağlı veri altyapısına köken takibinin nasıl entegre edilmesi gerektiği detaylı şekilde gösterilmeye çalışılmıştır.

Şekil 2.3 üzerinden açıklanacak olursa sorgulama servisi, sorgulama için Veri Vebi'ndeki örneklerden yararlanır. Köken farkında filtreleme servisi, Zenginleştirilmiş Organizasyonel VoID Ambarı, organizasyon verisi ve tercihleri kullanarak köken bazlı veriseti filtreleme yapabilir. İzleme servisi ise Zenginleştirilmiş Global VoID Ambarı üzerinde değişim takibi yaparak değişen VoID dokümanlarını Zenginleştirilmiş Organizasyonel VoID Ambarı'na bildirmekten sorumludur. Sorgulama servisi, Veri Vebi'ndeki veriyi sorgulamak için Filtrelenmiş Organizasyonel VoID Ambarı'nı kullanır. Organizasyonun sürekli izlenerek güncellenen ve köken kontrolü yapılan üstverisi, sorgulama servisinin hangi veri kaynaklarını sorgulayacağını belirler. Bunun yanında veri uzayındaki farklı birçok servis de veri uygulamaları için ortaklaşa kullanılarak organizasyonların taleplerini karşılar. Bahsedilen servislerin detayları Bölüm 3'de anlatılmıştır.

Organizasyonda bulunan aktörler, organizasyonel kullanıcı ve organizasyonel altyapı yöneticisidir. Organizasyonel altyapı yöneticisi, bağlı veri uygulamalarında kullanılacak veriyi sağlama amacı ile yönetimsel servisleri çalıştırır. Yönetimsel uygulamalar her organizasyonun kendi içerisindeki servisleri kullanır ve bu servisler organizasyon verisi ve tercihleri üretir veya o organizasyona dair yayımlanabilecek verisetlerini sunar. Organizasyonel kullanıcılar da organizasyonda kullanılan uygulamalar yardımı ile Veri Vebi'nde bulunan veriye erişerek veri tüketimini

gerçekleştirir.



Şekil 2.3: Tez çalışmasına adapte edilmiş veri tüketim akışı

Veri tüketimi için ihtiyaç duyulan temel servis sorgulama servsidir. Bununla beraber bağlı veri uzayında gerekli tüm servisler ve ihtiyaç duyulan üstveri zamanla eklenerek Veri Vebi'ndeki kaliteli veriye erişim sağlanmalıdır (Madhavan et al., 2007). Ayrıca, paylaşılan sözlükler ve veri bağlantılarının artması ile veri kaynakları arasındaki ilişki üstveri anlamında da netleşerek organizasyonların daha az eforla daha fazla ve uygun veriyi elde edebilmesi sağlanabilir.

2.2 Köken

Vebde bulunan büyük miktardaki verinin önemli bir kısmı, kopyalama, değiştirme veya birleştirme ile meydana getirilmiştir. Aynı bağlı veriseti

kopyalanabilir veya aynı/değişik ad uzayları ile farklı noktalarda sunulabilir. Verisetleri, farklı araçlar kullanılıp yaratılan RDF bağları ile birbirine bağlanabilir. Böylece aynı varlık kümelerinin çelişkili kopyaları birbirine bağlı olabilir. Bu şekilde üretilen verinin kalitesi üzerinde düşünülmesi gerekmektedir. Çünkü bu tip veri hızla webde yayılıp Veri Vebi'nin kalitesiz veriden oluşmasına neden olabilir.

Bir varlık bulunmaya çalışıldığında bu varlığı belirten birçok URI ve/veya URI'ye bağlanmış farklı kaynaklardan gelen veri dönebilir. Bu noktada cevaplanması gereken soru, hangi bağlantılar izlenerek istenilen veriye ulaşılabilecektir. Kullanılması gereken veri kaynağı, daha güvenilir veya güncel verinin sağlandığı kaynak olmalıdır. Veri kaynağına erişirken sadece varlık hakkındaki veri yeterli değildir. Veri kaynağının keşfi için VOID gibi bir sözlük ile ifade edilen üstveri de veri kaynağının nasıl/ne zaman üretildiğine dair soruları cevaplayamamaktadır. Bu yüzden veri veya keşif üstverisi yeterli olmamakta, ayrıca köken takibi için gerekli üstveriye de ihtiyaç duyulmaktadır. Bu doğrultuda veri ve kökene ait üstverinin, Veri Vebi'nde hangi süreçlerden geçerek kullanılması gerektiği üzerinde düşünülmelidir.

Hartig ve Zhao (Hartig and Zhao, 2010), köken üstverisini tüketmek amacıyla elde etmek için iki yaklaşım tanımlamıştır. İlk yaklaşımda verinin köken farkında değerlendirmesini gerçekleştiren uygulamalar, üstveriyi kaydeder. Veri sorgulama veya güncelleme gibi yapılan her operasyon uygulama tarafında kaydedilerek uzun vadede kullanılmak için saklanır. İkinci yaklaşımda ise veri sağlayıcısı tarafından veriseti ile beraber köken üstverisi yayımlanır. Veri kökeni, veriseti veya veri örneği seviyelerinde ele alınırsa birinci yaklaşım için veri örneği¹⁰, ikinci yaklaşım için ise veriseti seviyelerinin uygunluğu görülmektedir. Bununla beraber her veri örneğine dair yapılan operasyonun kaydının tutulup sorgulanması, performans açısından sorgulanma süresini olumsuz etkileyebilir. Bu yüzden tez çalışmasında, ikinci yaklaşım ele alınarak veriseti seviyesinde kökenin sorgulanması üzerinde çalışılmıştır.

Bölüm 2.1.3'de ifade edilen bağlı veri yayımlama, köken üstverisi için de Hartig ve Zhao (Hartig and Zhao, 2010) tarafından detaylandırılmıştır. Bu yaklaşıma göre dört farklı yöntemle köken verisi yayımlanabilir ve sorgulanabilir. İlk yöntemde, bir varlığın sahip olduğu HTTP URI için referanstan ayırma ile elde edilecek RDF çizgesinin içerdiği veriye ek olarak köken üstverisi tanımlanmaktadır. Köken üstverisi tanımı, verinin kaynağı ve verinin yaratılma süreci hakkında bilgi sahibi olunabilecek köken bilgisini sağlamalıdır. Ayrıca, varlığın dahil olduğu bağlı veriseti bilgisini de içermelidir. İkinci yöntemde göre, bağlı veriseti bir RDF yığını

¹⁰İngilizcesi: Instance

olarak bulunmaktadır. RDF yığını, tüm RDF verisetini içeren bir RDF dokümanıdır. Bu doküman ile beraber dokümana ait köken üstverisi sağlanabilir. Sağlanan üstveri tüm verisetinin kökenini tanımladığından bu durumda verisetinin VoID tanımlaması ile beraber sunulması önerilmektedir. Bu yaklaşımda tez çalışmasında kullanıldığı üzere bir bağlı verisetine dair üstveri, adlı çizgeler şeklinde yayımlanabilmektedir. Böylece her verisetine ait üstveri adlı çizgelerde tanımlanarak veri tüketicisinin yapacağı sorgu ile ilişkili köken takibi sağlanmaktadır. Üçüncü yöntemde, bir bağlı verisetine SPARQL uçnoktası yardımı ile erişilir. SPARQL uçnoktası, SPARQL sorgularını çalıştıran bir sorgu servsidir. Bu yaklaşımda SPARQL sorgu motoru, sorgu cevabına köken üstverisini de ekleyebilir. Bu yaklaşıma güncel olarak verilecek en uygun örnek, PROV köken modelinin PROV-AQ (Moreau et al., 2012) erişim ve sorgu modelidir. Bu model yardımı ile verisetine ek olarak farklı bir SPARQL uçnoktasından ekstra üstveri elde edilebilir. Dördüncü yöntemde bağlı veri yayımlama araçları ile (D2RServer (Bizer and Cyganiak, 2006), Triplify (Auer et al., 2009), Pubby (Cyganiak and Bizer, 2011) gibi) üstveri yayımlanır. Bu araçlar otomatik olarak bağlı veri varlıkları ile beraber köken üstverisi üreterek sunar. Bu araçlarda yapılacak konfigürasyon değişiklikleri ile köken üstverisi zenginleştirilerek sunulabilir.

Veri kökeninin sorgulanmasından önce köken verisinin yaratılması, vebde yayımlanması ve yayımlanmış verinin erişimi gibi eylemlerin nasıl ifade edileceğine dair bir köken veri modeline ihtiyaç vardır. Bu amaçla ihtiyaçların analiz edilerek köken modellemesi için geliştirilmiş sözlükler (ontolojiler) bulunmaktadır (Moreau et al., 2011; Hartig and Zhao, 2012; Lebo et al., 2013). Bu sözlükler yardımı ile RDF tabanlı köken tanımlamaları Veri Vebi'nde yayımlanarak veri tüketicileri tarafından kullanılabilir.

Köken modeli ifade edilirken öncelikle kökenin genel terimlerinin netleştirilmesi gerekmektedir. Bu elemanlar tüm sözlüklerde benzer şekilde ifade edilmekle beraber en güncel çalışma olarak PROV-O ontolojisi¹¹ (Lebo et al., 2013) bulunmaktadır. PROV-O, çalışma alanından bağımsız ve genel amaçlı kavramlar ile meydana getirilmiştir. Genel amaçlı oluşturulmasının en önemli nedeni, farklı ihtiyaçları kapsayabilecek bir sözlük ortaya çıkarmak ve kökeni ele alacak çalışmaları (Baillie et al., 2013; Ciccarese et al., 2013; Markovic et al., 2013; Missier et al., 2013) bu genel amaçlı sözlüğü kullanmaya teşvik etmektir. PROV-O'da temel kavramlar, *Varlık*¹², *Etmen*¹³ ve *Aktivite*¹⁴ olarak tanımlanmıştır. Varlık, başka bir varlıktan türeyebilir, bir aktivite tarafından üretilbilir veya bir

¹¹İngilizcesi: W3C Provenance Ontology

¹²İngilizcesi: Entity

¹³İngilizcesi: Agent

¹⁴İngilizcesi: Activity

etmene atfedilebilir. Örnek olarak bir haber ajansında çalışan muhabir, haberi yaptığında haber URI'si (varlık örneği) muhabir etmenine atfedilir. Haber ajansı da etmen olarak ifade edildiğinden muhabirin haber ajansında çalıştığı ifade edilebilir. Haber oluşturma aktivitesi, muhabir ile ilişkilendirilir ve bu aktivite oluşturulan habere bağlanır.

Kökeni tanımlamak için oluşturulan genel köken modelleri dışında veriseti seviyesinde kökenin ele alınması için VoID sözlüğüne ait özelleştirilmiş yaklaşımlar da ifade edilmiştir. Literatürde bu konuda yapılan çalışmalardan VoID için köken ilavesi (VoIDp) (Omitola et al., 2011b), VoID sözlüğünün köken bilgisi ile genişletilmiş hali olarak nitelendirilebilir. Yaklaşımın amacı, veriseti yayımlayıcılarının verisetlerinin kökeni ile ilgili üstveri sağlamasına yardımcı olmaktır. Böylece veri tüketen araçlar daha güvenilir ve kaliteli veriye erişebilir. VoIDp (Omitola et al., 2011a), bağlı verisetlerine köken üstverisi ekleme amacı ile oluşturulmuştur. VoIDp (Omitola et al., 2011b), VoID içerisinde köken açısından sadece kaynak ve sonuç verisetleri hakkındaki bilginin tutulmasına olanak veren üstveriyi sunmaktadır. Fakat verisetlerinin kökenine dair sadece kaynak ve sonuç verisetlerinin gözönüne alınması eldeki birçok veriseti arasından filtreleme yapılmak istendiğinde yeterli değildir. Bu yaklaşım yerine PROV-O gibi daha genel amaçlı sınıf ve özelliklerin genişletilebildiği bir çalışmanın VoID dokümanlarını zenginleştirmede kullanılması gereği bu tezin temel yaklaşımıdır. Tez çalışmasında VoID sözlüğüne ait kavramlar köken gözönüne alınarak filtreleme yapmak için genişletilmiştir. Böylece hem veriseti keşfi için kabul görmüş VoID sözlüğü kullanılmış, hem de verisetlerini filtrelemek için kökenin gözönüne alınması sağlanmıştır.

PROV-O çalışma alanından bağımsız ve genel amaçlı bir ontoloji olarak ortaya çıkmıştır. Vebdeki kaynakların versiyonlanması, kaynakların yazım veya güncellenme aşamalarında meydana gelen değişikliklerin izlenmesi için ortaya atılan temel bir metodolojiyi ifade etmektedir. VoIDp ise VoID için özelleştirilmiş bir köken genişletme çalışması olarak sunulmuştur. VoIDp ile verinin köken ilişkilerini belirtmek için veri yayımlayıcılarının üstveri oluşturmaya olanak sağlanmıştır. Tez çalışmasındaki hedef, genel amaçlı bir köken ontolojisini genişleterek VoID üzerinde kullanılabilir bir yaklaşım geliştirmektir. PAV¹⁵ (Ciccarese et al., 2013) ontolojisi PROV-O'yu genişleterek verinin yazımı, düzenlemesi ve dijital olarak yayımlanması üzerinde durmuştur. Bu yüzden VoID üzerinde üstveri sunabilmek için uygun bir yaklaşım olarak göze çarpmaktadır (Gray, 2014). PAV ontolojisi, varolan veb tabanlı sistemlere içerik sağlayan etmenlerin farklı rolleri arasında ayırım yapmak için terimler üretmiştir (Ciccarese

¹⁵Açılımı: Provenance, Authoring and Versioning

and Soiland-Reyes, 2015). Bahsedilen roller, yazarlık, düzenleyicilik, iştirakçilik ve dijital artifakt yaratıcılığı gibi bir dokümanın metin halinde oluşturulması ve vebde yayımlanması sürecindeki aşamalar ile ilgilenmektedir. Vebde yayımlanma tarafında tezi de ilgilendiren nokta, verinin erişilme, dönüştürülme ve tüketilme esnasında dijital varlıkların kökeninin izlenebiliyor olmasıdır. Daha geniş kapsamlı ve birlikte çalışabilir bir yaklaşım sunabilmek için PAV ontolojisi PROV-O ontolojisinin genel amaçlı modelini özelleştirmiştir. Bu özelleştirilmiş yaklaşım VoID üzerindeki veriseti üstverisi için de kullanılabileceği için PAV tarafından sağlanan ontolojik ifadeler ile PROV-O ontolojisinin beraberce kullanılabileceği görülmüştür.

PAV ontolojisinde Köken, dijital veri yaratılması tarafı ile ilgilenirken, Yazarlık¹⁶ verinin gerçek anlamda yaratılması ve bununla ilgili özelliklerden bahseder. Örnek vermek gerekirse İnce Memed romanı 1955 yılında yazılmış olmasına rağmen vebde ilk olarak e-kitap olarak yayımlanması 2010 yılında olmuştur denilebilir. Bu durumda romanın yazımı Yazarlık, veb üzerinde roman ile ilgili yapılan yayımlanma durumu Köken ile ifade edilir. Versiyonlama¹⁷ ise zaman içerisinde dijital varlıkların evrimini gösterir. Bu noktada çalışmada VoID ile adapte edilebilecek ontolojik çalışmanın Köken ve Versiyonlama ile ilgili olduğu görülmüştür. Böylece VoID dokümanının yaratılmasına ve ilişkili verisetinin hangi süreçlerden geçerek yayımlandığına dair üstveri PROV-O ve PAV ontolojisi kullanılarak geliştirilmiştir.

PROV-O ve PAV ontolojisi için veri yaratılması, veri türetilmesi, başka bir kaynaktan veri alınması, versiyonlama gibi kökene dair kavramlar bulunduğundan PAV ontolojisinin geliştiricileri PROV-O ontolojisinde bulunan benzeri terimleri genişletmişlerdir. Özellikle PROV-O'da atfetme amacıyla kullanılan *prov:wasAttributedTo* özelliği, veriyi yazma (*pav:authoredBy*), yazıya katkıda bulunma (*pav:contributedBy*), yazıyı düzeltme (*pav:curatedBy*), yazıyı vebde yayımlama (*pav:createdBy*), yayımlanma aşamasına yardımcı olma (*pav:createdWith*), farklı formattaki verinin dönüştürülmesinden sonra aktarma (*pav:importedBy*) ve belirli bir kaynaktan dönüştürülme olmadan veriyi çekme (*prov:retrievedBy*) şeklinde genişletilmiştir. Bu özellikler veriseti üstverisi kullanma aşamasında ihtiyaç duyulan tüm detayları kapsadığından dolayı VoID ontolojisini genişletme noktasında tez çalışmasında PROV-O ile beraber PAV ontolojisinden yararlanılmasının uygun olduğu görülmüştür.

Köken için ihtiyaç duyulan kavramların ifade edilebilmesi amacı ile 2010 yılında W3C Köken Kuluçka Grubu tarafından üç adet ana senaryo (Groth et al.,

¹⁶İngilizcesi: Authoring

¹⁷İngilizcesi: Versioning

2012) tanımlanmıştır. Haber Toplama Senaryosu, Hastalık Salgın Senaryosu ve İş Sözleşme Senaryosu adındaki bu senaryolar arasından bu tezin durum çalışması kapsamında Haber Toplama Senaryosu seçilmiştir. Bu senaryo farklı kaynaklardan aldığı haber elemanlarını (haber siteleri, ağ güncelleri ve haberle ilgili resimler gibi) birleştirme amacını gütmektedir. Bu doğrultuda senaryo kapsamında kaynakların kökeni ile ilgili güven, doğrulama ve lisanslama benzeri kavramların dikkate alınması gerekmektedir. Bu kavramların yanısıra haberlerin güncelliğinin sorgulanması gibi sahip olunmak istenen bilginin kalitesini arttırıcı noktalar üzerinde de durulmaktadır.

Haber Toplama Senaryosu'nda üretilecek örnek veri, fotoğraf, video gibi medya veya yapılan haberin içeriğidir. Medya veya haber içeriğinin veri kökenine bağlı olarak seçilmesi için üretilen verinin lisanslanmış olması veya güvenilir bir sağlayıcı tarafından yayımlanması gerekir. Seçim için ihtiyaç duyulan ise kökeni tanımlayan üstverinin tüm veri tüketicileri tarafından erişilebilir olmasıdır. Bunun için köken üstverisinin veri ile beraber yayımlanması gerekmektedir. Böylece erişilecek benzeri veriye sahip veri kaynakları arasından en uygunu seçilebilir. Tezde veri kaynaklarının organizasyonel kavramlar olduğu üzerinde durulmuştur. Dolayısıyla organizasyonel kavramların kökeni gözönüne alınmıştır. VoID üstverisinin organizasyonel kavramlara dair veri ile beraber nasıl ele alınıp işletilebileceği düşünülmüştür.

Tezde Haber Toplama Senaryosu, Bölüm 2.1.3'deki bağlı veri yayımlama ve Bölüm 2.1.4'de anlatılan bağlı veri tüketme akışı ile çalıştırılmıştır. Veri yayımlama sürecinde, VoID üstverisi zenginleştirilirken adlı çizgelerden (Carroll et al., 2005) yararlanılmıştır. VoID dokümanlarının detayları Bölüm 3.2'de anlatılmıştır. VoID üstverisine nasıl eklemeler yapıldığı aynı bölümde ifade edilmiştir. Veri tüketme süreci, köken farkında veriseti filtreleme, bağlı veri uygulaması, sorgulama, değişim izleme ve köken farkında filtreleme servislerinden meydana gelmektedir (bkz. Şekil 2.3). İç Veri'de varolan veriseti listesi, köken farkında filtreleme servisi ile filtrelenerek elde edilir. Sorgulama ile köken dikkate alınıp elde edilen verisetleri kullanılarak Dış Veri'ye erişilir.

2.3 Veri Uzayı

Günümüze kadar veri bütünleştirme genellikle bir veya birden fazla anlamsal olmayan veri kaynağına erişerek sağlanmıştır. Bu yaklaşım, ihtiyaç dahilinde kullanılacak her veri kaynağının elle eklenerek vakit harcanmasına neden olmakta ve tek bir global veri uzayı oluşturulmasını engellemektedir. Global veri uzayı,

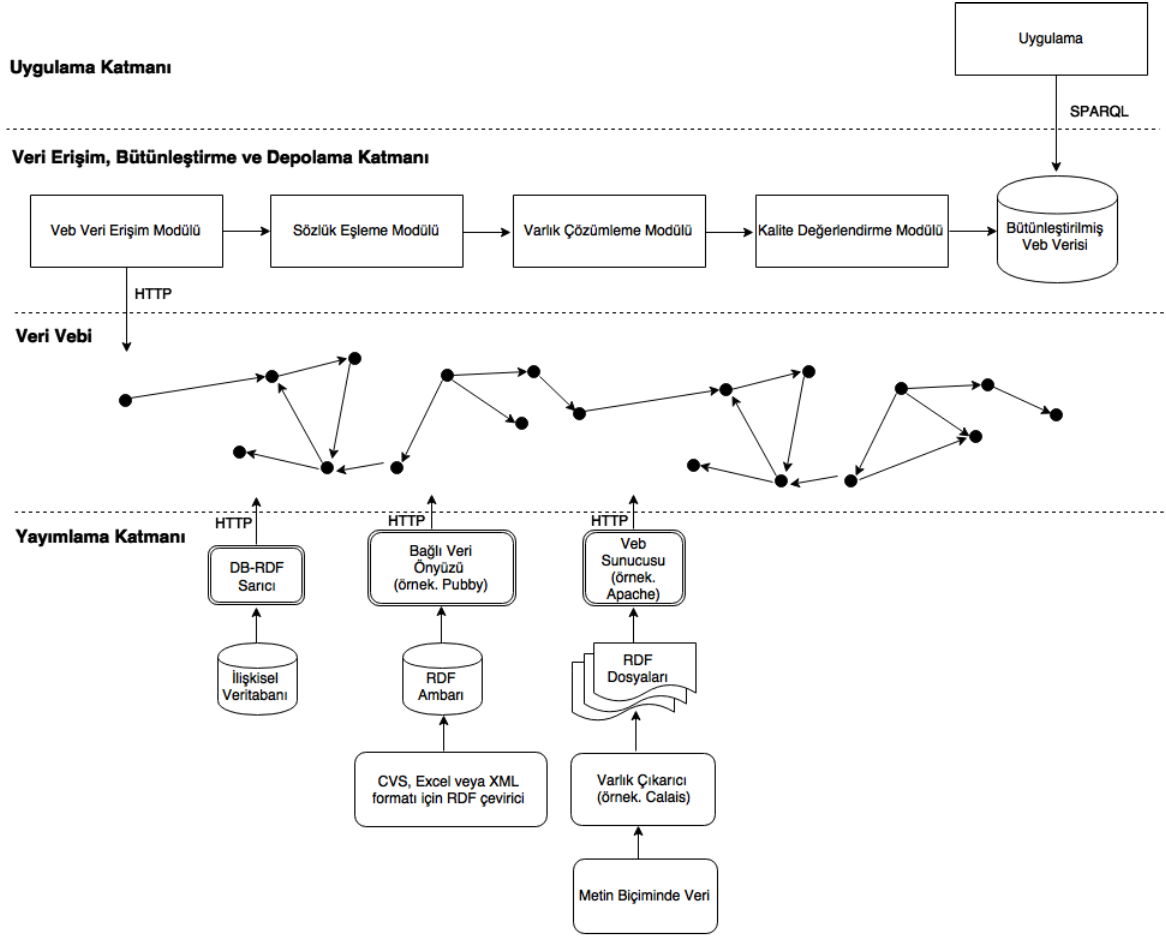
farklı veri kaynakları arasında bağlı veri yardımı ile bağlantılar kurularak meydana getirilir. Böylece tüm veri kaynaklarının birbiri ile bağlanması sağlanır. Vebi bir global veri uzayı olarak ele almak için öncelikle RDF kullanılarak yapısal verinin vebde yayımlanması gerekir. Yayımlanan verinin bağlanması ile geleneksel olarak her kaynağa erişmek için farklı arayüz kullanma yaklaşımı yerine dağıtık sorgu motorları ile tüm kaynaklara erişim sağlanmıştır.

Veri uzayının özellikleri, bir global şemaya bağlı kalınarak oluşturulmaması, ihtiyaç oluştuğunda veri entegrasyonunun sağlanması¹⁸ ve sorgulara en uygun cevapların sunulmasıdır (Franklin et al., 2005; Madhavan et al., 2007). Bağlı veri bakış açısı ile düşünüldüğünde oluşturulan global veri uzayının da temelde bu özelliklere sahip olması gerektiği görülmektedir. Tek bir şema tanımlamak yerine yayımlanan anlamsal veri arasında bağlantılar kurmak yayımlama aşamasında (bkz. Bölüm 2.1.3), ihtiyaç dahilinde entegrasyonun sağlanması ve sorgulara en uygun cevapların verilmesi de veri tüketimi aşamalarında (bkz. Bölüm 2.1.4) gözönüne alınmalıdır. Şekil 2.4'de Bizer (Bizer, 2011) tarafından oluşturulan katmanlı mimari, veri yayımlama ve tüketme bakış açılarının veri uzayındaki detaylarını göstermektedir. Bu yaklaşıma göre uygulamaya veri erişimi, sözlük eşlemesi, varlık çözümlemesi ve kalite değerlendirmesi modüllerinden geçerek bütünleştirilmiş veri döndürülür. Bu modüllerin her biri Şekil 1.1'de birer servis olarak gösterilmiş ve uygulamanın bu servislerin çalıştırılması ile gerçekleştirilebileceği ifade edilmiştir. Bütünleştirilmiş veriyi elde etmek için veri erişimi ve tüketimi aşamasında yayımlanmış verisetlerinden oluşan Veri Vebine erişilmesi gerekir. Şekil 2.4'in detaylı olarak incelendiği veri tüketimi ile ilgili elemanların tez çalışmasında nasıl düşünüldüğü Şekil 2.3'da gösterilmiştir. Özellikle İç Veri ve Dış Veri kavramlarının literatüre eklenmesi ve bu verilerin Veri Uzayı'nda hangi aşamalarda ihtiyaç duyulduğu Bizer (Bizer, 2011)'in çalışmasının üzerindeki tezin katkıları olarak görülmektedir. Kalite Değerlendirme Modülü olarak ifade edilen kısım tez çalışmasında köken farkında veriseti filtreleme şeklinde açıklanmış ve veriseti seviyesinde neler yapılması gerektiği incelenmiştir.

Kökenin bağlı veri uzayında üzerinde durulması gereken bir kavram olduğu ifade edilmekle beraber (Heath and Bizer, 2011) henüz bağlı veriyi gözönüne alarak geliştirilmiş köken farkında bir veri yayımlama ve tüketme altyapısı literatürde bulunmamaktadır. Bağlı veri uzayında köken farkında bir altyapı geliştirirken tez önerisine göre yedi ana madde üzerinde durulmuştur:

1. Veri uzayına köken farkında bir bağlı veri altyapısı eklemek.

¹⁸İngilizcesi: pay-as-you-go data integration



Şekil 2.4: Veri yayımlama ve tüketimini gözönüne alan veri uzayı katmanları(Bizer, 2011)

2. Mevcut verisetleri arasından köken verisi kullanarak en uygunları filtrelemek.
3. İç Veri ve Dış Veri kavramlarını netleştirerek literatüre kazandırmak.
4. Dış Veri'yi köken üstverisi ile genişletmek.
5. İç Veri'yi oluşturan organizasyonel kavramlar ile kökeni bağlı veri açısından ele alan bir ontoloji geliştirmek.
6. Organizasyonel köken tercihlerini gösteren bir ontoloji oluşturmak.
7. İç Veri, Dış Veri ve oluşturulan ontolojilerden yararlanarak köken kuralları meydana getirmek ve çalıştırmak.

Birinci madde, veri uzayına köken farkında bir bağlı veri altyapısı eklemektir. Bu altyapı oluşturulurken veriseti seviyesinde filtreleme yapılarak organizasyona en uygun verilerin sağlanması amaçlanmıştır.

Çalışmada referans alınan WIQA¹⁹ (Bizer and Cyganiak, 2009), vebde

¹⁹Açılımı: Web Information Quality Assessment Framework

erişilebilen büyük miktardaki veriyi kalite bilgisine göre filtreleyerek kullanıcıya sunmayı hedeflemiştir. Bu noktada WIQA, bilgi sistemlerinin yüksek kaliteli veri ile düşük kaliteli veriyi birbirinden nasıl ayıracağı, yüksek kaliteli veriye sahip olmak için nasıl filtreleme yapılacağı, filtreleme kararları ile ilgili kullanıcıların üzerine düşen görevlerin ne olduğu gibi sorulara cevap aramıştır. WIQA, filtreleme için veri tüketicilerine veri modeli üzerinde ifade edilen WIQA-PL dili ile geliştirilmiş politikalar sağlamaktadır. Bu politikalar ile veri kökenine dair sorgulamalar yapılmıştır.

İkinci maddede, verisetleri keşfedildikten sonra ve veriye erişim için sorgu yapılmadan önce, ilgili verisetlerinden en uygun olanlarının seçilmesine dair filtreleme üzerinde durulmaktadır. Bu noktada verisetlerinin VoID dokümanları üzerinde köken bilgisinin tanımlanması gerekliliği görülmüştür. Köken ontolojisi olarak PROV-O ontolojisi genişletilerek oluşturulan PAV ontolojisi ile kökenin üstveri seviyesinde verisetine adapte edilebileceği düşünülmüştür.

Tez çalışmasındaki yaklaşım olan PROV4FQ ile filtrelemede verisetlerinin üstverisini kullanıp kökeni gözönüne alarak bağlı veri sorgulamaktır. Bu amaçla PROV4FQ'ya en yakın çalışma olan WIQA ile karşılaştırma gereği oluşmaktadır. WIQA'da vebdeki herhangi bir veriye ait politikalar tanımlanırken, PROV4FQ bağlı veri tüketimi aşamasında VoID veriseti üstverisinin zenginleştirilmesi ile elde edilen üstveriyi sorgular. WIQA'dan farklı olarak PROV4FQ ayrıca PROV-O veri modelini kullanıp VoID üstverisini zenginleştirerek filtreleme üzerinde durmaktadır. Böylece farklı kalite boyutlarının PROV-O'nun genişletilerek adapte edildiği bir köken filtreleme çözümü sunulmaktadır. PROV4FQ, ayrıca İç Veri ve Dış Veri şeklinde veriyi sınıflandırarak veriseti filtrelemesi esnasında kullanılması gereken veriyi mimari seviyesinde netleştirmiştir. PROV4FQ, dağıtık sorgulamaya adapte edilmiş bir köken filtreleme çözümü olmakla beraber WIQA vebdeki bilgi sistemleri için kullanılan bir yaklaşımdır. PROV4FQ, bağlı veri mimarisinin temel servisi olarak ifade edilen sorgulama servisine yardımcı olarak köken farkında filtreleme servisini sunarken (bkz. Şekil 2.3) WIQA'da mimari seviyede detaylandırılan bu şekilde bir çalışma bulunmamaktadır.

Bağlı veri uzayında ele alınan üçüncü madde, veri uzayında İç Veri ve Dış Veri şeklinde ayırım yapılması gerekliliğidir. Organizasyonların bilgi tabanları, tercihleri ve keşfettiği verisetleri İçsel Veri olarak ifade edilmektedir (bkz. Şekil 1.1). Organizasyonun ihtiyaç duyacağı veri örneklerini içeren bağlı veri uzayı ise Dış Veri'yi tanımlamaktadır (bkz. Şekil 1.1). Organizasyon içinde tutulmayan fakat organizasyonun zamanla ihtiyaç duyacağı Dış Veri içerisinde VoID dokümanlarını genişleten sözlükler, VoID dokümanları ve veri örnekleri bulunmaktadır. Bağlı veri

uzayında şu ana kadar bu ayrımı ifade eden bir tanımlama yapılmamıştır. İç Veri ve Dış Veri'nin organizasyonel bakış açısı ile filtrelemede ele alınması tez önerisinin sağladığı yenilikler arasındadır.

Dördüncü madde, Dış Veri'de kökenin dikkate alınacağı tüm veriyi ifade edecek bir sınıflandırma ile köken üstverisi oluşturmaktır. Dış Veri, çalışmada yapılan sınıflandırmaya göre ham veri, anlamsal veri ve veriseti üstverisinden meydana gelmektedir. Dış Veri'deki herhangi bir veri, bu üç kategoriden birine dahildir. Bahsedilen veri durumuna ait farklı köken üstverisi oluşturularak Dış Veri'nin genişletilmesi gerektiği öngörülmüştür.

Bağlı veri uzayı ile ilgili olarak üzerinde durulan beşinci madde, köken ile organizasyonu beraberce gözönüne alan bir ontoloji oluşturulmasıdır. Organizasyon kavramı olarak, organizasyon verisi ve organizasyon dahilindeki çalışanları içerir. Ayrıca diğer organizasyonlar ile işbirliği halindedir. İşbirliği sırasında veri akışının belirli kalite kıstasları çerçevesinde gerçekleşmesi için organizasyonda kökene ait tanımların olması gerektiği görülmüştür. Bu yüzden organizasyonel kavramlar ve köken ile ilgili kavramların birleştirildiği bir ontoloji meydana getirilmiştir.

Tez önerisine göre veri uzayında ele alınan altıncı madde, VoID modeline eklenen köken üstverisine uygun organizasyonel bakış açısına göre tercih ontolojisi oluşturmaktır. Bu doğrultuda köken boyutlarından bazıları ele alınarak o boyuta dair PROV-O ve PAV ontolojileri yardımı ile tercihler meydana getirilmiştir.

Yedinci maddede ise, açık bağlı veri uzayında bulunan her organizasyonun, kendi tercihlerini gözönüne alarak elde etmek istediği cevapları filtreleyebildiği ifade edilmektedir. Bu noktada organizasyonlar tez önerisine göre adlı çizge halinde tercihlerini oluşturmaktadır. Organizasyonlar keşfettikleri veri kaynaklarından uygun olanları filtreleyerek organizasyonel alt-çizgeye erişmeye çalışmaktadır (bkz. Bölüm 3). SPARQL sorguları, bağlı veri uzayının bir alt-çizgesini döndürür. Bir alt-çizge ise bağlı veri uzayında yer alan farklı birçok veriseti tarafından kapsanan elemanları içerir. Organizasyonel tercihler hesaba katıldığında alt-çizgeye dair kısıtlamalar gelmekte ve organizasyonel alt-çizgeler elde edilmektedir.

Bahsedilen maddeler özetlenecek olursa tez çalışmasında İç Veri'nin kapsadığı VoID üstverisi ve organizasyonel tercihler kullanılarak Dış Veri'ye erişen bir köken farkında dağıtık sorgu çalıştırma mekanizması geliştirilmiştir. Adlı çizgeler şeklinde tutulan İç Veri'nin, SPARQL kural yapıları ile sorgulanması sağlanmıştır. Böylece veriseti köken üstverisi ve organizasyonel tercihlerin bağlı veri uzayında ele alındığı bir köken farkında sorgulama yaklaşımı literatüre kazandırılmıştır.

3 KÖKEN FARKINDA ORGANİZASYONLAR İÇİN OLUŞTURULAN ONTOLOJİLER

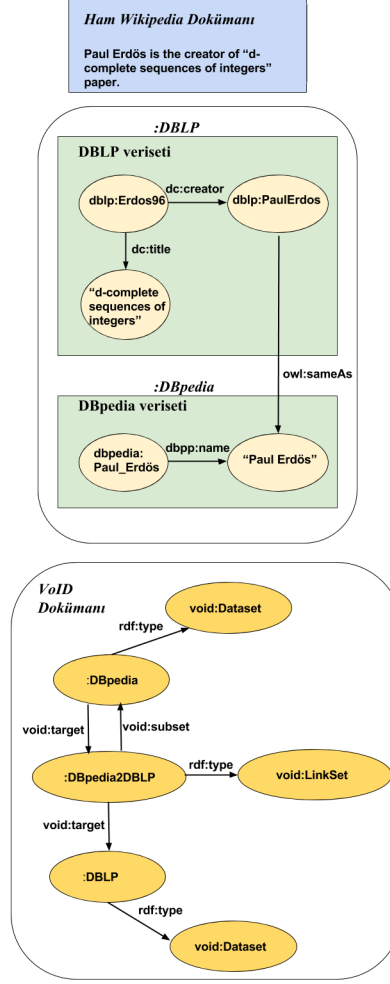
3.1 Örnek Senaryo

Vebdeki veri, vebdeki tüm veriyi anlamsallıktan bağımsız olarak ifade eden bir kavramdır. Veri Vebi ise büyük miktardaki veriyi içeren veri kaynaklarını birleştirebilmek amacı ile ortaya çıkan bir terimdir. Veri Vebi, anlamsal olmayan vebdeki veriden ve anlamsal veb teknolojilerinden yararlanarak birlikte kullanılabilir bir yapı oluşturmaya çalışmaktadır. Bu doğrultuda Veri Vebi, RDF çizgeleri biçiminde veri yayımlama, farklı veri kaynaklarındaki benzer veriyi tanımlayarak bu kaynaklar arasında bağlantı kurma ve yayımlanan bağlı veriyi ifade edecek ortak kavramları içeren sözlükler geliştirmek amacını güder.

Ham veri, insanların anlayabileceği ve RDF dışındaki bir gösterim formatında ifade edilen veridir (örneğin Wikipedia). Anlamsal veri ise makine tarafından da anlaşılabilir ve RDF ile gösterilen veridir (örneğin DBpedia). Vebdeki veri, ham veri ve anlamsal veriyi beraberce içerir. Bu durumda verisetleri arasında anlamsal bağlar çok fazla bulunmamaktadır. Benzeri özellikleri içeren veri arasında kurulan anlamsal bağlar bağlı verinin ortaya çıkmasına yol açmıştır. Ortamdaki uygun tüm bağların kurularak bu bağların açık ve keşfedilebilecek hale getirilmesi noktasında Veri Vebi meydana gelmiştir.

Standart olarak vebdeki dokümanlar HTML sayfaları ile sunulmaktadır. Gorlitz et al. (Görlitz and Staab, 2011a)'ın verdiği örnek geliştirilerek oluşturulan Şekil 3.1'de ham veri olarak dokümanların bir veb sayfasında ifade edildiği görülmektedir. Bu tip dokümanlarda doküman içerisinde linkler bulunmaktadır ve linklere tıklandığında HTTP protokolü yardımı ile internet üzerindeki diğer dokümanlara erişilmektedir. Vebdeki bağlantılı veri gözönüne alındığında doküman seviyesinden daha tanecikli bir yapı olan veri seviyesi üzerinde düşünülmelidir. Dolayısıyla bir dokümandaki tüm cümleyi okumaktan çok cümledeki elemanlara anlamsallık yükleyerek ilgili varlıklarla bağlantılar kurulmalıdır. Örnekteki DBpedia ve DBLP verisetlerinde bulunan verilerin üçlü biçiminde gösterimi bulunmaktadır. Bahsedilen örneklerdeki Paul Erdős kavramı ile farklı verisetlerinde yayınlanmış yüzlerce verisetinden ilgili olanlarla bağlar oluşturulabilir.

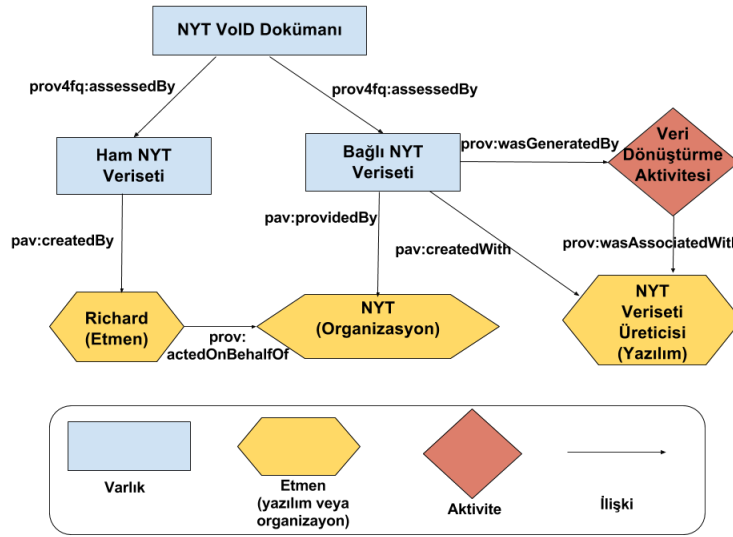
Açık veri olarak yayınlanmış verisetlerinin bağlanması bağlı açık veri kavramını ortaya çıkarmıştır. Bu noktada ise ilgili kavramın bağlı olduğu



Şekil 3.1: Ham veri, anlamsal veri ve VOID üstverisinin örnek gösterimi

verisetlerinin keşfedilmesi gerekmektedir. Keşfin sağlanması amacı ile verisetlerinin üstverisinin oluşturulması için VOID dokümanları (Alexander et al., 2009) oluşturulmaktadır. Şekil 3.1'deki örnekte DBpedia veriseti ile akademisyenlerin yayın bilgilerinin bulunduğu DBLP veriseti arasındaki ilişki veriseti seviyesinde oluşturulmuştur. Veriyi tüketmek isteyen kişinin (veya yazılımın) istediği verinin bu verisetlerinde olduğunu keşfetmesi ile beraber bağlı veriye erişim sağlanabilmektedir.

Ham verinin bağlı veri olarak yayımlanması ve bağlı veri üstverisi olan VOID tanımlarının oluşturulması ile beraber bağlı veriye dair kökenin de ifade edilmesi gerektiği görülmüştür. Bu noktada her üç veri seviyesi için de ne tip üstverilerin gösterilmesi gerektiğinin üzerinde düşünülmelidir. Bu amaçla Şekil 3.2'de New York Times (NYT) verisi için tez içerisinde ifade edilecek köken üstverisi hakkında bilgi verilmiştir. Ham NYT verisi için önemli olan kimin tarafından yaratıldığıdır. Ancak bağlı NYT verisetine bakıldığında verinin yaratılması değil yaratılan veriyi sağlayan kurum veya ham veriyi anlamsal veriye çeviren yazılım üzerinde



Şekil 3.2: VoID üzerinde kökeni ifade etmek için kullanılan senaryo

durulmalıdır. Çünkü genellikle anlamsal veri, ham veri dönüştürülerek meydana getirilir. Dolayısıyla bağlı veri için ham veriden farklı özelliklerin düşünülmesi gerekmektedir (bkz. Tablo 3.1). Veriyi yaratma ve diğer tüm operasyonlar için üç tip aktör olabilir. Bunlar kişi (etmen), organizasyon veya yazılım olabilir. Ayrıca tezdeki yaklaşımda VoID dokümanlarının ham veri veya bağlı verisetinin işlenmesine dair üstveriyi de tutması gerektiği öngörülmüştür. Böylece keşfedilecek verisetlerine dair VoID dokümanlarının içerisine köken üstverisinin de eklenmesi gerekmektedir. Tezde eklenen köken kavramları ve ilgili ontolojiler Bölüm 3.2’de detaylı olarak gösterilmiştir.

3.2 Köken Farkında Organizasyonlar için Oluşturulan Ontolojiler

Bağlı veri uzayında bilgiye sahip olmak için farklı verisetlerinden gelen veri birleştirilir. Bu birleşim esnasında verisetleri arasındaki bağlantılar esas alınır. Bağlantılar ile ilgilenirken benzeri veriyi sağlayabilecek verisetlerinden daha kaliteli olanın seçilmesi gerekir. Seçilimi yapmaya uygun sistemler, verinin nereden geldiği, ne zaman alındığı veya bağlantıların nasıl oluşturulduğu gibi kökene dair bilgiler yardımı ile kullanıcıya en uygun bütünleştirilmiş veriyi sunmalıdır. Verisetlerinin tanımlanması aşamasında veri yaratılması ve versiyonlaması ile ilgili üstverinin de sağlanması, bu bilgiler ile beraber bağlantıların oluşturulması gerekmektedir.

VoID tarafından sağlanan veriseti üstverisi (Dış Veri üstverisi), vebde ihtiyaç duyulan güven gibi gereksinimlerin (Groth et al., 2012) tümünü karşılamak için

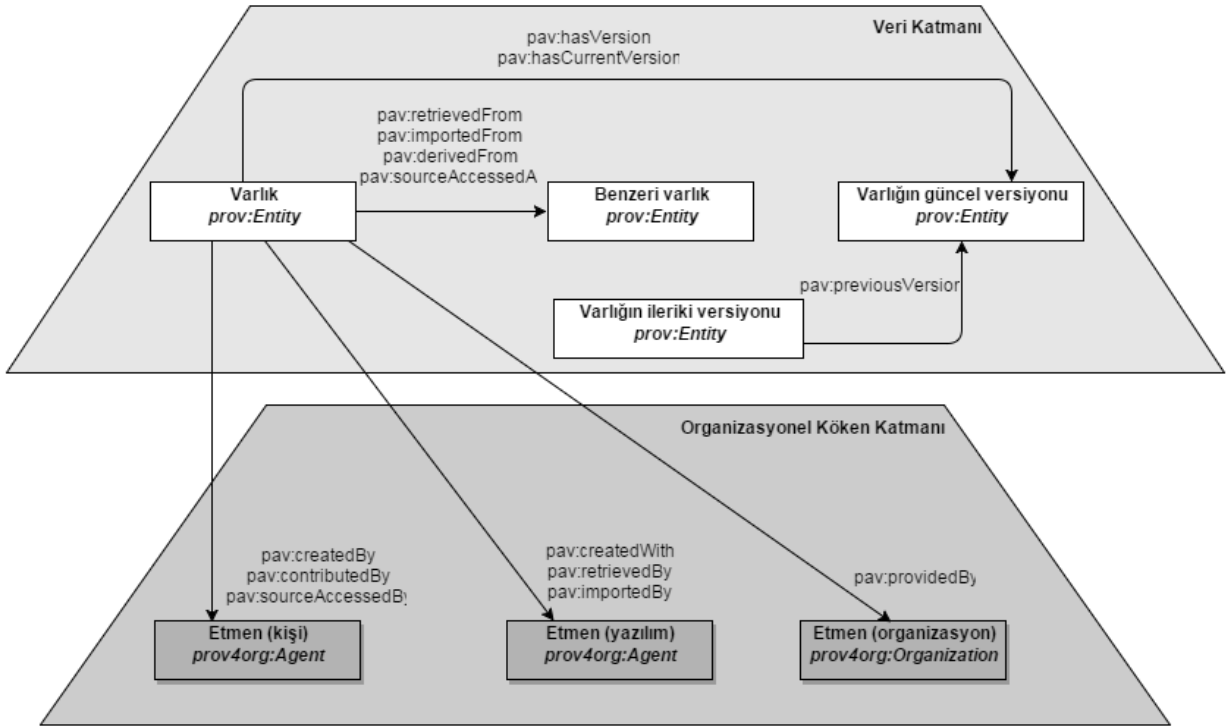
yeterli değildir. Veri yayımlayıcıları, veri ile beraber verisetlerinin üstverisini yeni ihtiyaçları adapte etmek için geniş bir perspektifte düşünmelidir. Bu yüzden köken gibi kaliteye dair ihtiyaçlar, veriseti yayımlayıcıları tarafından zenginleştirilen üstveri ile ifade edilebilmelidir. Omitola et al. (Omitola et al., 2011b), VoID'i genişleterek VoIDp adı verilen bir öneri sunmuştur. VoIDp, bir aktörün kaynak veriseti üzerinde gerçekleştirdiği süreçten sonra meydana gelen sonuç veriseti ile ilgilenir. Yalnız bahsedilen VoID zenginleştirmesi kapsamlı ihtiyaçlara dair organizasyonel tercihleri karşılayamayacağından başka bir öneri sunulması ihtiyacı ortaya çıkmıştır.

Veriseti üstverisi yanında organizasyon verisinin de gözönüne alınması için organizasyonların çalışma alanına uyabilecek genel kapsamlı bir sözlüğe ihtiyaç vardır (İç Veri üstverisi). Bu sözlük oluşturulurken sorgulamada İç Veri'de bulunan organizasyonel tercihler ve organizasyon verisinin temel köken kavramları ile eşleştirilmesi gerektiği düşünülmüştür. Aynı durum zenginleştirilmiş yerel VoID ambarında bulunacak dokümanlardaki genişletmeler için de geçerlidir. Dolayısıyla kökene ait kabul görmüş kavramları içeren ontolojilerin kullanılması ve ihtiyaçlara göre genişletilmesi gerektiği düşünülmüştür.

Köken için bir gösterim modeli olan PROV-O (Lebo et al., 2013) ve onu genişleten PAV (Ciccarese et al., 2013), köken farkında filtreleme metodolojisi yaratmak için tez çalışmasına uygun olduğu için seçilmiştir. PROV-O, PROV veri modelini RDF'e eşleyen PROV ontolojisidir. PAV ise veb kaynaklarının yazarlık, köken ve versiyonlama ile ilgili olarak izlenmesini sağlamak için PROV-O ontolojisini genişleterek geliştirilmiştir. VoID ile veriseti üstverisinin yaratılması aşamasında VoID dokümanına ve ilişkili verisetine dair üstverinin gözönüne alınması gerekmektedir. VoID dokümanının kim tarafından ve hangi zamanda yaratıldığı, en son ne zaman güncellendiği köken ile ilgili üretilmesi gereken üstveriyi meydana getirir. Veriseti ile ilişkili olarak ise verisetinin ham hali ve işlendikten sonraki durumu için köken bilgisi ayrı ayrı takip edilebilir. Verisetinin ham hali, veri RDF formatında olmadan önce de yayımlandı ise dikkate alınmalıdır. Bu durumda verisetinin kim tarafından yayımlandığı, nasıl dağıtımına verildiği, lisanslama bilgisi ve bu halin versiyon bilgisi tanımlanabilir. Verisetinin dönüşümlerden geçerek RDF formatında yayımlandığı noktada verisetinin yaratılmasında yardımcı olan araç veya yazılım hakkında da bilgi verilir. Veriseti, bazı dönüşümlerden geçerek bu noktaya geldi ise o dönüşümlere ait bilginin de VoIDp (Omitola et al., 2011b) üzerinde yapıldığı gibi ifade edilmesi gerekir. RDF ile ifade edilen verisetinin versiyonlamasının yapılması, küçük ve büyük değişiklikler ile meydana gelen verisetlerindeki değişimlerin anlaşılması açısından önemlidir.

3.2.1 Dış veri bakış açısı

Tez çalışmasında VoID genişletilerek verisetlerine dair üstverinin ve organizasyon verisine dair köken ifadelerinin tanımlanması gerektiği öngörülmüştür. Bu doğrultuda çalışma, Şekil 3.3’de görüldüğü üzere veri katmanı ve organizasyonel köken katmanı olmak üzere iki ayrı katmandan meydana gelmiştir. Veri katmanı, köken üstverisinin ifade edilmesi için gereken kavramları oluşturur. Organizasyonel köken katmanı, köken varlıklarını oluşması veya güncellenmesinde etkili olan etmenler (kişi, yazılım veya organizasyon) ile etmenlerin birbirleri ve diğer organizasyonlarla ilişkilerini ifade eder. Bu ontoloji, farklı bağlamları ifade eden senaryolara adapte edilebilmek amacı ile geliştirilmiştir.



Şekil 3.3: PAV Ontolojisi ile VoID Dokümanlarının Köken için Genişletilmesi (PEVoID)

Veri katmanı olarak ifade edilen kısımda kalite ile ilgili üstveri eklenmesi gerektiği görülmektedir. Üstverinin, PROV-O tarafından sağlanan temel köken ifadeleri ile oluşturulması aşamasında üç noktaya dikkat edilmelidir. Birincisi, oluşturulan VoID üstverisine ait köken bilgisi, ikincisi eğer var ise veriseti anlamsal hale gelmeden önceki haline dair köken bilgisi, üçüncüsü ise anlamsal olarak yayımlanan/yayımlanacak verisetine ait köken bilgisinin ifade edilmesi şeklindedir. VoID üstverisine ait köken bilgisi, dokümanın kim veya hangi araç tarafından üretildiği, hangi zamanda yaratıldığı ve ne zaman güncellendiğini içerir. Verisetinin anlamsal hale gelmeden önceki haline dair köken bilgisi,

ilgili verinin yayımlayıcısı, veri başlığı, lisans bilgisi ve versiyonunu içerir. Anlamsal olarak yayımlanan/yayımlanacak verisetine ait köken bilgisi, ham veriye ait üstveri yanında verisetinin hangi verisetinden, ne zaman, kim tarafından aktarılarak bu noktaya getirildiğini de gösterir. Her üç durum için de farklı adlı çizgelerin oluşturulup sorgulamaya dahil edilmesi ile veriseti seviyesinde filtreleme yapılabileceği görülmüştür. Bahsedilen üç durum için de kökenin ele alınmasına imkan veren bir sözlük ihtiyacı görülmektedir. Benzeri amaçla üretilen PAV (Cicarese et al., 2013) ontolojisi, veriseti ile ilgili üç farklı köken kontrolünün sağlanması için özellikler sunmuştur. Bu özelliklerin veriseti seviyesinde filtreleme için uyarlanmış hali Şekil 3.3’de görülmektedir.

Şekil 3.3’de öncelikle varlığın ve varlığın zaman içindeki değişiminin takip edilmesi gerekmektedir. Buna göre varlık olarak bahsedilen kavram, VoID dokümanı, ham veriseti veya yayımlanmış veriseti olabilir. Yani VoID dokümanı içerisinde VoID dokümanının kendi bilgisi, ham verisetinin üstverisi veya yayımlanmış verisetinin üstverisi farklı adlı çizgeler halinde bulunabilir. Her bir varlık için hangi PAV ontolojisi özelliklerinin ele alınabileceği Çizelge 3.1’de gösterilmiştir.

VoID dokümanı için oluşturulan köken tanımında öncelikle dokümanın kim tarafından yaratıldığı (*pav:createdBy*) ve yaratım sürecine kimin katkıda bulunduğu (*pav:contributedBy*) ifade edilir. Dokümanın yaratılma sürecinde hangi yazılım aracının kullanıldığı (*pav:createdWith*) da köken takibinde rol alır. Ayrıca dokümanın ne zamanın yaratıldığı (*pav:createdOn*) ve güncellenme anına ait zaman bilgisi (*pav:lastUpdateOn*) de kullanılan PAV özelliklerindendir. Doküman bir kişi tarafından yaratılabilirken bu sürece yazılımın katkıda bulunduğu Şekil 3.1’de ifade edilmiştir.

Ham veriseti, metin halinde oluşturulmuş verileri temsil etmektedir. Bu veri, Veri Vebi’nde herhangi bir formatta bulunuyor olabilir. Bahsedilen verinin bu şekilde bulunduğu anda köken ile ilgili hangi özelliklerinin takip edilmesi gerektiğine dair detaylar Çizelge 3.1’de açıklanmıştır.

VoID dokümanları için olduğu gibi ham verisetleri için de veri yaratıcısı ve ham veri oluşturma sürecine katkıda bulunan kişiler bulunmaktadır. Bahsedilen ham veri, herhangi bir veri kaynağından elde edilmiş ve değiştirilmeden kullanılıyor olabilir. Bu durumun ifadesi için veriyi çeken yazılımsal etmen özelliği (*pav:retrievedBy*) ve çekilen kaynak (*pav:retrievedFrom*), verinin aslen kim tarafından oluşturulduğuna dair bilgi sahibi olunması için önemlidir. Ham verisetinin elde edilmesinde değişiklik yapmadan alıp kullanmanın yanında veri

hakkında fikir sahibi olup yorumlanmış bir verisetine de sahip olunmuş olabilir. E-devlet veb sitesinden veri elde edip işlendikten sonra istatistiksel olarak Türkiye İstatistik Kurumu tarafından sunulan veri olduğunu varsayalım. Bu durumda asıl veri kaynağı e-devlet veb sitesidir. Bu veriye ulaşan yazılım (*pav:sourceAccessedBy*) ve bu verinin bulunduğu asıl kaynak (*pav:sourceAccessedAt*) köken takibi yapılabilmesi için VOID içerisinde ifade edilmelidir. Ham veriseti elde edilirken aynı veri formatında olmasına rağmen bazı değişiklikler sonucu bulunduğu noktaya gelmiş olabilir. Bu durumda verisetinin hangi verisetinden türediği (*pav:derivedFrom*) ile ilgili üstveri bilgisinin ifade edilmesi gerekmektedir.

Varlık olarak ifade edilen eleman	PAV ontolojisi özelliği
VOID dokümanı	<i>pav:createdBy</i> <i>pav:contributedBy</i> <i>pav:createdWith</i>
Ham veriseti	<i>pav:createdBy</i> <i>pav:contributedBy</i> <i>pav:retrievedBy</i> <i>pav:sourceAccessedBy</i> <i>pav:retrievedFrom</i> <i>pav:derivedFrom</i> <i>pav:sourceAccessedAt</i> <i>pav:providedBy</i> <i>pav:hasVersion</i> <i>pav:hasCurrentVersion</i> <i>pav:previousVersion</i>
Yayımlanmış anlamsal veriseti	<i>pav:createdBy</i> <i>pav:contributedBy</i> <i>pav:createdWith</i> <i>pav:importedBy</i> <i>pav:importedFrom</i> <i>pav:derivedFrom</i> <i>pav:providedBy</i> <i>pav:hasVersion</i> <i>pav:hasCurrentVersion</i> <i>pav:previousVersion</i>

Çizelge 3.1: VOID Dokümanına Eklenen Veri Yapılarına Uygun PAV Köken Özellikleri

Bir kaynaktan alınan veya türetilen verisetinin alındığı kaynağı yaratan kişiden başka o kişinin bağlı olduğu organizasyon bilgisi de önemlidir. Bu

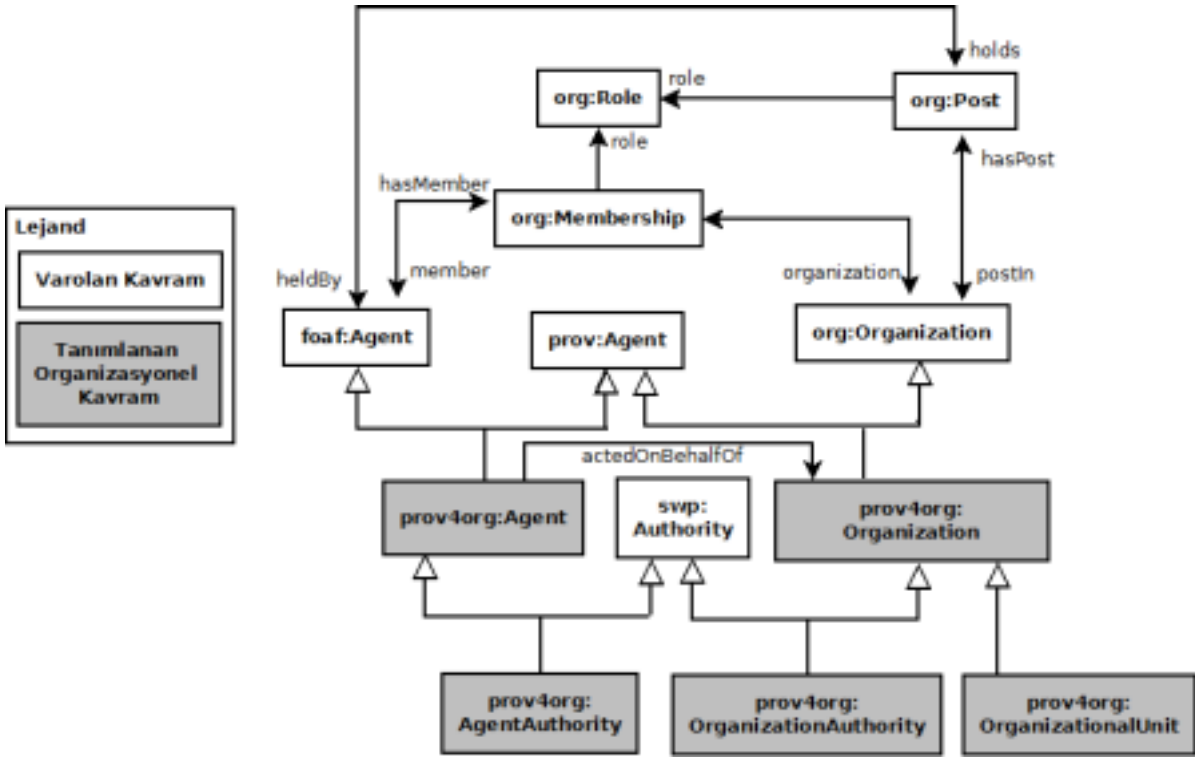
organizasyona (veya kuruluşa) olan güvenilirlik asıl kaynak ile ilgili filtrelemede esas alınır. Bu amaçla kaynağı sağlayan organizasyon bilgisi (*pav:providedBy*) de üstveri olarak ifade edilmelidir. Bu değişimlere maruz kalan verisetlerinin versiyonu ile ilgili bilgilerin tanımlanması (*pav:hasVersion*, *pav:hasCurrentVersion*, *pav:previousVersion*) veri değişimi esnasında büyük veya küçük değişimlerin olup olmadığına dair bilgi sahibi olunmasına yardımcı olur.

Yayımlanmış veriseti için, ham veriseti içeriğinin aynı kalıp formatın anlamsal hale getirilmesi ile ilgili bilginin tutulması gerekmektedir. Gösterim modelinin değiştirildiği kaynak verisetinden (*pav:importedFrom*) anlamsal verisetine dönüştürmeyi yapan yazılımın (*pav:importedBy*) belirtilmesi köken takibi için gereklidir. Ayrıca yayımlanmış verisetleri sadece varolan verisetinin dönüştürülmesi ile meydana gelmemektedir. Yeni olarak üretilen anlamsal verisetlerinin yaratıcısı (*pav:createdBy*), verisetine yaratma aşamasına katkıda bulunan kişi (*pav:contributedBy*) ve verisetini yaratırken hangi yazılım aracının (*pav:createdWith*) kullanıldığının da belirtilmesi önemlidir. Ham verisetinde olduğu gibi yayımlanmış veriseti için de asıl kaynağı sağlayan organizasyon bilgisinin ifade edilmesi (*pav:providedBy*) köken takibi için önem arz eder.

3.2.2 İç veri bakış açısı

İç Veri’de kullanılmak üzere her organizasyonel alan için ihtiyaç duyulan kavramları tanımlayan ontoloji Şekil 3.4’de gösterilmiştir. Şekil 3.3’de gösterilen *prov4org:Agent* ve *prov4org:Organization* tanımlarının ilişkileri Şekil 3.4’de bulunmaktadır. Bahsedilen tanımlar, organizasyon ontolojisi (ORG) (Reynolds, 2014), köken ontolojisi (PROV-O) (Lebo et al., 2013) ve SWP ontolojisi (Bizer et al., 2010) genişletilerek kökene adapte etmek amacı ile oluşturulmuştur.

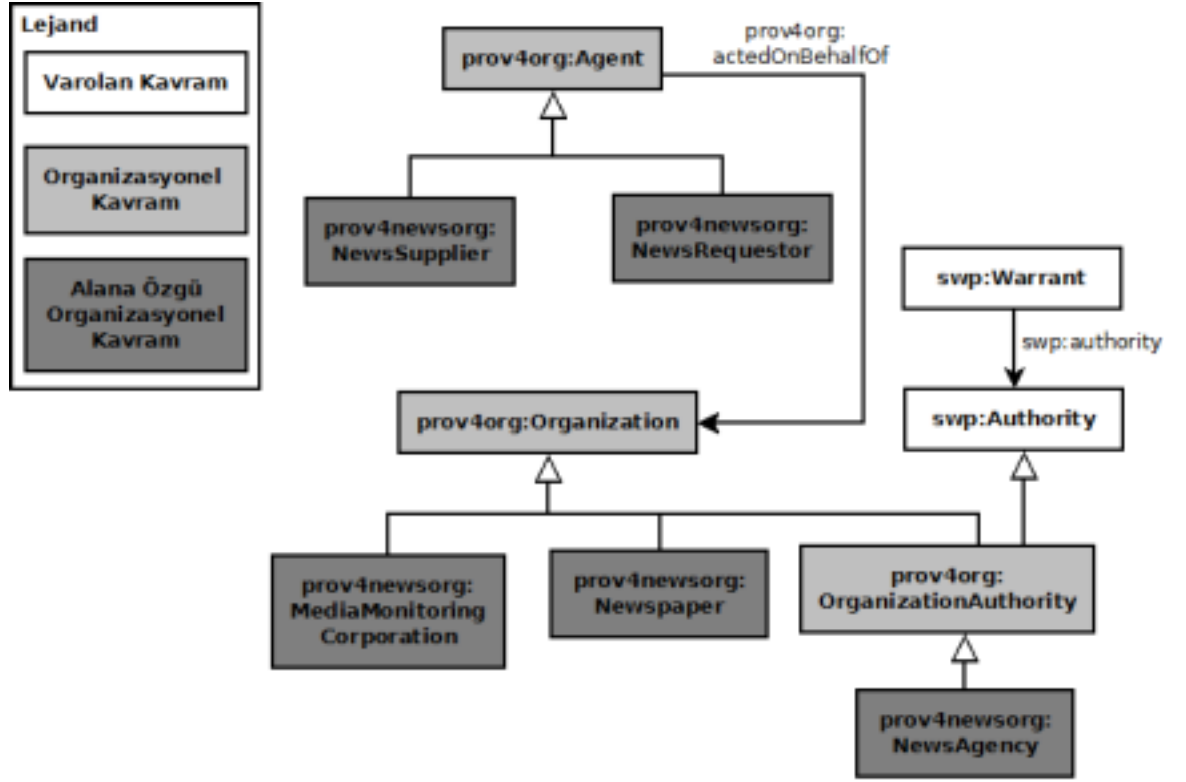
prov4org:Agent ifadesi organizasyonda bulunan bir kişiyi ifade edebilmekle beraber bir yazılım olarak da tanımlanabilir. Kişi olarak düşünüldüğünde etmenin, organizasyonun bir üyesi olduğunun (*org:Membership*) gözönüne alınması gerekmektedir. Üyelik kavramı bir firmadaki hiyerarşik pozisyonu temsil eden ilişkiyi gösterir. Üyelik, kişinin rolü (*org:Role*), bulunduğu organizasyon (*org:Organization*) ve kişisel bilgilerinin (*foaf:Agent*) ifade edildiği ilişkisel yapı ile ifade edilmektedir. Dolayısıyla organizasyondaki her çalışanın bir üyeliği vardır. Üyeliğin sürdüğü zaman diliminin veya rollerin değişiminin (çalışanların zaman içinde terfi etmesi gibi) de takip edilmesi için kökende üyelik kavramının hesaba katılması gerektiği görülmüştür. Pozisyon bilgisi ise (*org:Post*), organizasyonel hiyerarşide ifade edilen fakat pozisyona dair görevli bir kişinin olmasının gerekli



Şekil 3.4: Organizasyonel Köken Ontolojisi

olmadığı durumları ifade etmektedir. Bu doğrultuda pozisyon, kendisini dolduran bir kişi olmadan da varolabilir. Ancak üyelik, organizasyon ile etmeni birbirine bağlayan ve bunlar olmadan var olmayan bir ilişkiyi temsil eder. Bir organizasyon profili meydana getirilirken çoğunlukla pozisyon veya üyelikten biri tercih edilerek çalışma alanında tanımlanır. Organizasyonda çalışan kişilerden bağımsız bir yapı ifade edilmek isteniyorsa pozisyonlar tanımlanır. Eğer amaç organizasyonu oluşturan kişilerin kaydını tutmak ve bu kişilerin yetilerine uygun kökene dair sorgulamalar yapmak ise üyelik kavramı oluşturulması gerekir. Organizasyonel birim (*org:OrganizationalUnit*), büyük bir organizasyondaki departman gibi daha küçük birimleri temsil eder. Organizasyonel olarak bazı operasyonların yapılmasına dair oluşturulacak kurallar, organizasyonel birim seviyesinde meydana getirileceğinden köken için organizasyonel birim kavramının ifade edilmesi gerektiği görülmüştür. Organizasyon içerisinde güvene dair ifadelerin tanımlanması için de etmen otoritesi (*prov4org:AgentAuthority*) ve organizasyon otoritesi (*prov4org:OrganizationAuthority*) olmak üzere iki kavram tanımlanmıştır.

Organizasyonel köken katmanına dair ontoloji, her organizasyonun kapsamı birbirinden farklı olduğundan dolayı her çalışma alanı için farklı oluşturulmalıdır. Şekil 3.5’de Haber Toplama Senaryosu için organizasyonel köken ontolojisi ele alınmaktadır.



Şekil 3.5: Haber Toplama Senaryosu için Organizasyonel Köken Ontolojisi

Köken için üstverinin ontoloji mühendisliği perspektifine göre oluşturulması aşamasında öncelikle veriyi kullananlar tanımlanmalıdır. Örnekteki Haber Toplama Senaryosu'nda kullanıcılar ve kullanıcıların dahil olduğu organizasyonların tanımlanması gereklidir. Bu senaryoya uygun farklı organizasyonlar ve organizasyonel kişiler veri yayımlayıcıları tarafından oluşturulmaktadır. Bunun yanında etmen ve organizasyonların eriştiği varlıklara dair üstveri, varlık ve etmen arasındaki bağlantıya göre (veriyi yaratma, sağlama veya veri erişimi gibi) Şekil 3.3'deki gibi ifade edilmelidir.

Şekil 3.5'de etmen (*prov4org:Agent*), organizasyon (*prov4org:Organization*) ve organizasyon otoritesi (*prov4org:OrganizationAuthority*), Haber Toplama Senaryosu'nu oluşturan temel elemanlardır. Senaryoda etmen, veriyi üreten kişi olarak haber sağlayıcısı (*prov4newsorg:NewsSupplier*) veya veri tüketen kişi olarak haber isteyen (*prov4newsorg:NewsRequestor*) şeklinde genişletilebilir. Organizasyon ise bir gazete (*prov4newsorg:Newspaper*), Medya İzleyen Kurum (*prov4newsorg:MediaMonitoringCorporation*) veya haber ajansından (*prov4newsorg:NewsAgency*) biridir. Organizasyon otoritesi (*prov4org:OrganizationAuthority*), gazetede haberin yayımlanması veya değiştirilmesi esnasındaki sorumluluğun kime ait olduğunu gösterir. Örnekte otorite görevini üstlenen birimin haber ajansı olacağı öngörülmüştür. Otoriteye

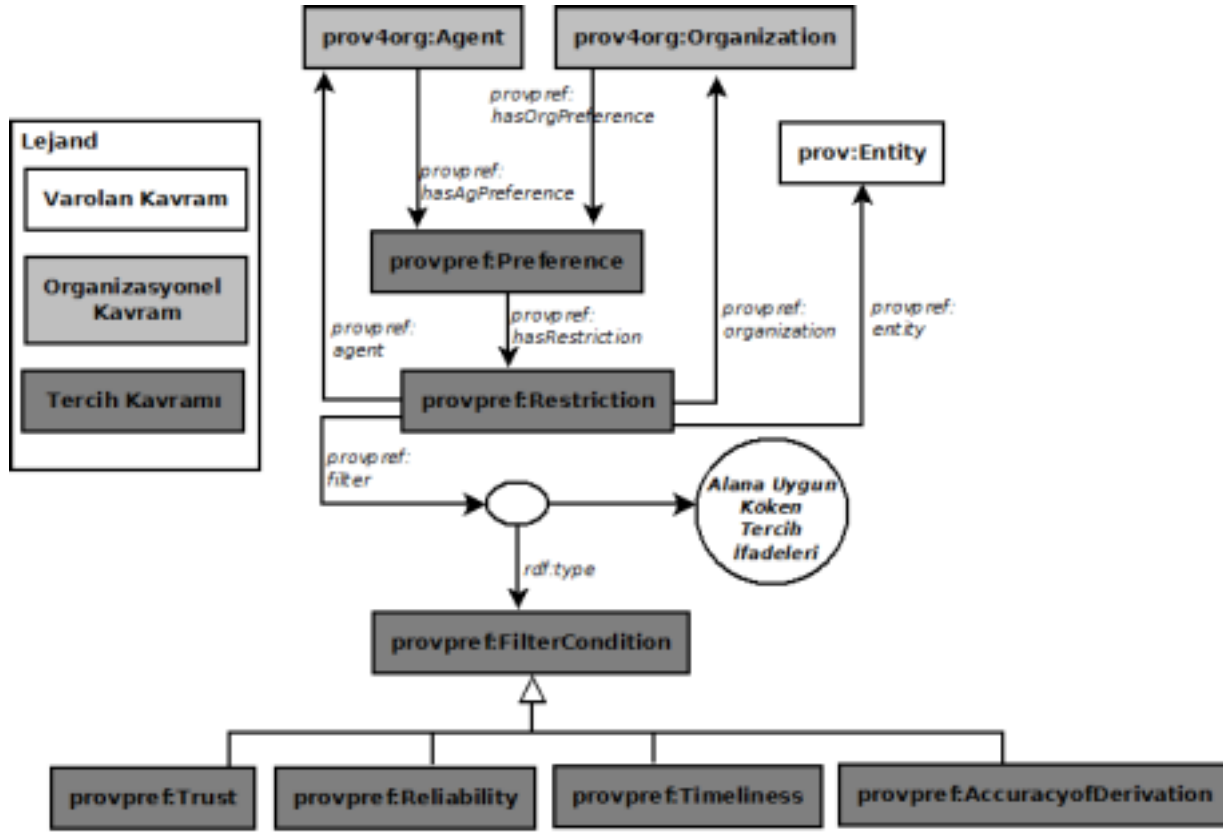
dair kavramları güven gereksinimini sağlamak amacı ile ifade ederken seçilen senaryo için Anlamsal Veb Yayımlama Sözlüğü²⁰ (SWP) kullanılmış ve bu sözlük organizasyonel köken ontolojisi ile ilişkilendirilmiştir. Köken boyutu olarak sadece güveni gözönüne almak yeterli değildir. Dolayısıyla organizasyonel köken ontolojisi farklı ihtiyaçları gözönüne alacak şekilde daha da genişletilebilir. Tez kapsamında bu ontoloji örneğe uygun biçimde ifade edilmeye çalışıldığından sadece güven gözönüne alınmıştır.

İç Veri bakış açısına göre önemli bir nokta da organizasyonun kendisi veya organizasyondaki diğer etmenlerin tercihlerinin dikkate alınması gerekliliğidir. Bu yüzden tercihlerin ifade edilmesi için alandan ve uygulamalardan bağımsız bir ontoloji gerekmektedir. Böylece organizasyonel tercihler de dikkate alınarak Dış Veri’de bulunan verisetlerinden organizasyonun ihtiyaçlarına en uygun olanı seçilebilir. Tercihler gözönüne alınırken bağlı veri uzayının (Dış Veri) organizasyona (İç Veri) dair tercihleri veya organizasyonun (İç Veri) bağlı veri uzayına (Dış Veri) dair tercihleri olmak üzere iki durum görülmektedir. Dış Veri ile ilgili tercihler genellikle gizlilik konusundadır. Çünkü veriyi sağlayan yayımlayıcılar, sağlanan verinin herkes tarafından erişilmesini istemeyebilir (Dava dosyalarını sağlayan Adalet Bakanlığı’nın dosyaları sadece ilgili hakim, savcı ve avukatlar tarafından görülmesini istemesi gibi). Bu durumla ilgili geliştirilen PPO (Privacy Preference Ontology for Linked Data) (Sacco and Passant, 2011) vebde dağıtık olarak bulunan verinin kullanıcılar tarafından erişimine dair bir çözüm getirmiştir.

Tez çalışmasında ise gizlilik gibi belirli bir boyuttan öte daha genel ve organizasyonun bağlı veri uzayına dair sahip olabilecekleri ile ilgili, FRAP ontolojisi (Polo et al., 2014) benzeri bir yaklaşım sunulmaktadır. Şekil 3.6’da gösterildiği üzere herhangi bir çalışma alanından bağımsız şekilde organizasyonel olarak ifade edilen ontolojik tanımlara tercihler adapte edilebilmelidir. Genel bir tercih ontolojisinin ifade edilmesi ile kişilerin veya organizasyonel birimlerin sahip olduğu tercihlerin taşınabilmesi de sağlanmaktadır (Bir organizasyonel birimden diğerine geçen bir çalışanın tercihlerinin devam etmesi veya bir organizasyonel birimin tercihlerinin farklı birimlerde de kullanılması gibi).

Şekil 3.6’daki sınıfları incelediğimizde *prov4org:Agent*, *prov4org:Organization* ve *prov:Entity*, Şekil 3.5’deki temel elemanlar görülmektedir. *prov4org:Agent* ve *prov4org:Organization*, tercihlerin etmen veya organizasyon seviyesinde ifade edilebileceğini göstermektedir. Tercihler aslında İç Veri’deki elemanların kısıtlarını temsil etmektedir. Dolayısıyla *provpref:Restriction*, organizasyon veya organizasyondaki bir etmenin başka bir etmen, organizasyon

²⁰İngilizcesi: Semantic Web Publishing Vocabulary



Şekil 3.6: Organizasyonel Tercih Ontolojisi

veya varlık üzerindeki tercihiine dair kısıtları göstermek için kullanılır. Kısıtın içerisindeki filtreleme koşulu (*provpref:FilterCondition*) üzerinde tercih yapılan elemana dair beklentiyi göstermektedir. Filtreleme koşulu, ifadesiz düğüm²¹ ile farklı tercihlere bağlanmıştır. İfadesiz düğümler, bir veb kaynağı birden çok veriyi tanımlamak istediğinde kullanılan veri yapılarıdır. Örneğin bir profesörü ifade ederken profesörün ismi, çalıştığı kurum, departmanı gibi verileri ifade etmek için bu yapı kullanılabilir. Şekil 3.6'da ifadesiz düğüm kullanılmasının nedeni, bir tercihe ait farklı üçlü tanımlarının oluşturulabileceği veri yapılarına ihtiyaç duyulmasıdır. Bu üçlü tanımları, Çizelge 3.2 ve Şekil 3.7'de detaylandırılmıştır. Bu ontoloji ile ilgili olarak Şekil 5.4'de NYT'nin sadece Amerika'daki haber ajanslarından gelen veriyi almasına dair bir örnek bulunmaktadır.

Filtreleme koşulu, köken ile ilgili Ek-4'de belirtilen farklı köken boyutlarını ele alacak şekilde ifade edilebilir. Şekil 3.6'daki tercih ontolojisinde Toniolo ve arkadaşlarının (Toniolo et al., 2014) köken için tercih kriteri olarak ele aldıkları Güven (*provpref:Trust*), Güvenilirlik (*provpref:Reliability*), Güncellik (*provpref:Timeliness*) ve Türeme Doğruluğu (*provpref:AccuracyofDerivation*) kullanılmıştır. Güven, köken takibinde bulunurken verinin üretilmesinde rol

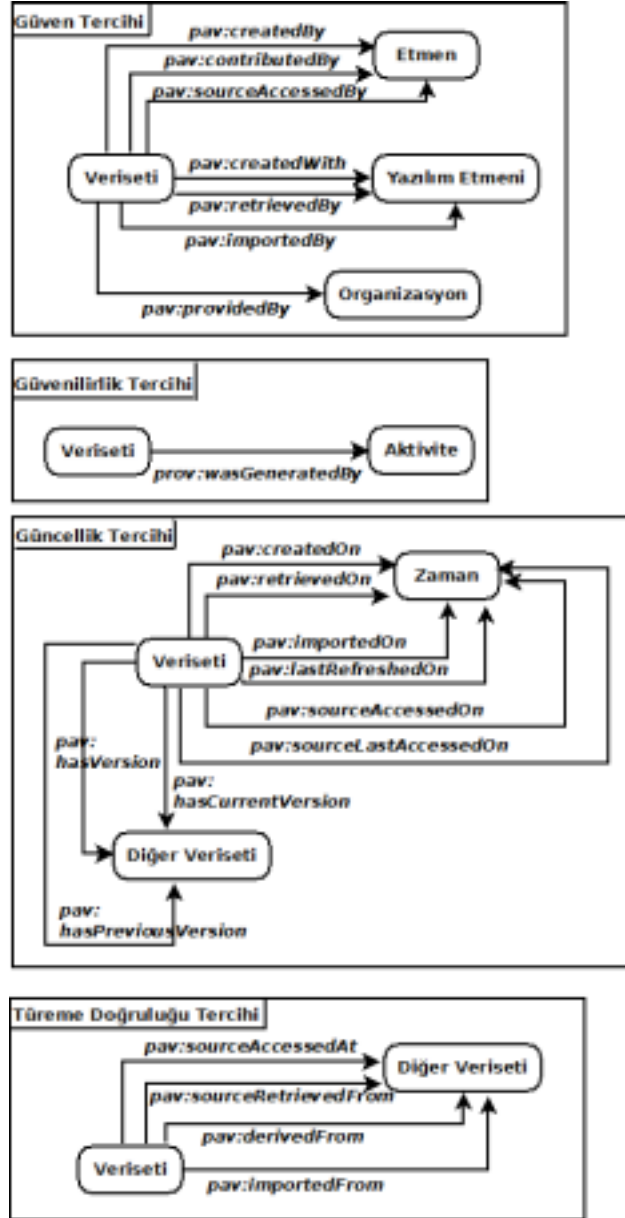
²¹İngilizcesi: blank node

oynayan aktörün güvenli olup olmadığına dair tercihi ifade eder. Güvenilirlik ise verinin türetilmesi esnasındaki (ham verinin yaratılması, ham verinin anlamsal veriye dönüştürülmesi gibi) aktivitelerin tercih edilmesi noktasında devreye girer. Güvenilir olan veri, bir aktivite tarafından farklı biçimde dönüştürülerek güvenilir olmayan biçimde yayımlanmış olabilir. Bu durumda aktivite güvenilirliğinin de organizasyon tarafından tercih olarak tanımlanması önem kazanır. Güncellik, verinin oluşturulma ve güncellenme zamanlarının takibi ve buna göre organizasyonun tercihlerini oluşturması ile ilgilidir. Türeme Doğruluğu, verinin meydana getirildiği ilk andan itibaren üzerinde oluşturulan değişikliklerin takibi üzerinde durur. Değerlendirmede veri, en az sayıda türeme ile şu ana geldiyse tercih edilebilir. Bahsedilen tercih kriterleri tez çalışmasında ayrı ayrı ele alınıp filtrelemeye dahil edilmiştir. İleriki çalışmalarda köken tercihlerine ayrı ayrı ağırlık vererek (Dividino et al., 2012) veya tercihlerin birbirlerine olan üstünlüğünü tanımlayarak (Toniolo et al., 2014; Toniolo et al., 2015) daha çok tercih edilebilir verisetlerinin nasıl oluşturulabileceğinin gösterimi üzerinde durulacaktır.

Tercih boyutu	PAV ontolojisi özelliği	Verisetini Etkileyen Köken Sınıfı
Güven	pav:createdBy pav:contributedBy pav:createdWith pav:retrievedBy pav:sourceAccessedBy pav:providedBy pav:importedBy	Etmen
Güvenilirlik	prov:wasGeneratedBy	Aktivite
Güncellik	pav:createdOn pav:retrievedOn pav:importedOn pav:lastRefreshedOn pav:sourceAccessedOn pav:sourceLastAccessedOn pav:hasVersion pav:hasCurrentVersion pav:hasPreviousVersion	Varlık
Türeme Doğruluğu	pav:sourceAccessedAt pav:retrievedFrom pav:derivedFrom pav:importedFrom	Varlık

Çizelge 3.2: Organizasyonel Tercihler Oluşturulurken Dikkate Alınacak Elemanlar

Tez çalışmasında kritik olan nokta, veriseti ile ilgili üstverinin köken takibine olanak sağlamasıdır. Böylece verisetinin üstverisine etki eden Etmen, Aktivite ve Varlık örneklerinin ne şekilde değişiklikler yaptığı veriseti üstverisinde oluşturulabilmektedir. Çizelge 3.1’de farklı veri seviyeleri için ne tip bağlantılar kurulabileceği ifade edilmiştir. Tercih bakış açısından önemli olan ise veriseti üstverisini etkileyen Etmen, Aktivite ve Varlık ile ilgili hangi özelliklerin organizasyon tarafından gözönüne alınması gerektiğidir. Bu amaçla Çizelge 3.2 oluşturulmuştur.



Şekil 3.7: Farklı organizasyonel köken tercihlerinin gösterimi

Çizelge 3.2’den yola çıkarak Şekil 3.7’de tercihlerin veriseti seviyesinde nasıl ifade edileceği gösterilmiştir. Şekildeki Güven tercihi, organizasyonun kullanacağı

verisetlerini yaratan, yaratılmasında yardımcı olan veya verisetine erişimde yardımcı olan etmenlerin hangileri olduğunu ifadesinde kullanılır. Etmen, bir yazılım etmeni ise verinin yaratılması, elde edilmesi ve organizasyona aktarılması aşamalarında da güven tercihleri belirtilebilir. Eğer veriyi sağlayan bir organizasyon ise ilgili organizasyona ait tercihler de meydana getirilebilir. Güvenilirlik tercihinde verisetini oluşturan aktiviteye dair tercihler belirtilir. Güncellik tercihi, verisetinin organizasyona alınması, oluşturulması veya erişilmesi anındaki zamanlar ile tercihler ortaya koyar. Ayrıca verisetinin güncel versiyonu ile ilgili tercihler de tanımlanabilir. Türeme Doğruluğu tercihi ise, verisetinin organizasyona alınmadan önceki halinden şu ana kadar geçen zamandaki operasyonları inceleyerek en az değişimi olan verisetinin alınmasını öngörür.

Tercihler için kendi boyutu içerisinde veya farklı tercihlerle beraber karmaşık bağlı kurallar (Khandelwal et al., 2010; Jacobi et al., 2011; Khandelwal et al., 2011) üretilebilir. Bu kurallar bağlı veri ile ilgili üstverinin bağlanmasını sağlamaktadır. Bu doğrultuda tercihler arasındaki öncelikler gözönüne alınarak farklı kurallar da öncelikler ifade edilerek birbirine bağlanabilir. Gelecek çalışmalarda bağlı kurallar (Khandelwal et al., 2011) ve üstveri yardımı ile oluşturulan köken tercihleri birleştirilerek daha kapsamlı köken takibi üzerinde çalışılacaktır.

3.3 Kullanılan Senaryo ve Ontolojiler

Bağlı açık veri çalışmalarının gelişmesi ile verisetleri oluşturularak Veri Vebi'nde yayımlanması üzerine yoğunlaşmıştır (Möller et al., 2010). Bu verisetleri farklı sözlükler yardımı ile kendi alanına ait veri ve üstveriyi tanımlayarak akademik veya ticari çalışmalarda kullanılmak üzere sunulmuştur. Köken ihtiyacına dayalı üç ana senaryo (Haber Toplama Senaryosu, Hastalık Salgın Senaryosu ve İş Sözleşmesi Senaryosu) W3C Köken Grubu tarafından ele alınarak Groth et. al. (Groth et al., 2012) tarafından anlatılan gereksinimler üzerinde bu senaryolar detaylandırılmıştır. Tez için bu senaryolar arasından varolan verisetlerine uygunluk incelenmiş ve New York Times veriseti (TheNewYorkTimesCompany, 2010) seçilerek Haber Toplama Senaryosu üzerinde çalışılmıştır.

RDF kullanan veri sağlayıcılarının, bir çizgede tanımlanan verinin doğruluk değerini ifade edebilecekleri kesin bir yol bulunmamaktadır. Fakat veri tüketicileri, reddetmek, kabul etmek, üzerinde yorum yapmak veya fikir sahibi olabilmek için bu bilgiyi talep etmektedir. Dolayısıyla bu talebi karşılayacak bir sözlük ihtiyacı doğmuştur ve bu amaçla SWP sözlüğünün (Carroll et al., 2005) kullanılmasının

uygun olacağı düşünülmüştür.

SWP sözlüğünde (Carroll et al., 2005) bulunan *swp:authority* özelliği çalışmada aynen kullanılmaktadır. Fakat bu sözlükte bahsedilen *rdg:Graph* ve *swp:Warrant* arasındaki bağlantı tez çalışmasında *prov:Entity* ve *swp:Warrant* arasında kurulmuştur. Bağlı veri sorgulamanın farklı seviyelerinde bulunan varlıklar köken bazlı olarak sorgulanmalıdır. Bu varlıkların her biri bir adlı çizge değildir, sadece bir URI ile temsil edilen varlıklardır ve *prov:Entity*'yi genişletmektedir.

Varolan ontolojilerin yeniden kullanılması dışında tez çalışmasına dair oluşturulan ontolojiler ve bu ontolojilerden yararlanan adlı çizgeler ise;

- Zenginleştirilmiş VoID üstveri ontolojisi (PEVoID) (bkz. Şekil 3.3),
- Organizasyonel kavramları içeren ontoloji (bkz. Şekil 3.4 ve Şekil 3.5),
- Organizasyonel tercih ontolojisi (bkz. Şekil 3.6),
- Organizasyon verisinin ifade edilmesi için kullanılan alan ontolojileri (TheNewYorkTimesCompany, 2010),
- Organizasyonların kullanacağı verisetlerinin üstverisini içeren VoID dokümanları (bkz. Şekil 5.1, Şekil 5.2, Şekil 5.8),
- Organizasyon verisini belirten adlı çizgeler (bkz. Şekil 5.3, Şekil 5.9),
- Organizasyonel tercihleri belirten adlı çizgeler (bkz. Şekil 5.4).

3.4 Adlı Çizgeler

Organizasyonel adlı çizgeler ve VoID köken üstverisini içeren adlı çizgeler, PEVoID ve organizasyonel ontolojilerde bulunan kavramlar ile örneklendirilirler. Organizasyon verisi ve VoID'e dair zenginleştirilmiş üstveri, çalışmada SPARQL kural çalıştırması için adlı çizgelere dönüştürülmektedir.

Sistemin çalışması esnasında organizasyonlar, tercihlerini VoID üstverisi ile ilişkili her gereksinime eşleştirir. Eşleştirilen adlı çizgelerin uygunluk durumuna göre filtreleme gerçekleştirilir. VoID köken üstverisini içeren adlı çizgelerin EBNF dilbilgisi Şekil 3.8'de gösterilmektedir.

```

PEVoIDAdlıÇizgesi ::= ÇizgeAdı '{'
    ÜçlüDeseni+
    VarlıkTanımlamaÜçlüsü
    HamveyaYayımlanmışVerisetineAitÜçlüÇizgesi+
'}'
ÜçlüDeseni ::= URI_Değişken
    URI_İfadesizDüğüm_Değişken
    URI_İfadesizDüğüm_Literal_Değişken '.'
VarlıkTanımlamaÜçlüsü ::= ÇizgeAdı ' rdf:type prov:Entity '
HamveyaYayımlanmışVerisetineAitÜçlüÇizgesi ::= ÇizgeAdı
    ' prov4fq:assessedBy ' KökenÇizgesi
ÇizgeAdı, KökenÇizgesi ::= URI
URI_Değişken ::= URI | Değişken
URI_İfadesizDüğüm_Değişken ::= URI | İfadesizDüğüm | Değişken
URI_İfadesizDüğüm_Literal_Değişken ::= URI | İfadesizDüğüm |
    Literal | Değişken

```

Şekil 3.8: VoID Köken Üstverisini İçeren Veriseti Tabanlı Adlı Çizgenin EBNF Dilbilgisi

Şekil 3.8’da gösterilen üçlü desenleri, VoID dokümanı içerisinde bulunan veriseti başlığı, tanımı, örnek kaynakları gibi temel elemanlar ile verisetinin yayımlayıcısı, verisetinin güncellenme zamanı gibi *dcterms* aduzayında ifade edilen özelliklerden meydana gelmektedir. Varlık tanımlama üçlüsü, Şekil 3.3’de gösterildiği üzere varlığı ifade etmektedir. Bu noktada varlık, VoID dokümanınınıdır. Ayrıca bu çizgede ham veriseti ve yayımlanmış verisetinin köken takibinin yapılması için gerekli adlı çizgeler de ifade edilir. Bu adlı çizgeler, ayrı ayrı oluşturulup Şekil 4.3’deki veri kaynağı katmanında işlenir.

```

DışVeriAdlıÇizgesi ::= ÇizgeAdı '{'
    ÜçlüDeseni+
    HamveyaYayımlanmışVerisetineAitKökenÜçlüsü+
'}'
ÜçlüDeseni ::= URI_Değişken
    URI_İfadesizDüğüm_Değişken
    URI_İfadesizDüğüm_Literal_Değişken '.'
HamveyaYayımlanmışVerisetineAitKökenÜçlüsü ::= URI_Değişken
    Köken_URI_İfadesizDüğüm_Değişken
    URI_İfadesizDüğüm_Literal_Değişken '.'
URI_Değişken ::= URI | Değişken
URI_İfadesizDüğüm_Değişken ::= URI | İfadesizDüğüm | Değişken
URI_İfadesizDüğüm_Literal_Değişken ::= URI | İfadesizDüğüm |
    Literal | Değişken
Köken_URI_İfadesizDüğüm_Değişken ::= 'pav:' URI | 'prov:' URI

```

Şekil 3.9: Ham veya Yayımlanmış Verisetine Ait Köken Adlı Çizgesinin EBNF dilbilgisi

Şekil 3.9’de bulunan ham veya yayımlanmış verisetine ait köken adlı çizgesi, Şekil 3.8’dekine benzer üçlü desenlerinin yanında verisetlerinin üstverisini detaylı şekilde ifade eden köken bilgisi içerir. Köken verisi, PROV-O veya PAV ile ifade edilen Şekil 3.3’deki özellikleri kapsamaktadır.

İç Veri ile ilgili adlı çizge tanımları Şekil 3.10’de ve Şekil 3.11’de görüldüğü üzere ifade edilmektedir.

```

İçVeriOrganizasyonAdlıÇizgesi ::= ÇizgeAdı '{' ÜçlüDeseni+ '}'
ÜçlüDeseni ::= URI_Değişken
               URI_İfadesizDüğüm_Değişken
               URI_İfadesizDüğüm_Literal_Değişken '.'
URI_Değişken ::= URI | Değişken
URI_İfadesizDüğüm_Değişken ::= URI | İfadesizDüğüm | Değişken
URI_İfadesizDüğüm_Literal_Değişken ::= URI | İfadesizDüğüm | Literal
                                         | Değişken

```

Şekil 3.10: Alana Dair Organizasyonel Adlı Çizgenin EBNF dilbilgisi

```

İçVeriTercihAdlıÇizgesi ::= ÇizgeAdı '{'
               TercihİfadeÜçlüsü
               KısıtİfadeÜçlüsü
               KısıtDeseni
               FiltrelemeKoşuluDeseni
               '}'
ÜçlüDeseni ::= URI_Değişken
               URI_İfadesizDüğüm_Değişken
               URI_İfadesizDüğüm_Literal_Değişken '.'
TercihİfadeÜçlüsü ::= URI_Değişken ' provpref:hasPreference ' URI_Değişken
KısıtİfadeÜçlüsü ::= URI_Değişken ' provpref:hasRestriction ' İfadesizDüğüm
KısıtDeseni ::= KısıtTanımlamaÜçlüsü
               KısıtElemanÜçlüsü
               KısıtFiltrelemeÜçlüsü
KısıtTanımlamaÜçlüsü ::= İfadesizDüğüm ' rdf:type provpref:Restriction . '
KısıtElemanÜçlüsü ::= URI_Değişken
               ( ' provpref:agent ' | ' provpref:organization '
               | ' provpref:entity ' )
               URI_Değişken
KısıtFiltrelemeÜçlüsü ::= İfadesizDüğüm ' provpref:filter ' URI_Değişken
FiltrelemeKoşuluDeseni ::= FiltrelemeKoşuluÜçlüsü ÜçlüDeseni+
FiltrelemeKoşuluÜçlüsü ::= URI_Değişken ' rdf:type provpref:FilterCondition .
URI_Değişken ::= URI | Değişken
URI_İfadesizDüğüm_Değişken ::= URI | İfadesizDüğüm | Değişken
URI_İfadesizDüğüm_Literal_Değişken ::= URI | İfadesizDüğüm | Literal | Değişken

```

Şekil 3.11: Tercih Adlı Çizgesinin EBNF dilbilgisi

Şekil 3.10 ve Şekil 3.11'deki adlı çizgeler, organizasyon verisi veya organizasyonel tercihleri temsil etmektedir. Organizasyon verisi, organizasyon içerisindeki çalışanlar ve çalışanların bulunduğu organizasyonel birimlere ait veri ile diğer organizasyonlara dair genel tanımları kapsar. Organizasyonel tercihler ise organizasyonun erişmek istediği veriseti için farklı köken boyutları gözönüne alarak oluşturulur.

Organizasyonlar, filtrelenmiş elemanları elde edebilmek için SPARQL kuralları işletirler. SPARQL kuralları Şekil 3.12'de görüldüğü üzere CONSTRUCT sorguları (Polleres, 2007) şeklinde kaydedilir. CONSTRUCTCümlesi, WHERECümlesi içerisinde bulunan tüm adlı çizge desenlerini kullanarak elde etmek istediği çizgeyi birleştirerek oluşturur. Dış Veriyi içeren (bkz. Şekil 3.8, Şekil 3.9) ve İç Veriyi içeren adlı çizgeler (bkz. Şekil 3.10, Şekil 3.11), Bölüm 5'de detaylandırıldığı üzere köken kurallarında ilklendirilir ve bu kurallar PROV4FQ tarafından işletilir. Şekil 3.12'de İç Veri ve Dış Veriye ait adlı çizgelerin sorgulanma biçimi görülmektedir.

```

KuralCümlesi ::= CONSTRUCTCümlesi WHERECümlesi
CONSTRUCTCümlesi ::= 'CONSTRUCT' '{'
    ÜçlüDeseni+
    '}'
WHERECümlesi ::= 'WHERE' '{'
    DışVeriAdlıÇizgesi*
    İçVeriOrganizasyonAdlıÇizgesi*
    İçVeriTercihAdlıÇizgesi* '}'

```

Şekil 3.12: Köken kuralının EBNF dilbilgisi

4 BAĞLI VERİ ALTYAPISINI GELİŞTİREN KÖKEN FARKINDA FİLTRELEME

4.1 Köken Farkında Filtreleme Süreci

Şekil 1.1’de ifade edildiği üzere global veri uzayına organizasyonel olarak bakıldığında organizasyonel çalışma alanı ve organizasyonel bağlı veri uzayı olarak bir ayırım yapılmaktadır. Bu ayrıma dair bileşenler Şekil 4.1’de detaylandırılmıştır. Tez çalışması köken farkında bağlı veri sorgulama altyapısı geliştirdiği için organizasyonel çalışma alanında sorgulama yapan bir bağlı veri uygulaması gösterilmiştir. Bu uygulama, bağlı veri uzayında ihtiyaç duyduğu servislere erişir ve organizasyonun istediği bilgiye dair sorguları çalıştırır. Organizasyonel çalışma alanının sahip olduğu veri, İç Veri olarak ifade edilir. Sorgular çalıştırılırken kökene ait kontroller sağlanması için İç Veri’den yararlanılır. İç Veri, organizasyon verisi, organizasyonel tercihler ve zenginleştirilmiş organizasyonel VoID ambarından meydana gelmektedir. Organizasyon verisi ve tercihler her organizasyonun kendine özgüdür. Organizasyonel VoID ambarında ise o organizasyonun sorgulayabileceği verisetlerinin üstverisinin zenginleştirilmiş hali bulunmaktadır ve bu üstveri zenginleştirilmiş global VoID ambarı tarafından sağlanır.

Bağlı veri uzayı, bağlı veri servisleri ve Dış Veri’den meydana gelmektedir. Çeşitli servisler içerisinde tez çalışmasında sorgulama servisi, veri değişimi izleme servisi ve köken farkında filtreleme servisi gözönüne alınmıştır. Sorgulama servisindeki dağıtık sorgu motoru, sorguları analiz eder ve sorgulayacağı verisetlerini bulur, analizden sonra elde edilen verisetlerine göre sorguyu tekrar organize eder ve tekrar organize edilmiş alt-sorguları Şekil 4.1’deki gibi farklı uçnoktalarına gönderir.

Sorguların gönderileceği uçnoktaların üstverisinin ve organizasyonel tercihlerin bulunduğu İç Veri’de zaman içerisinde değişiklikler olabilmektedir. Bu değişiklikler, organizasyonun varolan tercih ve bilgisini değiştirmesi, yeni organizasyon verisi eklemesi veya organizasyonu ilgilendiren global VoID dokümanlarının güncellenmesinden dolayı aynı değişikliğin İç Veri’de de yapılmasıdır.

Veri uzayı bakış açısı ile bu değişikliklerden veri seviyesinde haberdar olunabilmesi için bir servisin eklenmesi gerekliliği görülmektedir. Veri uzayı tanım olarak, organizasyonların sadece iş mantığına dair neler yapılabileceğini

et al., 2007) veri uzayı soyutlaması (Franklin et al., 2005; Halevy et al., 2006) ile benzerlikler göstermesinden dolayı ortam soyutlamasında ifade edilen A&A üstmodeli²² kullanılmıştır.

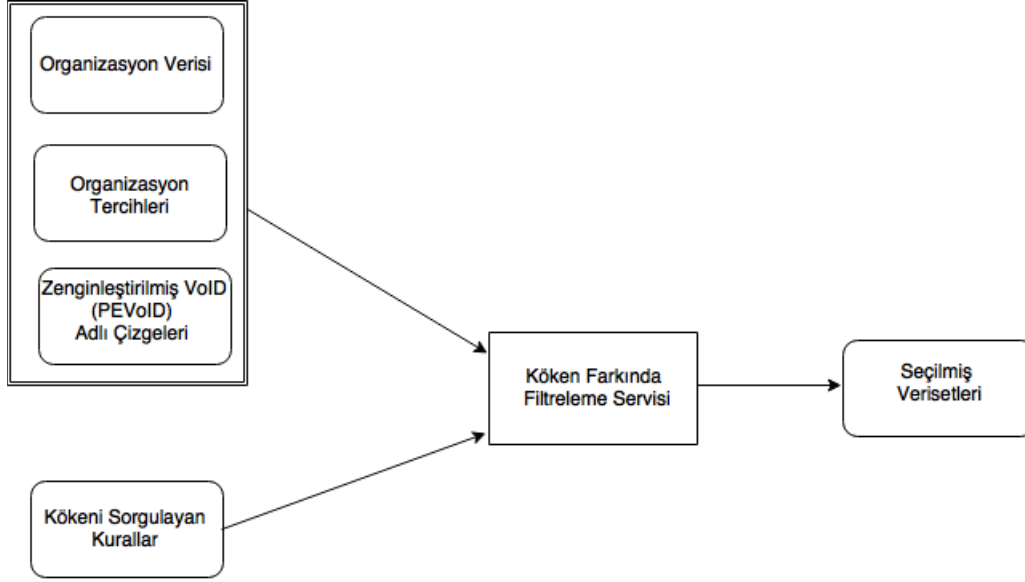
A&A üstmodeli ile geliştirilen gerçekleştirim modeli olan CArtAgO (Ricci et al., 2011) ile Erdur et. al. (Erdur et al., 2013) tarafından ifade edilen mimari oluşturulmuştur. Bu mimari veri izleme servisini ifade ederken çekme²³ ve itme²⁴ yaklaşımlarını beraberce kullanmayı benimsemiştir. Çekme yaklaşımında veri tüketicisi yayımlanan veriyi elde etmek için belirli aralıklarla sorgular atar. İtme yaklaşımında ise veri tüketicisi yayımlayıcıya abone olur. Yayımlanan veride değişiklik olduğunda meydana gelen değişiklikler abone olanlara iletilir. Tez çalışmasında organizasyon verisi ve tercihlerin izlenmesi için itme yaklaşımının, zenginleştirilmiş global VoID ambarına erişim için de çekme yaklaşımının uygun olduğu düşünülmüştür. Organizasyon verisi ve tercihlerin izlenmesi için her organizasyon ile ilgili SPARQL sorguları, izleme mekanizmasında kaydedilmekte ve bu sorgular izlenen artifaktlar üzerinde çalıştırılmaktadır. İzlenen anlamsal artifaktlardan gelen değişim bilgisi, ilgili sorgunun bilgi ve tercihler üzerinde çalıştırılması ile Veri İzleme Servisi'nden Köken Farkında Filtreleme Servisi'ne iletilir. Zenginleştirilmiş global VoID ambarındaki değişimin takibi için çekme yaklaşımının uygun olduğu düşünülmüştür. Çünkü organizasyonun ihtiyaç duyduğu üstveri dokümanları zaman içerisinde değişmekle beraber yeni VoID dokümanları da Veri Vebi'ne eklenmiş olabilir. Dolayısıyla belirli aralıklarla Veri Vebi'nin taranması için sorgulamanın yapılarak zenginleştirilmiş organizasyonel VoID ambarının güncellenmesi gerektiği görülmüştür. VoID dokümanları ile ilgili değişimler hakkında da Köken Farkında Filtreleme Servisi'ne haber verilmektedir.

Şekil 4.2'de gösterilen Köken Farkında Filtreleme Servisi ile veriseti seçimi aşamasında ilgili birden fazla veriseti bulunabilir veya verisetlerinin bazılarının organizasyonel tercihler ile değerlendirilerek filtrelenmesi gerekebilir. Veriseti filtrelemesi, analiz edilen aday verisetlerinin sorgu en iyileme safhasına geçmeden önce ele alınarak seçilmiş verisetlerinin elde edildiği süreçtir. Köken Farkında Filtreleme Servisi, zenginleştirilmiş organizasyonel VoID ambarında bulunan dokümanlar ile organizasyon verisi ve tercihleri kullanarak filtrelenmiş organizasyonel VoID ambarında organizasyona özgü kullanılabilir veri kümelerini oluşturur. Böylece veri uzayı yaklaşımına uygun bir şekilde bağımsız ve birbirinden faydalanan servislerin meydana getirdiği organizasyonel bağlı veri uzayı ifade edilmiştir.

²²Açılımı: Agents&Artifacts Meta-Model

²³İngilizcesi: pull

²⁴İngilizcesi: push



Şekil 4.2: Köken Farkında Filtreleme Süreci

Tez çalışmasında kullanılan dağıtık sorgulama yaklaşımı, veriseti seçimi adımıyla aday verisetlerinin seçimi için farklı veri kaynaklarının üstverisinden yararlanır. Bir sorgu, herhangi bir dağıtık sorgu motoruna gönderilmeden önce veriseti seçiminden sonra aday verisetleri elde edilir. Aday verisetleri, bağlı veri bulutundan gelen analiz edilmiş verisetleri ve bağsetlerinin listesidir. Seçilmiş verisetleri, organizasyonel tercihler, organizasyon verisi ve verisetinin içerdiği üstveri düşünülerek filtrelenmiş verisetleri ve bağsetlerinin listesidir. Seçilmiş verisetlerinin elde edilmesi için aday veriseti kümesi ile filtrelenmiş organizasyonel VoID ambarındaki üstveri kümesi karşılaştırılır. Filtrelenmiş organizasyonel VoID ambarında da bulunan aday verisetleri, seçilmiş verisetleri olarak Sorgu En İyileme aşamasına aktarılır ve sorgulama operasyonu devam eder.

Köken farkında filtreleme servisi, seçilmiş verisetlerini elde etmek için zenginleştirilmiş organizasyonel VoID ambarındaki üstveri ile beraber organizasyon verisi ve organizasyonel tercihlerde bulunan adlı çizgeleri organizasyonel kural tanımlamalarına atar. Servis, uygun verilerin atandığı kural tabanlı operasyonları gerçekleştirir. İlgili veride meydana gelen değişimler sonucunda kural tabanlı operasyonlar da güncellenen veri ile tekrar çalıştırılmaktadır. Böylece filtrelenmiş organizasyonel VoID ambarının sürekli güncel tutulması sağlanır.

Dış Veri, Veri Vebi'ndeki tüm verisetlerinin üstverisini ifade eden birçok global VoID ambarındaki üstveri dokümanları ve bahsedilen verisetlerindeki veriyi sağlayan RDF ambarlarından meydana gelmektedir. Sorgulama servisinin ana elemanı olan dağıtık sorgu motoru, seçilmiş verisetlerini sorgu en iyileme adımıyla ele alır ve alt-sorguları SPARQL uç noktaları üzerinde çalıştırır. Şekil 2.4'de

görüldüğü üzere Veri Erişim, Bütünleştirme ve Depolama gibi farklı aşamalardan geçerek SPARQL uçnoktaları yardımı ile Dış Veri'den veri çizgeleri elde eder. Sorgulama sonucunda elde edilen çıktı, organizasyonel çalışma alanına iletilir.

Bahsedilen yaklaşım ile bağlı verisetlerini filtreleyebilen bir köken çözümü sunulmakla beraber aslında farklı servisleri ifade eden bir bağlı veri uzayı yaklaşımı geliştirilmiştir. Böylece zaman içerisinde geliştirilen farklı servislerin de uyarlanabileceği bir çözüm altyapısı sunulmuştur.

4.2 Köken Farkında Bağlı Veri Sorgulama Yazılım Mimarisi

Köken farkında dağıtık sorgu altyapısının yazılım mimarisi, mimari belgeleme stilleri kullanarak modül görünüm belgelemesi²⁵ içerisindeki katmanlı biçim²⁶ ile çizilmiştir (Clements et al., 2002). Clements et. al. (Clements et al., 2002)'ın tanımladığı modül biçimlerinden tez çalışması dahilinde ayrışma²⁷, kullanma²⁸ ve katmanlı biçim yazılım mimari biçimleri kullanılmıştır. Ayrışma biçimi, temelde modüller arasındaki içermeye ilişkilerini gösterir. Kullanma biçimi, modüller arasındaki fonksiyonel bağımlılıkları ifade eder. Katmanlı biçim, modül grupları (katmanlar) arasındaki ilişkiler ile ilgilenir. Clements et. al. (Clements et al., 2002), tez çalışmasındaki Veri Kaynağı Katmanı ve İç Veri Katmanı'nda bulunan elemanların gösterimi için herhangi bir simgeleme içermemektedir. Bu nedenle, ontoloji için yuvarlak köşeli dikdörtgen, adlı çizge için altıgen, kural tanımı için paralelkenar ve veri ambarları için silindirik kullanımı Şekil 4.3'de görüldüğü üzere eklenmiştir.

Köken farkında filtreleme servisini ifade eden PROV4FQ yazılım mimarisi ise iki katmanda incelenmektedir. Bu katmanlar, Köken Uygulama Katmanı ve Veri Kaynağı Katmanı olarak adlandırılmıştır. İç Veri Katmanı ise sorgulama servisi ve köken farkında filtreleme servisinin tükettiği (veya veri sağladığı), organizasyon dahilinde bulunan veriyi ifade eder.

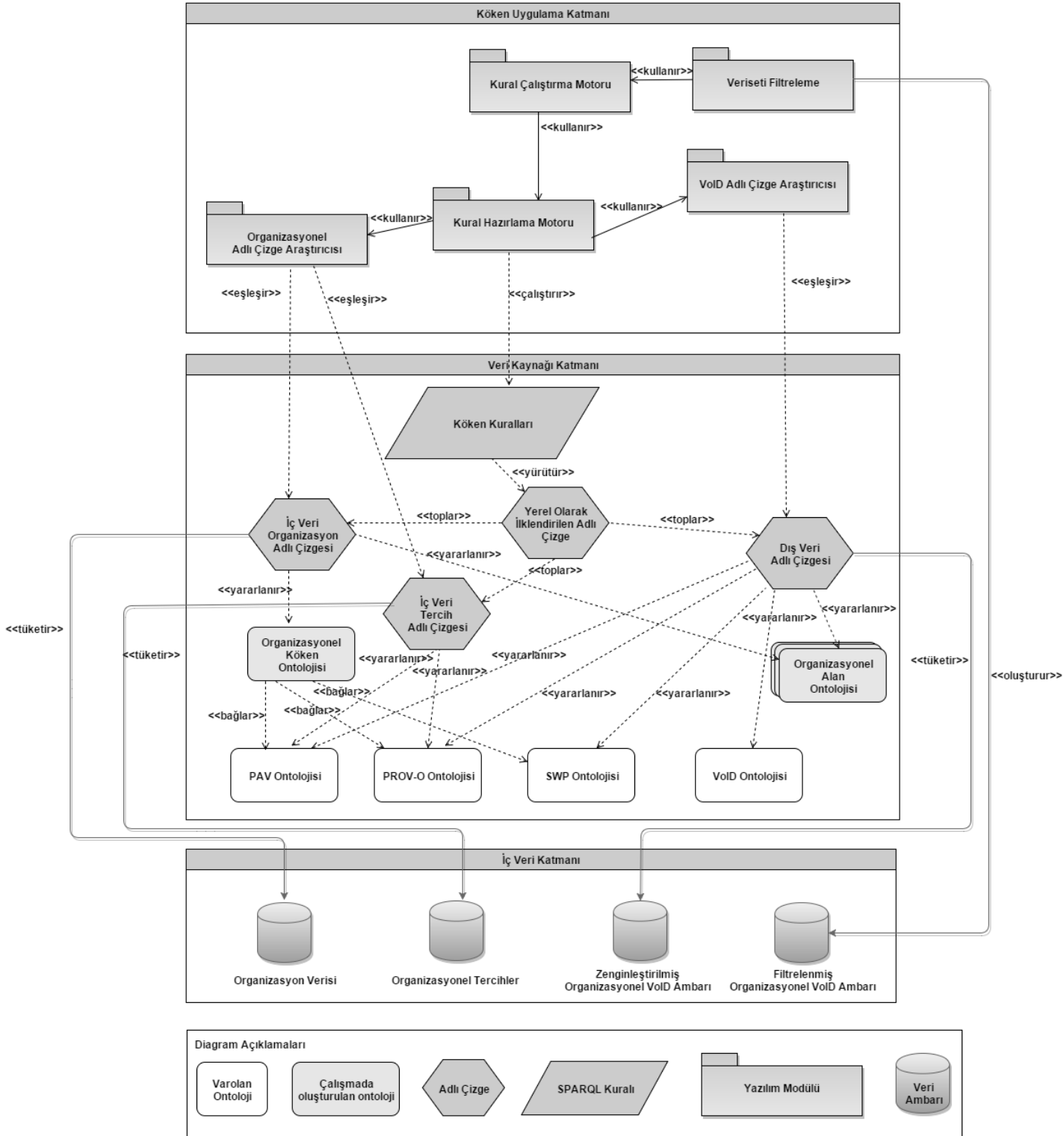
Şekil 4.3'de gösterilen modüller, köken farkında dağıtık sorgulama yazılım mimarisinin operasyonel bölümleridir. *kullanır* ilişkisi iki modülü birbirine bağlar ve ok yönündeki çıktıyı döndürür.

²⁵ İngilizcesi: Module View Documentation

²⁶ İngilizcesi: Layered Style

²⁷ İngilizcesi: Decomposition

²⁸ İngilizcesi: Use



Şekil 4.3: PROV4FQ Yazılım Mimarisi

Köken Uygulama Katmanı, köken farkında veriseti seçiminden sorumludur. *eşleşir* ilişkisi modül ve adlı çizge arasındaki bağlantıyı tanımlar. Modülün eşleştirdiği talep edilen adlı çizgeleri ok yönünde çıktı olarak iletir. *çalıştırır* ilişkisi, ilgili kuralları çalıştırmak için modül ve kurallar arasında bir bağlantı oluşturur.

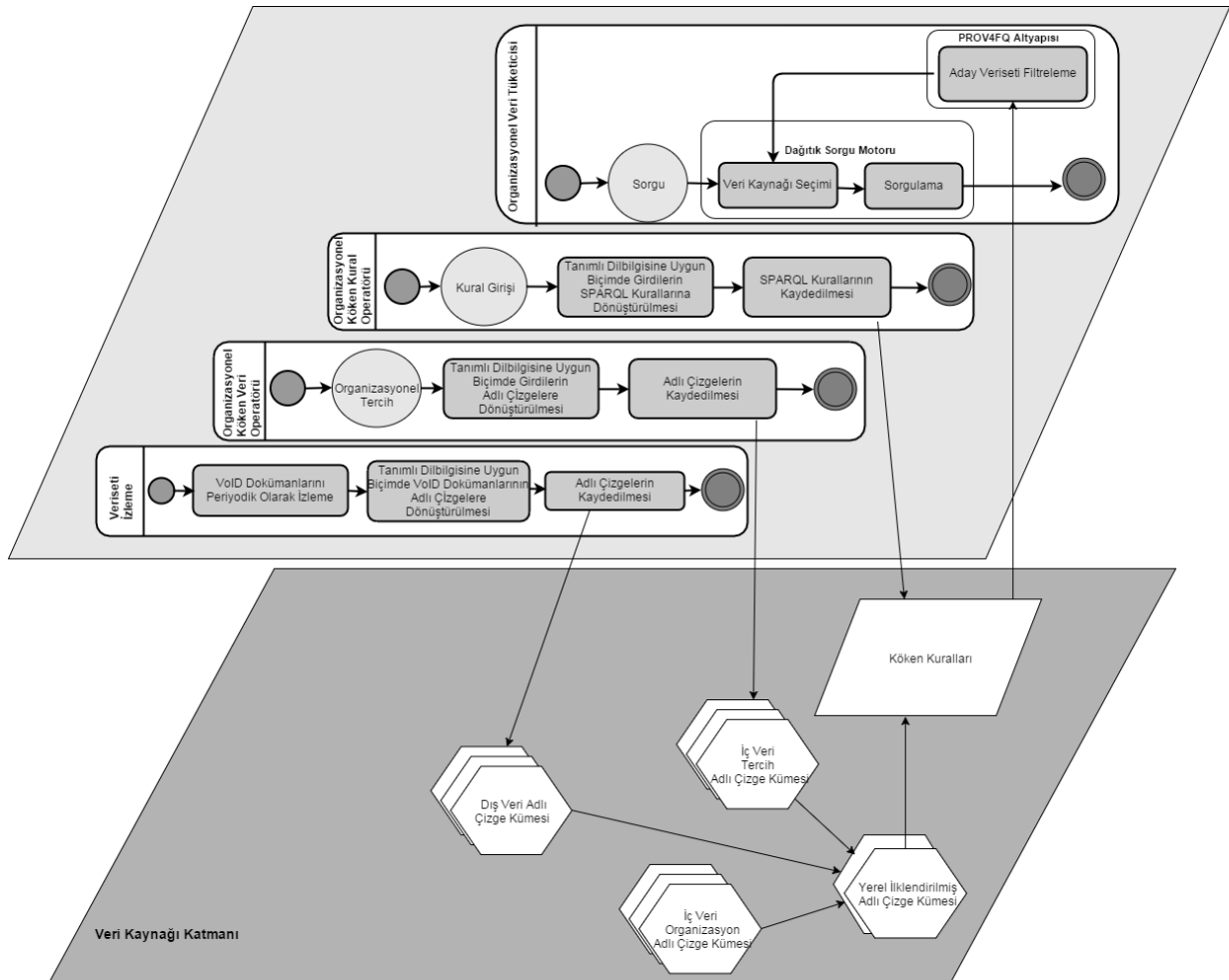
Veri Kaynağı Katmanı, SPARQL kuralları, adlı çizgeler, çalışmada oluşturulan ontolojiler ve varolan ontolojilerden meydana gelir. *yararlanır* ilişkisi ontolojileri adlı çizgelere bağlar. *bağlar* ilişkisi, var olan ontolojiler ile kullanıcı-tanımlı ontolojiler arasındaki bağlantıyı temsil ederken *toplar* adlı çizge kümesinin adlı çizgelerden oluştuğunu gösterir.

İç Veri Katmanı, Şekil 4.1 anlatılırken de ifade edildiği gibi organizasyona ait veriyi gösterir. *tüketir* ilişkisi adlı çizgelerin organizasyon verisi veya zenginleştirilmiş VoID ambarına erişerek veri kullanımını gösterir. *oluşturur* ise filtreleme sürecinin sonunda oluşan filtrelenmiş veriseti verisinin organizasyonel VoID ambarına ilgili yazılım modülü tarafından iletildiği ilişkiyi ifade eder.

Köken Uygulama Katmanı, filtrelenmiş verisetlerini oluşturmaktan sorumludur. VoID Adlı Çizge Araştırmacı, VoID dokümanlarında bulunan boyutlar ile ilgili adlı çizgeleri keşfeder. Dış Veri Adlı Çizgeleri, yerel olarak ilklendirilmiş adlı çizge kümesine eklenir. Organizasyonel Adlı Çizge Araştırmacı, organizasyona ait tüm tercihleri içeren adlı çizgeleri inceler. Organizasyonel Adlı Çizge Araştırmacı, İç Veri Organizasyon ve Tercih Adlı Çizgeleri'nden ilgili olanları kullanarak yerel adlı çizge kümesini günceller. Tez çalışmasındaki kurallar SPARQL CONSTRUCT sorguları şeklinde oluşturulmuştur (Polleres, 2007) ve seçilen verisetleri hakkında bilgi içeren bir çizge döndürürler. Kural Hazırlama Motoru, Köken Kuralları'nı ele alır. Daha sonra yerel adlı çizge kümesindeki elemanları (Dış Veri Adlı Çizgeleri, İç Veri Organizasyon Adlı Çizgeleri ve İç Veri Tercih Adlı Çizgeleri) bu kurallarda yerine koyar. Böylece İç Veri ve Dış Veri eşleştirilmeye çalışılır. Eğer uygun bir eşleştirme var ise kural, Kural Çalıştırma Motoru'nda çalıştırılır ve veriseti bilgileri Veriseti Filtreleme Modülü'ne iletilir. Veriseti Filtreleme Modülü, seçilen verisetlerini sorgu çalıştırmanın ileriki adımlarında kullanılması için Filtrelenmiş Organizasyonel VoID Ambarı'na geçirir. Bu süreç, Veri İzleme Servisi'nden gelen etki ile Zenginleştirilmiş Organizasyonel VoID Ambarı'nda değişime neden olan anlarda çalışır. Böylece, Filtrelenmiş Organizasyonel VoID Ambarı'nın güncel olarak organizasyondaki veri tüketicileri tarafından sürekli kullanılması ve kaliteli veriye erişimi sağlanır. Her sorgu çalıştırıldığında Dağıtık Sorgu Motoru'nun veriseti analizi safhasından sonra kullanılacak verisetleri ile Filtrelenmiş Organizasyonel VoID Ambarı'ndaki veri karşılaştırılarak eşleşenler seçilir ve Sorgu En İyileme safhasına geçilir (bkz. Şekil 4.1).

4.3 İş Akışı

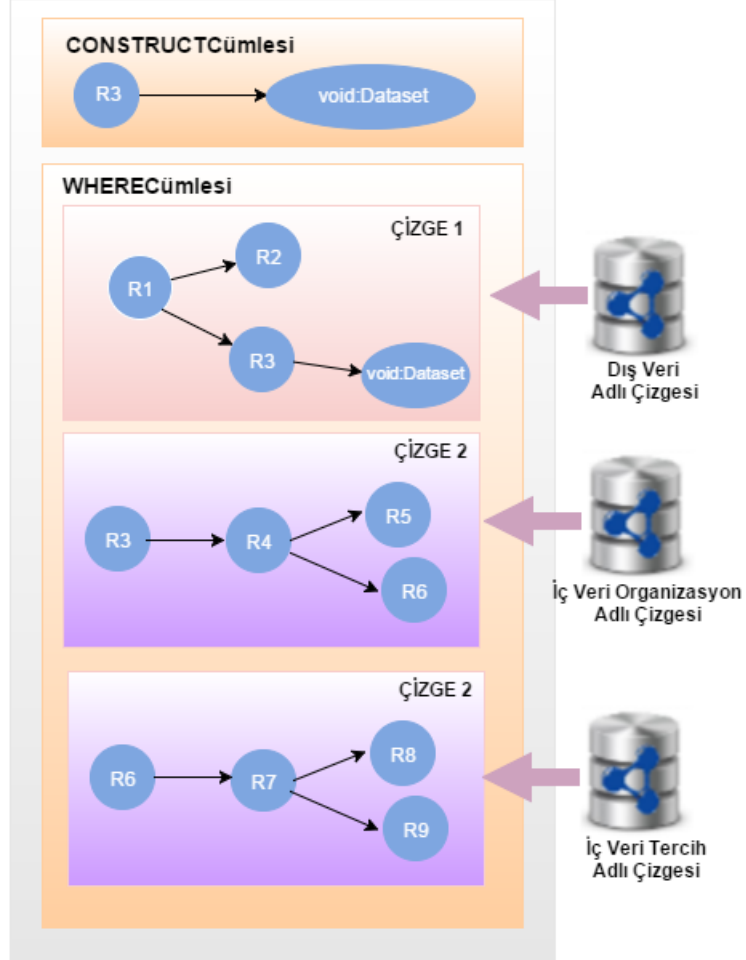
PROV4FQ, VoID-tabanlı köken üstverisi, organizasyon verisi ve organizasyonel tercihleri içeren kuralları işleterek aday verisetlerini filtreleyen bir altyapıdır. Aday verisetlerinin filtrelendiği PROV4FQ altyapısının iş akışı veriseti-tabanlı operasyonlar (güncellenen verisetlerinin takibi), organizasyon-tabanlı operasyonlar (organizasyonel ve köken bilgisinin girişi/güncellenmesi) ve verisetlerinin filtrelenmesi için veri tüketicisinin (organizasyonun) attığı sorguya dair adlı çizgelerin sorgulanmasına dair operasyonlardan meydana gelmektedir. Şekil 4.4’de ifade edilen Organizasyonel Veri Tüketicisi, bağlı veri bulutunda yapacağı sorguya dair en uygun cevapları elde etmek istemektedir. Dağıtık sorgulama esnasında ne yapıldığını önemsemeyen talep ettiği sorguya dair sonuçları alır.



Şekil 4.4: PROV4FQ İçerisinde SPARQL CONSTRUCT Sorgusunun İlkendirilmesi

Bir organizasyondaki yeni veya değişen tercihler, köken hakkındaki kural içeriğinin de değişimine neden olabilmektedir. Şekil 4.4’de görüldüğü üzere

Organizasyonel Köken Kural Operatörü, sorgular hakkındaki kural detaylarını girmekten sorumludur. Operatörün veri modelinin yapısını detaylı olarak bilmeden girdiği kurallar, SPARQL kural yapısına dönüştürülerek organizasyonda saklanmaktadır.



Şekil 4.5: SPARQL CONSTRUCT Sorgusunun İlkendirilmesi

Organizasyonel Köken Veri Operatörü, organizasyonun tercihlerini girmekle yükümlüdür. Girilen veri, dilbilgisine uygun biçimde adlı çizgelere dönüştürülmekte ve depolanmaktadır (bkz. Şekil 3.9). Depolanan bilgi sorgulanmak istendiğinde PROV4FQ tarafından oluşturulan Yerel İlkendirilmiş Adlı Çizge Kümesi'ne eklenir. Köken Kuralları'na erişilip bu kurallardaki ilgili kısımlara Yerel İlkendirilmiş Adlı Çizge Kümesi'ndeki adlı çizgeler yerleştirilir. Şekil 3.12'de gösterilen köken kuralları üzerinde Şekil 4.5'deki gibi adlı çizgeler koyularak filtrelemeler gerçekleştirilmektedir.

Anlamsal veb sorgulama yaklaşımları, merkezi ve dağıtık olarak sınıflandırılmaktadır. Merkezi sorgulama, merkezi bir veri ambarından sorgulanan bağlı veriyi birleştirme prensibinden meydana gelmektedir. Bu yaklaşımın en

büyük sorunu güncel verinin gösterimini yapamamaktır. Bu sorundan kurtulmak için Veriseti İzleyicisi²⁹ (Erdur et al., 2013) iki yaklaşımı birleştirmektedir. Organizasyonlar, ilişkili dağıtık VoID-tabanlı Köken Üstveri Ambarlarında³⁰ bulunan adlı çizgelerdeki değişiklikleri fark eder ve VoID dokümanlarını kendi ambarında günceller. Bu şekilde Veriseti İzleme aşamasında dağıtık veri hakkındaki üstveri merkezi olarak güncel şekilde depolanır.

Organizasyonel kurallar çalıştırılmadan önce ilgili İç Veri ve Dış Veri, Yerel İlkendirilmiş Adlı Çizge Kümesi'nde bulunmalıdır. Çünkü kurallar bu kümede bulunan çizgeler üzerinden işletilir. İlkendirme işlemi, SPARQL sorguları üzerinde (bkz. Şekil 3.12) Şekil 4.5'deki gibi yapılmaktadır.

²⁹*İngilizcesi:* Monitoring Element

³⁰*İngilizcesi:* VOID-based Provenance Metadata Store

5 DURUM ÇALIŞMASI

Günlük hayatta karşılaşılan durumlarda kişisel filtrelemeler yapılmaktadır. Örneğin, takip edilen dergilerde sadece çalışılan alanlarda makaleler yayınlandığında veya ilgilenilen alana dair özel bir sayı çıktığında akademisyenler bilgilendirilmek ister. Başka bir örnek ele alınırsa, yabancı haber kaynaklarında yayımlanan haberler farklı güvenilir kaynaklardan yayımlanıyorsa itibar edilip haberler incelenir. Bahsedilen ve benzeri durumlar gerçek hayatta bireysel olarak takip edilirken bağlı veri bulutunda ise çizgelerin filtrelenmesi şeklinde ele alınmaktadır. Tez içerisinde literatürde kabul görmüş olan Haber Toplama Senaryosu'nun köken farkında bir altyapı kullanılarak bağlı veriyi filtreleme amacı ile kullanılması düşünülmüştür.

5.1 Durum Çalışmasında Köken İhtiyacının Ele Alınması

Köken için ihtiyaç duyulan kavramların ifade edilebilmesi amacı ile 2010 yılında W3C Köken Kuluçka Grubu (Cheney et al., 2010b) tarafından üç adet ana senaryo (Cheney et al., 2010a) üzerinde durulmuştur. Haber Toplama Senaryosu, Hastalık Salgın Senaryosu ve İş Sözleşme Senaryosu adındaki bu senaryolar arasından Haber Toplama Senaryosu seçilmiştir. Bu senaryo farklı kaynaklardan aldığı haber elemanlarını (haber siteleri, ağ günceleri ve haberle ilgili resimler gibi) birleştirme amacını gütmektedir.

Çoğu veb kullanıcısı bir veb dokümanı veya kaynağın kullanımının uygunluğuna dair mekanizmalara ihtiyaç duymaktadır. İçeriğin orjinal kaynağı, kaynakla ilgili lisans bilgisi ve içeriğe dair kullanım kısıtları bahsedilen ihtiyaçlar arasındadır. Toplanan içerik düşünüldüğünde içeriği oluşturan süreçlerin yaratım ve iletimi ile ilgili güven kıstasları da gözönüne alınmalıdır. Güven ile alakalı olarak bir organizasyon, olayın oluştuğu ülkedeki habere ait farklı haber kaynaklarından güven üstverisine sahip olabilir. Dolayısıyla bu organizasyon, sadece güvenerek sorgulayabileceği haber siteleri(veya haber kaynakları)ndeki veriye erişen bir filtreleme mekanizması kurulmasını talep edebilir.

Haberin güvenli bir kaynaktan gelmesi veya eğer kaynak zinciri şeklinde (Twitter'da retweet yapılması gibi) bir yapı varsa kök kaynağın güvenilir olması gibi durumlar veri tüketicisi (organizasyon) açısından önemlidir. Bunun yanında veri tüketicisi de güvendiği kaynakları belirleyebilir. Her iki durumda da veri kalitesinin artmasına dair farklı bakış açılarına göre çözümlemeler yapılmaya çalışılmıştır (Zaveri et al., 2015). Bakış açılarından güven ele alındığında eğer bir haber iki veya

daha fazla haber kaynağı (veya otoritesi) tarafından güvenli olarak ifade ediliyorsa bu haber talepte bulunan kişiye veya organizasyona ulaştırılabilir. Bunun yanı sıra Veri Dağıtımı³¹ ile ilgili durumlarla da karşılaşılabilmektedir. Örneğin, Associated Press haber ajansı tarafından sağlanan bir haber eğer New York Times'ta yayımlandı ise New York Times'a güvenen kişi veya organizasyonlar bu haberin kökeninde güvenilir kaynaklar olduğu için habere itibar eder. Bahsedilen örnekte Anlamsal Veb Yayımlama Sözlüğü³²'nde (SWP) bulunan iddia edilme³³ (veriyi yayımlayan kaynağın ana kaynak olması) ve alıntı yapma³⁴ (yayımlayıcının başka bir kaynaktan aldığı veriyi yayımlaması) gibi durumlar ele alındığından tez çalışmasına yardımcı olması açısından bu sözlük kullanılmıştır.

5.2 Güven Senaryo Adımları

5.2.1 Dış veri tabanlı operasyonlar

Çalışmada veriseti tabanlı operasyonlarda kullanılan VoID üstverisinin adlı çizgeye dönüştürüldüğü ve güvene dair bilgilerin eklendiği doküman Şekil 5.1'de görülmektedir. Dokümanda, alan modeline uygun olarak eklenen VoID elemanının *prov:Entity* tipinde olduğu fark edilmektedir. Böylece PROV-O kullanılarak üzerinde sorgulama yapılabilecek bir varlık olduğu anlaşılmaktadır. *prov4fq:assessedBy* tanımlaması ise bu VoID dokümanının bir adlı çizge ile kalite ölçütüne tabi tutulacak şekilde sorgulanacağını ifade etmektedir. Örnekte, obje değeri olarak *provreqgraph:RawNYTDatasetNG* olmakla beraber gereksinimlere göre eklenecek tüm adlı çizge URI referansları aynı şekilde sorgulama ve filtreleme mekanizmasına girecektir. Bahsedilen adlı çizge, VoID dokümanı ile beraber yayımlanabileceği gibi bu dokümandaki URI sayesinde PROV-AQ (Moreau et al., 2012) sorguları ile birlikte farklı bir veri deposundan da erişilebilir. Tez çalışmasında adlı çizgeye VoID dokümanı ile beraber erişilmektedir. Adlı çizgeye dair detaylar Şekil 5.2'de gösterilmiştir.

Şekil 5.2'de ifade edilen *provreqgraph:RawNYTDatasetNG* güven gereksinimini(kalite boyutunu) karşılamaktadır. VoID üstverisi, tanımlanan verisetine dair hangi otorite ile bağlantılı sorgulamalar yapılabileceğini içermektedir. Örnekte, NYT verisetine dair otoritenin BBC olduğu ile ilgili New York Times'ın VoID dokümanı içerisinde bir adlı çizge meydana getirilmiştir. Atılan sorgularda

³¹İngilizcesi: Data Syndication

³²İngilizcesi: Semantic Web Publishing Vocabulary

³³İngilizcesi: asserted by

³⁴İngilizcesi: quoted by

```

<file://nytprov4fqvoid.trig> {
  <http://data.ege.edu.tr/endpoint/repositories/nytprov4fqvoid>
    a void:Dataset ;
    dcterms:modified "2014-06-22"^^xsd:date ;
    dcterms:rightsHolder "The New York Times Company" ;
    dcterms:created "2014-08-20"^^xsd:date ;
    dc:creator "The New York Times Company" ;
    cc:license <http://creativecommons.org/licenses/by/3.0/us/> ;
    cc:attributionURL <http://data.nytimes.com/
                          47452218948077706853/> ;
    cc:attributionName "The New York Times Company" ;
    pav:version "1.0" ;
    a prov:Entity ;
    prov4fq:assessedBy provreqgraph:RawNYTDataSetNG .
    prov4fq:assessedBy provreqgraph:LinkedNYTDataSetNG .
}

```

Şekil 5.1: NYT verisetine dair VoID dokümanındaki adlı çizgelerin ifade edilmesi

adlı çizgelerin her ikisinde de aynı otorite bilgisinin olup olmayacağı kontrol edilmektedir. Organizasyonlar kendi ihtiyaçlarına uygun olarak farklı sorgular oluşturabileceklerinden dolayı köken gereksinimlerine dair adlı çizgelerin VoID içerisinde mümkün olduğunca fazla olması o verisetinin seçilmesi için avantaj sağlamaktadır. Çünkü organizasyon, kökenine dair hiçbir bilgiye sahip olmadığı verisetlerini doğrudan kullanmayacağını belirten kurallar tanımlamış olabilir. Bu durumda veriseti ne kadar güncel ve yararlı bilgi içerirse içersin ihtiyacı olan kişi veya organizasyonlara verisini ulaştıramayacaktır.

```

provreqgraph:RawNYTDataSetNG {
  <http://data.ege.edu.tr/endpoint/repositories/nytprov4fqvoid>
    swp:assertedBy provreqgraph:NytWarrant ;
    dcterms:title "New York Times dataset"@en ;
    dcterms:publisher <http://www.nytimes.com/> ;
    dcterms:issued "2015-01-01T00:00:00T"^^xsd:dateTime ;
    pav:version "1.0" ;
    foaf:page <http://www.nytimes.com> ;
    pav:createdBy orgs:NYT.
    provreqgraph:NytWarrant
      swp:authority orgs:BBC;
      dc:date "2005-10-21".
}

```

Şekil 5.2: VOID dokümanı içerisindeki güvene dair ifadeler

5.2.2 İç veri tabanlı operasyonlar

Şekil 5.3'deki örnekte farklı organizasyonların temel bilgilerini içeren bir adlı çizge bulunmaktadır. Bu çizgede gösterilen BBC Haber Ajansı, Şekil 5.2'deki VoID içerisinde bulunan güvene dair adlı çizge ile eşleşmektedir. Otoriteye dair bilgilerin (otoritenin ülkesi gibi) de organizasyon tarafından tutulması, Dış Veri tabanlı operasyonlardan gelen adlı çizgeler ile İç Veriye ait adlı çizgelerin eşleştirilmesini sağlamaktadır.

```
@prefix od: <http://ege.edu.tr/seagent/ont/org/od#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov4org: <http://ege.edu.tr/seagent/ont/prov4org#> .
@prefix prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#> .
@prefix orgs: <http://ege.edu.tr/seagent/ont/orgs#>.
@prefix prov: <http://www.w3.org/ns/prov#>.
@prefix iso: <http://www.daml.org/2001/09/countries/iso-3166-ont#> .

od:OrganizationalDomainInformation{
  orgs:ConsumerOrg rdf:type prov4org:Organization .

  orgs:BBC rdf:type prov4newsorg:NewsAgency;
            prov4newsorg:country iso:US;
            foaf:name "BBC".

  orgs:AssociatedPress rdf:type prov4newsorg:NewsAgency;
            foaf:name "Associated Press".

  orgs:NYT rdf:type prov4newsorg:NewsAgency;
            foaf:name "The New York Times";
            foaf:page <http://www.nytimes.com> ;
            prov4org:authority orgs:BBC.
}
```

Şekil 5.3: Organizasyonun güven tanımlamaları içeren alana dair adlı çizgesi

Organizasyonun tercihlerinin de Dış Veri'yi elde etme esnasında gözönüne alınması gerekmektedir. Şekil 5.4'deki örnekte NYT'nin sadece Amerika'daki haber ajanslarından gelen veriyi almasına dair bir tercih bulunmaktadır.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix prov4pref: <http://ege.edu.tr/seagent/ont/provpref#> .
@prefix prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#> .
@prefix orgs: <http://ege.edu.tr/seagent/ont/orgs#>.
@prefix prov: <http://www.w3.org/ns/prov#>.
@prefix iso: <http://www.daml.org/2001/09/countries/iso-3166-ont#> .
@prefix ex: <http://ege.edu.tr/seagent/ont/org/ex#> .

ex:NYTPreferences{
  orgs:NYT prov4pref:hasOrgPreference ex:p1.

  ex:p1 rdf:type prov4pref:Preference .

  ex:p1 prov4pref:hasRestriction ex:r1 .

  ex:r1 rdf:type prov4pref:Restriction;
    prov4pref:organization prov4newsorg:NewsAgency;
    prov4pref:filter _:pfc1 .

  _:pfc1 rdf:type prov4pref:FilterCondition;
    prov4newsorg:country iso:US.
}

```

Şekil 5.4: Organizasyonun tercihlerine dair adlı çizge

5.2.3 Adlı çizgelerin sorgulanması

Şekil 5.5’de gösterilen örnekte Amerika’daki bir Haber Ajansını otorite olarak kabul eden verisetlerinin çizgesi elde edilmek istenmektedir. Bunun yanında VoID dokümanını yaratan kuruluş ile ham veriseti kuruluşunun da aynı olduğunun PAV ontolojisi kullanılarak bulunmaya çalışılmaktadır. Sorgu çalıştırılması esnasında Şekil 5.5’de bulunan *?graph* içerisindeki *?authority* tanımlaması *?graph2* içerisindeki aynı tanıma bağlanarak çizgeler eşleştirilmeye çalışılır. Eğer Şekil 5.6’de görüleceği gibi bir eşleştirme gerçekleşirse dönen üçlü örgüleri Filtrelenmiş Organizasyonel VoID Ambarı’na iletilir (bkz. Şekil 4.1).

Şekil 5.6’de Dış Veri ve İç Veri’yi içeren adlı çizgeler güven senaryosu için örneklendirilmiştir. Şekilde bulunan ilk adlı çizge NYT’ye ait VoID dokümanının temel bazı özellikleri ile beraber bir güven çizgesi tarafından değerlendirildiğini ifade etmektedir. İkinci adlı çizge, ilgili güven çizgesindeki üçlülere göstermektedir. Bu çizgede ifade edilen organizasyon örneği, organizasyonel alan bilgisini içeren üçüncü adlı çizge içindeki organizasyonlarda detaylandırılmıştır. Dördüncü adlı

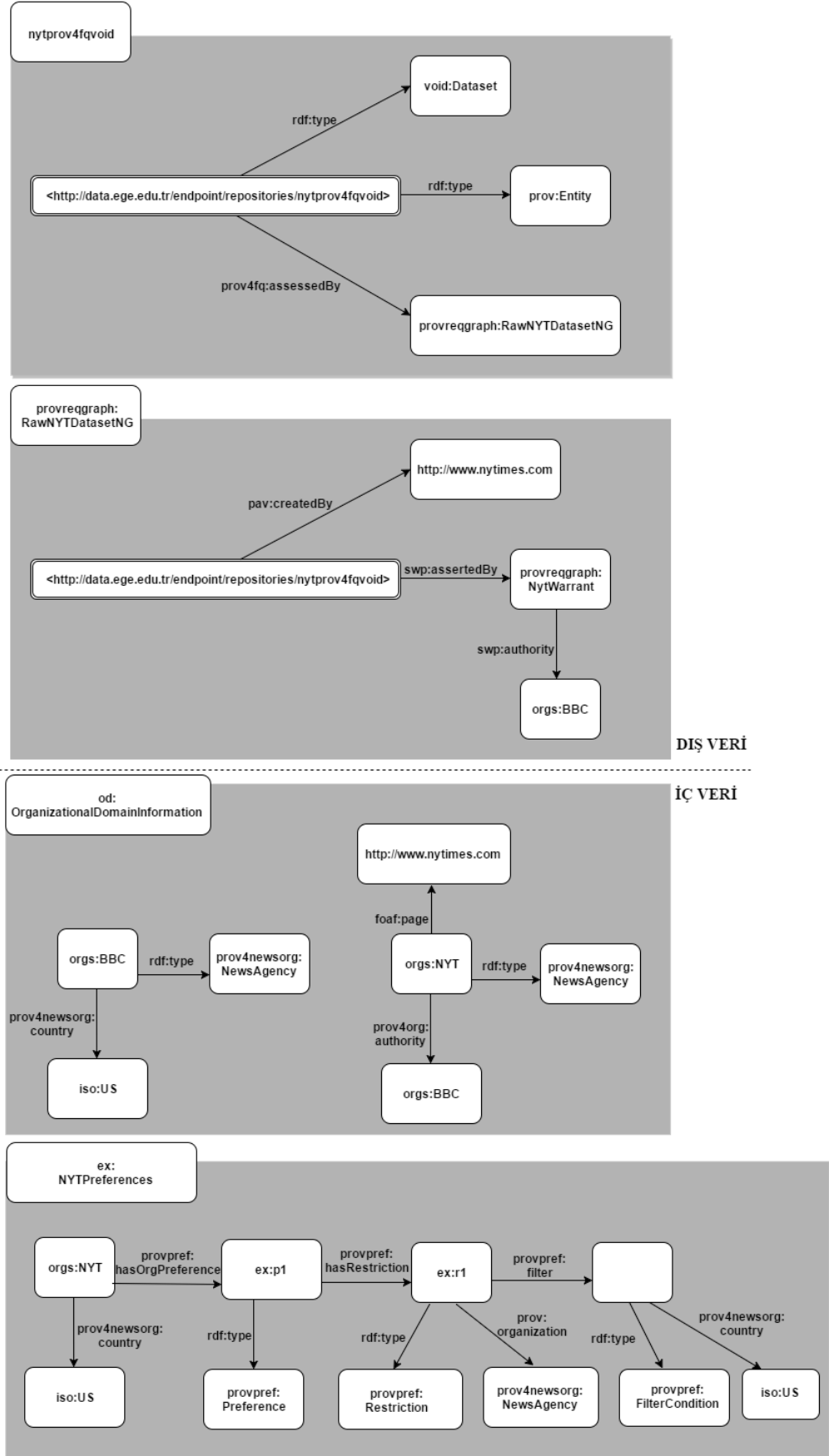
çizge ise organizasyonun tercihlerini göstermektedir.

Şekil 5.7’de bir sorgu üzerinden sistemdeki veri akışı incelenmektedir. Tez çalışmasında hedef aday verisetlerinin filtrelenerek seçilmiş verisetlerinin sorgulama esnasında kullanılır şekilde tutulmasıdır. Bunun için her aday veriseti kökene dair organizasyonun tercihlerine ve kurallara uygun şekilde işlenir. Bu işlem sonucunda kullanılacak verisetleri (VoID içerisindeki tanımı ile uçnoktasını) filtrelenerek Birleştirilmiş Sorgu Motoru (WoDQA)’na ihtiyaç dahilinde kullanılmak üzere sunulur. Bundan sonraki aşamalarda Birleştirilmiş Sorgu Motoru, kendi çözümlediği verisetlerinin arasından filtrelenmiş verisetlerini kullanarak tüm uçnoktalarına sorgu atar.

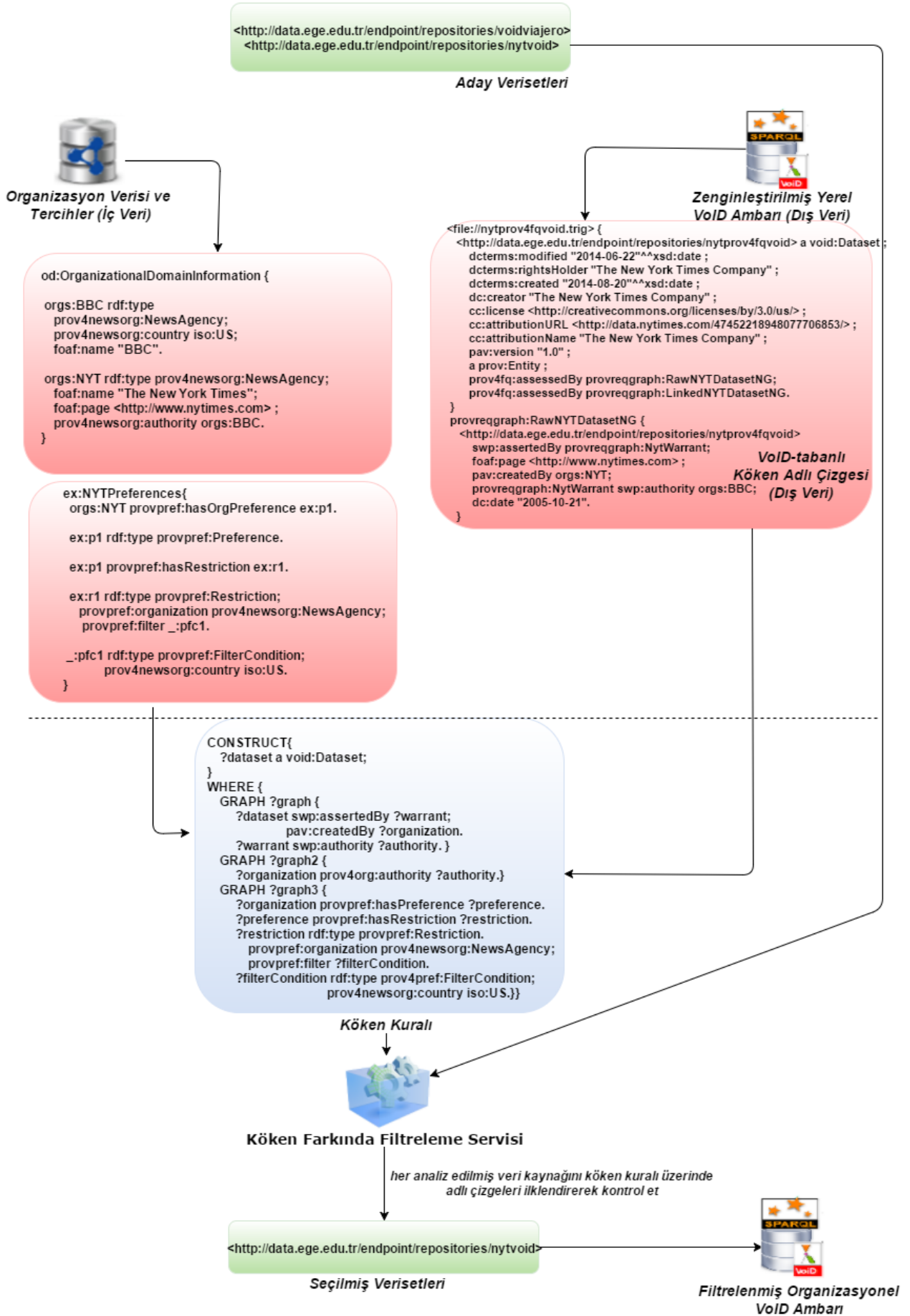
```
PREFIX swp: <http://www.w3.org/2004/03/trix/swp-2/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX prov4org: <http://ege.edu.tr/seagent/ont/prov4org#>
PREFIX prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#>
PREFIX iso: <http://www.daml.org/2001/09/countries/iso-3166-ont#>
PREFIX void: <http://rdfs.org/ns/void#>

CONSTRUCT{ ?dataset a void:Dataset. }
WHERE {
  GRAPH ?graph {
    ?dataset swp:assertedBy ?warrant;
              pav:createdBy ?organization.
    ?warrant swp:authority ?authority. }
  GRAPH ?graph2 {
    ?organization prov4org:authority ?authority.
  }
  GRAPH ?graph3 {
    ?organization provpref:hasPreference ?preference.
    ?preference provpref:hasRestriction ?restriction.
    ?restriction a provpref:Restriction;
                 provpref:organization prov4newsorg:NewsAgency;
                 provpref:filter ?filterCondition.
    ?filterCondition a provpref:FilterCondition;
                     prov4newsorg:country iso:US.
  }
}
```

Şekil 5.5: Adlı çizgeleri sorgulayan örnek kural tanımı



Şekil 5.6: Dış Veri ve İç Veri'yi içeren adlı çizgeler



Şekil 5.7: PROV4FQ Veri Akışı

5.3 Güven Ağı Senaryo Adımları

Güven ile ilgili verilecek karar geçişli bir süreç olabilir yani bir bilgi parçasına veya bilgi kaynağına güvenmek ilgili diğer bir kaynağa güvenmeyi gerektirebilir. Örneğin, bir kitaba ve kitabın yazarına yayımlayıcısından dolayı güvenilebilir. Yayımlayıcıya da bir arkadaşın tavsiyesinden dolayı güvenilmiş olabilir (Artz and Gil, 2007).

SWP, adlı çizgeler üzerinde güven kıstaslarını gerçekleyebilmek için Açık Anahtar Altyapısı(AAA)³⁵ kullanarak güvenli kriptografik dijital imzalar sağlamaktadır. Açık anahtar hakkında örnek vermek gerekirse Bob ile ilişkili bir açık anahtarın Bob'dan geldiği bilinmektedir. Bu durum AAA ile sağlanmaktadır ve açık anahtar-kullanıcı ilişkisi AAA tarafından sağlanan bir otorite (Sertifika Otoritesi(SO)³⁶) tarafından doğrulanmaktadır. Bir açık anahtar sertifikası, açık anahtarın sahipliğini ispat etmek için kullanılan elektronik dokümandır. Sertifika, açık anahtar, anahtar sahibinin kimlik bilgisi ve sertifika içeriğinin doğru olduğuna dair varlığın dijital imzası hakkında bilgi içerir. Eğer imza doğru ise ve sertifikayı ele alan kişi (talep eden) imzalayana güveniyorsa imza sırasında kullanılan açık anahtar artık imzalayan ve talep eden kişi arasında kullanılabilir. Tipik bir AAA mekanizmasında imzayı atan SO, müşterilerine kullanabilecekleri sertifikalar sunar.

Her zaman sertifika yönetim araçları yazmak istenmediğinden dolayı OpenSSL kullanarak sertifikalar üretilebilmekte ve yönetilebilmektedir. OpenSSL dağıtımları bir Perl komut dosyası ile SO oluşturarak yapılması düşünülen işi kolaylaştırmaktadır. Çünkü sistemler kendi sertifikalarını üretmeden önce sertifikanın doğruluğunu garanti etmek için bir SO'ya sahip olmalıdır. SO'ya sahip olunmayan durumlarda sistemin kendi SO'sunu oluşturmak için bahsedildiği şekilde OpenSSL dağıtımı ile oluşturması öngörülmektedir. Kullanıcı sertifika oluşturma aşamasında ise kendi açık-gizli anahtarlarını oluşturup Açık Anahtar Kriptografi Standardı³⁷'na uygun şekilde SO'ya bir sertifika talebinde bulunur ve SO talebin bütünlüğünü inceleyerek talep edenlere imzalı sertifikayı döndürür.

5.3.1 Dış veri tabanlı operasyonlar

Şekil 5.8'de ifade edilen *provreqgraph:LinkedNYTDatasetNG* adlı çizgesi güven ağı gereksinimini karşılamaktadır. Bu adlı çizgede, anlamsal olarak ifade

³⁵İngilizcesi: Public Key Infrastructure - PKI

³⁶İngilizcesi: Certificate Authority - CA

³⁷İngilizcesi: Public Key Cryptography Standard - PKCS

```

provreqgraph:LinkedNYTDataSetNG {
  prov4newsorg:nytrdf a void:Dataset;
    pav:importedFrom <http://data.ege.edu.tr/endpoint/repositories/nytprov4fqvoid>;
    pav:importedOn "2014-11-10T20:50:29"^^xsd:dateTime;
    pav:createdWith <http://www.example.com/nytlinkeddatagenhelper/> ;
    pav:providedBy orgs:NYT.
}

```

Şekil 5.8: VoID dokümanı içerisindeki güvene dair ifadeler

edilen New York Times verisetinin bahsedilen veriseti yardımıyla aktararak anlamsallaştırıldığı ve zamanı, hangi etmen tarafından anlamsallaştırıldığı ve orijinal veriyi sağlayan organizasyon ifade edilmiştir. Bu adlı çizgeyi kullanan güven ağı senaryosuna göre iki aşama bulunmaktadır. Birinci aşama, Bölüm 5.2’de gösterilen adlı çizgeler ve kurallara benzer biçimde filtrelemenin yapılmasıdır. İkinci aşama ise filtrelemeden çıkan veriseti bilgisinin sertifikalarda bulunan dijital imza ile doğrulanmasıdır.

5.3.2 İç veri tabanlı operasyonlar

Şekil 5.9’daki adlı çizgelerde organizasyonlardaki kişilerin bilgileri bulunmaktadır. Bu bilgiler ile organizasyonel köken adlı çizgesi ve VoID adlı çizgeleri kullanılıp Şekil 3.12’de gösterilen yapıya uygun kurallar çalıştırılarak (bkz. Şekil 4.5) adlı çizgeler sorgulanır.

5.3.3 Adlı çizgelerin sorgulanması

Veri tüketicisinin örnek olarak attığı sorguda New York Times tarafından yayımlanan bir haber ile beraber Associated Press ve BBC tarafından sağlanan resimler talep edildiği düşünülmüştür. Güven ağı senaryosunda birinci aşama kuralların çalıştırılması ile gerçekleştirilir. SPARQL kuralları, Şekil 3.12’de bulunan dilbilgisine uygun şekilde oluşturularak varolan adlı çizgeler ile çalıştırılır. Şekil 5.10 ve Şekil 5.11’de örnek kurallar bulunmaktadır.

Şekil 5.10’da *?graph1* VoID dokümanı ile, *?graph2* organizasyon verisi ile, *?graph3* ise organizasyonel tercihler ile karşılanmaktadır. Dolayısıyla *?graph1* Dış Veri ile *?graph2* ve *?graph3* ise İç Veriden gelen adlı çizgeler ile sorguya dahil edilecektir. Kurala göre verisine atfedilen organizasyon dahilinde çalışan bir kişinin bulunması talep edilmektedir. Çalışan kişinin sahip olduğu sertifikasyon bilgisi, Güven Ağı senaryosunun ikinci aşamasında kontrol edilerek seçilmiş veriseti

elde edilir.

Şekil 5.11’de *?graph1* VoID dokümanı ile, *?graph2* ve *?graph3* organizasyon verisi ile, *?graph4* ise organizasyonel tercihler ile karşılanmaktadır (Bkz. Şekil 3.12). Dolayısıyla *?graph1* Dış Veri ile *?graph2*, *?graph3* ve *?graph4* ise İç Veriden gelen adlı çizgeler ile sorguya dahil edilecektir. Kurala göre verisetine atfedilen organizasyonun otorite kuruluşunda çalışan kişi, sorguyu atan veri tüketicisi tarafından tanınmıyor ise veriseti filtrelenmektedir. Bu kural, Şekil 5.3, Şekil 5.8, Şekil 5.9 ve organizasyonel tercihlerde bulunan adlı çizgelerdeki veri ile ilklendirilerek New York Times’a dair veriseti seçimi gerçekleştirilir.

```
@prefix od: <http://ege.edu.tr/seagent/ont/org/od#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov4org: <http://ege.edu.tr/seagent/ont/prov4newsorg#> .
@prefix prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#> .
@prefix orgs: <http://ege.edu.tr/seagent/ont/orgs#>.
@prefix person: <http://ege.edu.tr/seagent/ont/person#>.
@prefix prov: <http://www.w3.org/ns/prov#>.

od: APOrganizationalGraph{
    person:Bob rdf:type prov4org:Agent;
                prov4org:actedOnBehalfOf orgs:AssociatedPress.
}
od: BBCOrganizationalGraph{
    person:Sam rdf:type prov4org:Agent;
                prov4org:actedOnBehalfOf orgs:BBC;
                foaf:homepage "http://www.bbc.com#sam".
                foaf:mbox "mailto:sam@bbc.com".
}
od: ConsumerOrganizationalGraph{
    person:Alice rdf:type prov4org:Agent;
                prov4org:actedOnBehalfOf orgs:ConsumerOrg;
                foaf:knows person:Sam.
}
od: NYTOrganizationalGraph{
    person:Richard rdf:type prov4org:Agent;
                prov4org:actedOnBehalfOf orgs:NYT;
                foaf:homepage "http://www.nytimes.com#richard".
                foaf:mbox "mailto:richard@nytimes.com".
}
```

Şekil 5.9: Organizasyonel veriye dair adlı çizgeler

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#>
PREFIX provpref: <http://ege.edu.tr/seagent/ont/provpref#>
PREFIX pav: <http://purl.org/pav/>
PREFIX void: <http://rdfs.org/ns/void#>

CONSTRUCT{ ?dataset a void:Dataset. }
WHERE {
  GRAPH ?graph1 {
    ?dataset pav:providedBy ?publisherOrganization. }
  GRAPH ?graph2 {
    ?org_person prov4org:actedOnBehalfOf ?publisherOrganization. }
  GRAPH ?graph3 {
    ?consumerOrganization provpref:hasPreference ?preference.
    ?preference provpref:hasRestriction ?restriction.
    ?restriction a provpref:Restriction;
                  provpref:organization prov4newsorg:NewsAgency;
                  provpref:filter ?filterCondition.
    ?filterCondition a provpref:FilterCondition;
                     provpref:organization ?publisherOrganization.
  }
}

```

Şekil 5.10: Örnek Güven Ağı kural tanımı-1

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX prov4org: <http://ege.edu.tr/seagent/ont/prov4org#>
PREFIX prov4newsorg: <http://ege.edu.tr/seagent/ont/prov4newsorg#>
PREFIX pav: <http://purl.org/pav/>
PREFIX void: <http://rdfs.org/ns/void#>

CONSTRUCT{ ?dataset a void:Dataset. }
WHERE {
  GRAPH ?graph1 {
    ?dataset pav:providedBy ?publisherOrganization. }
  GRAPH ?graph2 {
    ?consumer foaf:knows ?org_person. }
  GRAPH ?graph3 {
    ?org_person prov4org:actedOnBehalfOf ?authority. }
  GRAPH ?graph4 {
    ?consumerOrganization provpref:hasPreference ?preference.
    ?preference provpref:hasRestriction ?restriction.
    ?restriction a provpref:Restriction;
                  provpref:organization prov4newsorg:NewsAgency;
                  provpref:filter ?filterCondition.
    ?filterCondition a provpref:FilterCondition;
                     provpref:organization ?authority.
    ?publisherOrganization prov4org:authority ?authority.
  }
}

```

Şekil 5.11: Örnek Güven Ağı kural tanımı-2

5.3.4 Sertifika sahibi kişilerle ilgili verisetlerinin doğrulanması

Güven ağı senaryosunun ikinci aşamasını sertifika sahibi kişilerle ilgili verisetlerinin dijital imzalarının doğrulanması oluşturmaktadır. Tez çalışmasında güven ağı üzerinden filtreleme yapmak amacı ile SWP API yardımı ile oluşturulan örnek sertifikalar (Bizer et al., 2010), Şekil 5.12’de görüldüğü gibi VoID verisetlerinde bulunan uçnoktası bilgisi ile eşleştirilmiştir.

```
@Before
public void setUp(){
    final Certificate[] chain = PKCS12Utils.getCertChain(keystore,
                                                         password);

    DatasetAuthority ds1 = new DatasetAuthority();
    ds1.setEndpoint(QueryVocabulary.BBC_ENDPOINT_URL);

    final SWPAuthority auth1 = new SWPAuthorityImpl(
        Node.createURI("http://www.bbc.com#sam"));

    auth1.setEmail("mailto:sam@bbc.com");
    auth1.setCertificate((X509Certificate) chain[0]);

    ds1.setAuth(auth1);

    DatasetAuthority ds2 = new DatasetAuthority();
    ds2.setEndpoint(QueryVocabulary.NYT_ENDPOINT_NAMED_GRAPH_URL);

    final SWPAuthority auth2 = new SWPAuthorityImpl(
        Node.createURI("http://www.nytimes.com#richard"));
    auth2.setEmail("mailto:richard@nytimes.com");
    auth2.setCertificate((X509Certificate) chain[1]);

    ds2.setAuth(auth2);

    authoritiesOfDatasets.add(ds1);
    authoritiesOfDatasets.add(ds2);
}
```

Şekil 5.12: Verisetine dair VOID’de bulunan uçnoktası bilgisi ile sertifika bilgisinin eşleştirildiği test kod parçası

Kriptografik objeleri tek bir dosyada saklamaya yarayan PKCS#12 ile organizasyondaki sertifika sahibi tüm kişilerin verisi tutulmaktadır. Kişilerin verisi,

organizasyon verisi adlı çizgelerinde bulunmaktadır. Bahsedilen adlı çizgeler, Şekil 5.9’de görülebileceği üzere hangi organizasyonların altında kimin çalıştığını gösteren bilgiler içermektedir. Sertifikaların oluşturulması ve saklanması SWP API kullanımına uygun şekilde sağlanmıştır.

SWPAuthority tanımı, otorite URI bilgisi, email ve sertifikadan oluşmaktadır. Bahsedilen otorite URI bilgisi, aslında organizasyonel adlı çizgelerdeki organizasyonlar içerisinde çalışan kişilerdir. Verisetinin oluşturulmasında rol alan kişilerin sahip olduğu ve kurumun PKCS#12 anahtar deposunda bulunan sertifikalar sayesinde dijital imzalar doğrulanarak Şekil 5.12’deki örnekte gösterilen uçnokta bilgisine ait VoID dokümanları filtrelendir. SWP içerisinde birçok dijital imza algoritması bulunmakla beraber SWP örneklerinde üzerinde durulan SHA1 ve RSA tez çalışmasında da kullanılmıştır. Senaryoda verilen örneğe göre Şekil 5.12’de görüldüğü üzere varolan NYT ve BBC verisetlerine dair sertifikalardaki dijital imzalar doğrulanarak bu verisetleri seçilmiş verisetleri olarak Filtrelenmiş Organizasyonel VoID Ambarı’na döndürülür.

6 TARTIŞMA VE ÖNERİLER

Kalite değerlendirme aktivitesinin hedefi, Zaveri et. al. (Zaveri et al., 2015) tarafından ifade edilen kalite boyutlarına göre veri analizi yaparak veri kalitesini ölçmektir. Bu boyutlardan biri olan köken için Ek-4’de görülen kavramlar göz önüne alınıp veri analizi yapılarak verisetlerinin kalitesi değerlendirilebilir. Değerlendirme farklı veri tiplerini dikkate almaktadır. Bu veri tipleri, bir yazılım bileşeni tarafından depolanan, alınan veya işlenen bir formatta bulunur ve vebde kullanılır. Bağlı açık veri seviyesinde düşünüldüğünde hangi seviyelerde değerlendirmenin yapılacağı ve bu seviyelerdeki veri tiplerinin ayrımı üzerinde düşünülmelidir. Değerlendirme, RDF üçlülere gibi küçük birimler ile bağdaştırılabileceği gibi verisetinin tümünü ele alacak bir yaklaşım da sunabilir. Kalite değerlendirmesi sonucunda elde edilen sonuçlar, tez çalışmasında olduğu gibi kullanılacak verinin filtrelenerek sunulmasına yardımcı olur.

Bağlı açık veri çalışmalarında değerlendirme süreci üç ayrı veri tipine ayrılarak incelenmektedir (Zaveri et al., 2015):

- Üçlü değerlendirmesine odaklanan *RDF üçlülere*,
- RDF üçlülerinin bir koleksiyonu tarafından tanımlanan varlıklara odaklanan *RDF çizgeleri*,
- Verisetinin adlı çizgelerin kümeleri yardımı ile tanımlanabildiği değerlendirmeye odaklanan *Verisetleri*.

Şekil 6.1’de görüldüğü gibi çalışmaların birçoğu üçlü veya çizge seviyesindedir ve veriseti seviyesindeki çalışmalar fazla bulunmamaktadır. Ancak değerlendirme üçlü seviyesinde oluyor ise daha üst seviye olarak adlandırılan çizge veya veriseti seviyesine de yayılabilir. Tez çalışmasında üstveri üzerinde yapılan çalışmalarda yapıldığı gibi veri, daha ince taneli olarak düşünülüp varlık, üçlü veya çizge bazındaki aşamalardan geçerek veriseti üzerinde sonuca ulaşılmaktadır.

Şekil 6.1’de bulunan çalışmaların bazılarında geliştirilen araçların otomatikleştirme seviyeleri gösterilmiştir. Çalışmaların kullandığı boyutlar ve metriklerin doğası gereği araçlar *otomatik*, *yarı-otomatik* veya *manuel* olarak sınıflandırılmıştır. Otomatikleştirme seviyesi çalışmadaki kullanıcı katılımının miktarına göre belirlenmektedir.

Work	Goal	Type of data			Degree of automation			Tool support
		RDF Triple	RDF Graph	Dataset	Manual	Semi-automated	Automated	
Gil et. al., 2002	Approach to derive an assessment of a data source based on the annotations of many individuals	✓	✓			✓		✓
Golbeck et. al., 2003	Trust networks on the semantic web		✓			✓		✓
Mostafavi et. al., 2004	Spatial data integration	✓	✓					
Golbeck, 2006	Algorithm for computing personalized trust recommendation using the provenance of existing trust annotations in social networks	✓	✓	✓				
Gil et. al., 2007	Trust assessment of web resources		✓					
Lei et. al., 2007	Assessment of semantic metadata	✓	✓					
Hartig, 2008	Trustworthiness of Data on the Web	✓				✓		✓
Bizer et. al., 2009	Information filtering	✓	✓	✓	✓			✓
Bohun et. al., 2010	Data integration	✓	✓			✓		✓
Chen et. al., 2010	Generating semantically valid hypothesis	✓	✓					
Flemming et. al., 2010	Assessment of published data	✓	✓			✓		✓
Hogan et. al., 2010	Assessment of published data by identifying RDF publishing errors and providing approaches for improvement	✓	✓	✓		✓		✓
Shekarpoor et. al., 2010	Method for evaluating trust		✓					
Fürber et. al., 2011	Assessment of published data	✓	✓					
Gamble et. al., 2011	Application of decision networks to quality, trust and utility assessment		✓	✓				
Jacobi et. al., 2011	Trust assessment of web resources		✓					
Bonatti et. al., 2011	Provenance assessment for reasoning	✓	✓					
Gueret et. al., 2012	Assessment of quality of links		✓	✓			✓	✓
Hogan et. al., 2012	Assessment of published data	✓	✓	✓				
Mendes et. al., 2012	Data integration	✓			✓			✓
Rula et. al., 2012	Assessment of time related quality dimensions	✓	✓					
Tekbacak, 2014	Provenance-aware querying	✓	✓	✓	✓			✓

Şekil 6.1: Kaliteyi ve kökeni ele alan çatıların karşılaştırılması(Zaverı et al., 2015)

Gueret et al. (Gu  ret et al., 2012) herhangi bir kullanıcı etkisi olmadan   alışan bir ara   sunmuştur. Ara  , *otomatik* olarak kaynak se  imi yapmakta, veb verisi SPARQL u  noktaları veya referanstan ayrılabilcek kaynaklar ile elde edilmekte ve sisteme giren u  l   k  mesi   zerinde kalite de  erlendirme raporları oluşturulmaktadır.   alışmadaki ama  , d  ş  k kalitedeki ba  lantıları ve uyumsuz a   yapılarını ortadan kaldıracak bir otomatik ba  lantı oluşturma s  reci meydana getirmektir. Oluşturulan ba  lantıların ba  lı veri kalitesini arttıracak  ı d  ş  n  lmektedir. A   metrikleri kullanılarak oluşturulan iyi veya k  t   linklerin k  mesinin uygunlu  u test edilmektedir.

Yarı-otomatik   alışmalara   rnek olarak Flemming (Flemming, 2010), kullanıcıya verisetine dair birkaç soru sorarak   nceden tanımlanmış veri kalite metriklerine a  ırlık verilmesi   zerinde   alışmıştır. B  hm et al. (B  hm et al., 2010), ilerideki incelemeler i  in   nceden tanımlanmış veri k  melerinin keşfedilmesine yardımcı olacak bir aracı kullanıcılara sunmaktadır. TRELLIS kullanıcı aray  z   (Gil and Ratnakar, 2002), de  işik kullanıcıların bir veri kayna  ı ve kendileri ile ilgili veri   zerinde g  veni ifade etmesine izin verir. Bir kaynak   zerinde kullanıcı tarafından alınan kararlar belirtim olarak kaydedilerek uyuşturmayan bilgiyi analiz etmek veya eksik bilgiyi ele almak i  in kullanılır. tRDF yaklaşımlı (Hartig, 2008), RDF verisinin g  veni ile ilgili bir   atı sa  lamaktadır. Bu bileşenlerin t  m  nde kullanıcı veri ile beraber   atı ile de etkileşim halindedir.

Manuel   alışmalara   rnek vermek gerekirse, kullanıcı katılımına y  ksek seviyede ihtiya   duyan WIQA (Bizer and Cyganiak, 2009) ve Sieve (Mendes et al., 2012) ara  ları bulunmaktadır. WIQA Bilgi Kalite De  erlendirme   atısı, kullanıcıların farklı alanlardaki kalite tabanlı bilgi filtreleme politikalarını kullanarak bilgiyi filtrelemesine olanak sa  lamaktadır. Benzeri şekilde Sieve kullanarak bir kullanıcı spesifik kalite de  erlendirme g  revi i  in farklı metrikler ve ilgili notlandırma fonksiyonları tanımlayabilir. Bu metrik tanımlaması kalite de  erlendirme g  revi i  in farklı konfig  rasyonlar i  erebilen bir XML dosyası yaratarak yapılmaktadır. Tez   alışmasında geliştirilen PROV4FQ ile WIQA'daki yaklaşıma benzer bir mantıkla k  ken g  z  n  ne alan bir sorgulama mimarisi ortaya koymaya   alışılmıştır. Geniştletilmiş VoID ile beraber ele alınacak organizasyon verisi ile organizasyonel tercihlerin kullanıcı tarafından girilmesi gereklili  i   alışmayı manuel   alışmalar kategorisine katmaktadır.

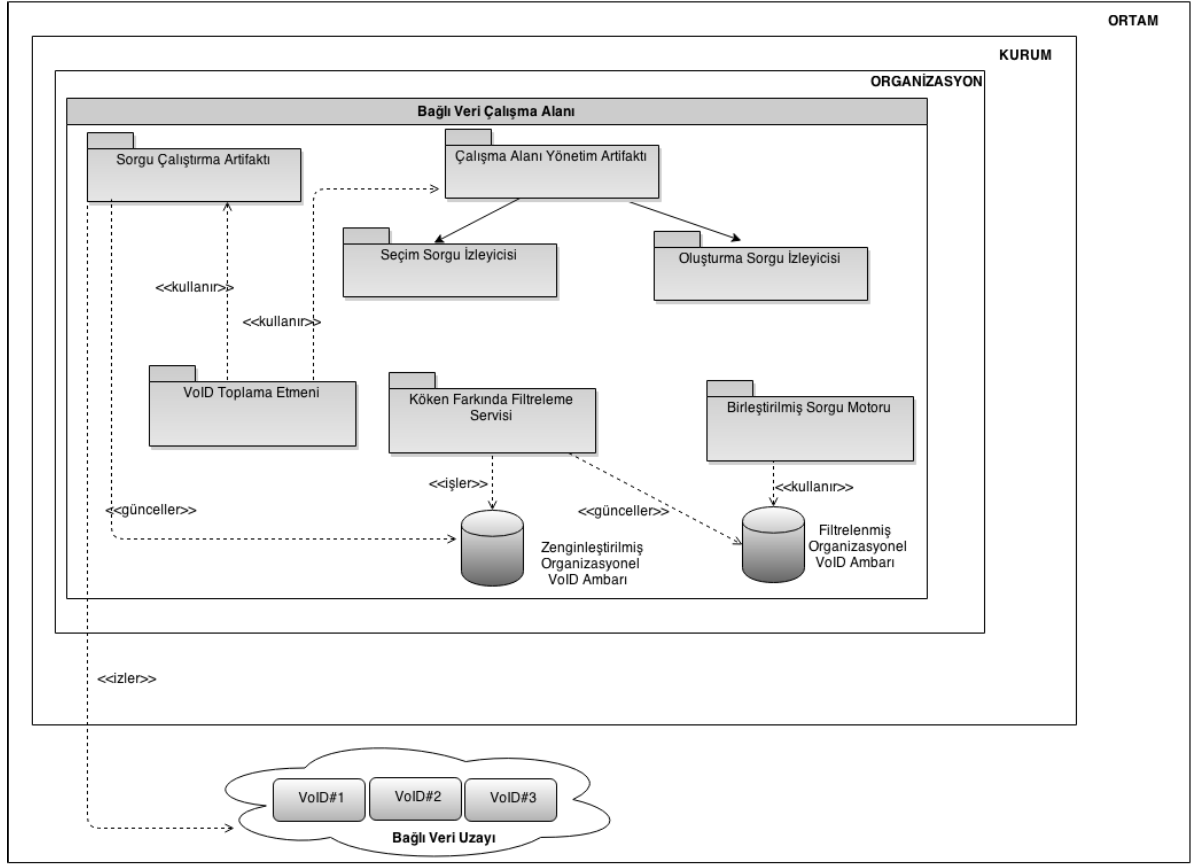
Dinamik verinin y  netimi ba  lı veri uygulamaları i  in temel sorunlardan biridir. Dinamik veri y  netimi ile ilgili yaklaşımlardan veri t  keticisinin ana eleman olarak g  z  n  ne alındı  ı durumlarda, veri t  keticisi belirli periyotlarla de  işim hakkında bilgi sahibi olmak i  in sorgulama yapmak zorundadır. Di  er

yaklaşımında ise yayımlayıcılara kaydolan veri tüketicilerine değişim anında bilgi iletimi yapılmaktadır. Birinci yaklaşım ciddi bir ağ yükü getirmekte ikinci yaklaşım ise bağlı verinin özgür bir şekilde veri dağıtımını sağlayan doğasına uymamaktadır. İlgili olarak Bölüm 4.3’de verisetlerinin izlenmesine dair açıklamalar yapılmıştır.

Ortam, çoklu etmen sistemlerindeki her bir etmenin algılayıcıları yardımı ile farkına vardığı ve etkileycileri yardımı ile eylemde bulunabildiği bir platform sunmaktadır. Ortam tanımına uygun olarak, verisetleri veya veri örneklerindeki değişimin algılanması gerektiği görülmüştür. Bununla birlikte bağlı veri sorgulamanın veriseti değişimi algılanması ile adaptasyonunun sağlanması gerektiği düşünülmüştür. Ortam soyutlaması kullanılarak bağlı veri uzayındaki dinamik verinin izlenmesine dair yapılan çalışma (Erdur et al., 2013) Bölüm 4.3’de ifade edilen kısmın detaylandırılması için yol gösterici niteliktedir. Erdur et al. (Erdur et al., 2013) tarafından geliştirilen bağlı veri ortam altyapısı, bağlı veri bulutundaki herhangi bir değişime dair sorgular ile değişim takibini ortam üzerinde gerçekleştirmektedir. İzleme altyapısı tasarımı ve gerçekleştirimi süresince çoklu-etmen sistemi araştırma alanında yaygın olarak kullanılan ortam soyutlaması (Weyns et al., 2007) üzerinde durulmuştur. Etmenler ve Artifaktlar³⁸ (Omicini et al., 2008), etmen ortamlarını modellemede iyi bilinen bir üstmodel yaklaşımıdır. CArtaGO (Ricci et al., 2011), A&A üstmodeli ile etmen ortamını programlamak için gerekli yapıları sağlayan açık kaynak kodlu bir implementasyondur. Erdur et al. (Erdur et al., 2013), bu yaklaşımı kullanarak bağlı veri uygulama geliştiricilerine SPARQL sorgu izleme imkanı sağlamaktadır.

Erdur et al.’ın (Erdur et al., 2013) üzerinde durduğu ortama dair bahsedilmesi gereken kavramlar olarak artifakt, etmen (kökendeki artifakt ve etmen tanımından farklı bir anlam ifade etmektedir) ve çalışma alanı bulunmaktadır. En basit tanımı ile bir ortam, çalışma alanlarında bulunan artifaktlar ile tasarlanabilir. Bir çalışma alanı, önceden tanımlı artifaktları içermekle beraber uygulama alanında diğer artifaktların oluşturulmasına (fabrika artifaktı) da neden olabilir. Artifaktlar, ortamın bir bölümünün durumu hakkında bilgi vermek amacı ile farklı değerlere sahiptir ve ortamın durumunu değiştirmek için operasyonlar gerçekleştirir. Bu operasyonlar, ortamdaki olayları algılayarak ortaya çıkan değişimlere dair yapılan operasyonlardır. Bir artifakt üzerindeki değişimler hakkında haberdar olmak için etmen ilgili artifakta odaklanmalıdır. Çünkü artifakt içerisinde herhangi bir özellik değiştiğinde artifakt kendine odaklanan etmenlere algılamaları için bilgi gönderir. Artifaktın değişen özelliği hakkındaki yeni değer, etmen tarafından algılanır ve kendi iş mantığına uygun biçimde yorumlanır.

³⁸ İngilizcesi: Agents&Artifacts(A&A)



Şekil 6.2: Veriseti Değişimini Ele Alan Bağlı Veri Ortam Altyapısı

Tez çalışmasına Erdur et al.'ın (Erdur et al., 2013) önerisinin adapte edildiğinde Şekil 6.2'de ifade edildiği üzere değişim takibinin VoID deposundaki VoID verisetlerinin değişimi için de geçerli olduğu öngörülmektedir. Tez çalışmasında önceden bahsedilen Organizasyon kavramına ek olarak kullanılan Kurum ve Ortam kavramları Şekil 6.2'de beraberce ifade edilmektedir. Bu öngörüye göre altyapı, ilgili çalışmada ifade edildiği üzere bir Organizasyon içerisinde bulunan Bağlı Veri Çalışma Alanı³⁹ ve Veri Kaynağı Katmanı'ndan meydana gelmektedir. Erdur et al. (Erdur et al., 2013)'a göre bağlı veri ortamındaki değişimleri takip edip sorgu izleme artifaktlarının yaratılmasından Çalışma Alanı Yönetim Artifaktı sorumludur⁴⁰. Değişim takibi için Seçim Sorgu İzleyicisi⁴¹ ve Oluşturma Sorgu İzleyicisi⁴² yardımı ile VoID değişimleri hakkında sorgular çalıştırılmaktadır. Seçim Sorgu İzleyicisi, organizasyonla ilgili verisetleri listesinde değişiklik olup olmadığı ile ilgilenirken Oluşturma Sorgu İzleyicisi odaklanılan verisetleri üzerinde değişiklik olup olmadığı ile ilgilenir. Yeni veya güncellenen

³⁹ İngilizcesi: Linked Data Workspace

⁴⁰ İngilizcesi: Workspace Management Artifact

⁴¹ İngilizcesi: Select Query Monitor

⁴² İngilizcesi: Construct Query Monitor

verisetleri, VOID Toplama Etmeni⁴³ yardımı ile artifaktlardan gelen bilgiler doğrultusunda toplanmaktadır. Sorgu Çalıştırma Artifaktı⁴⁴, belirli aralıklarla VOID değişimine dair sorguları çalıştırarak herhangi bir değişiklik olup olmadığına dair kontrolde bulunur. Bu artifakta odaklanan VOID Toplama Etmeni değişimlerden sürekli olarak haberdar olur. Böylece güncel VOID verisi üzerinde sorgu çalıştıran Birleştirilmiş Sorgu Motoru (WoDQA) atılacak sorguları Şekil 4.3’de ifade edilen biçimde çalıştırır.

⁴³*İngilizcesi:* VOID Gathering Agent

⁴⁴*İngilizcesi:* Query Executor Artifact

7 SONUÇ

Bağlı veriseti yayımlanmasının yaygınlaşması ile beraber benzeri veri kaynaklarının sağlandığı durumlarla karşılaşmaktadır. Daha kaliteli veya kişisel tercihlere göre bağlı veriye dair benzeri verisetlerinden filtreleme yapılması için veri kaynaklarının bazı süreçlerden geçmesi gerekmektedir. Bu tezde, birleştirilmiş sorgu motorlarının kulanacağı verisetleri için verisetlerine dair VoID üstverileri kullanılarak ve genişletilerek filtreleme yapılmasına dair süreçler ortaya konmuştur. Köken üstverisi yardımı ile geliştirilen filtreleme mekanizmasının sağladığı yenilik, ontolojik olarak ifade edilen köken bazlı kavramların bir köken farkında filtreleme servisi ile uygulanarak elde edilen işlenmiş verinin birleştirilmiş sorgu motoru tarafından kullanılmasıdır. VoID üzerinde yapılacak ek ontolojik tanımlarla ilgili sözlükler genişletilerek VoID'in bu sözlükler kullanılarak kökeni gözönüne alan kalite kıstaslarına göre üstveri tutması gerektiği düşünülmüştür. Genişletilen VoID dokümanlarındaki köken farkında üstveri ile organizasyon verisi ve organizasyonların tercihleri kurallarda ele alınarak filtreleme sağlanmaktadır. Köken farkında filtreleme servisi, eline gelen VoID'e dair her uç noktası ile ilgili kuralları işleterek filtrelemeyi gerçekleştirmekten ve birleştirilmiş sorgu motorunda kullanılacak verisetlerinin seçiminden sorumludur.

Tez içerisindeki durum çalışmaları, varolan ontolojiler ile belirtilmesi kolay olduğundan dolayı seçilmiş olmakla beraber kalite değerlendirmesi ve kökeni gözönüne alan çok fazla alt başlık bulunmaktadır (Bkz. Ek-4). Farklı uygulama alanlarına dair organizasyonların ihtiyaç duyduğu bu boyutlar daha detaylı düşünülerek yeni sözlükler yardımı ile VoID üstverisi zenginleştirilebilecektir.

KAYNAKLAR DİZİNİ

- Akar, Z., Halaç, T. G., Ekinçi, E. E., and Dikenelli, O.**, 2012, Querying the web of interlinked datasets using void descriptions., in C. Bizer, T. Heath, T. Berners-Lee, and M. Hausenblas (eds.), *LDOW*, Vol. 937 of *CEUR Workshop Proceedings*, CEUR-WS.org
- Alexander, K., Cyganiak, R., Hausenblas, M., and Zhao, J.**, 2009, Describing linked datasets: On the design of void, the "vocabulary of interlinked datasets", in *Linked Data on the Web (LDOW 2009)*, *Proceedings of the WWW2009 Workshop*, CEUR Workshop Proceedings
- Artz, D. and Gil, Y.**, 2007, A survey of trust in computer science and the semantic web, *Web Semant.* **5(2)**, 58 pp.
- Auer, S., Dietzold, S., Lehmann, J., Hellmann, S., and Aumüller, D.**, 2009, Triplify: Light-weight linked data publication from relational databases, in *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, New York, NY, USA, ACM, 621–630 pp.
- Auer, S., Lehmann, J., Ngomo, A.-C. N., and Zaveri, A.**, 2013, Introduction to linked data and its lifecycle on the web, in *Reasoning Web*, 1-90 pp.
- Baillie, C. C., Edwards, P., Pignotti, E., and Corsar, D.**, 2013, Short paper: Assessing the quality of semantic sensor data., in s. Corcho, C. A. Henson, and P. M. Barnaghi (eds.), *SSN@ISWC*, Vol. 1063 of *CEUR Workshop Proceedings*, CEUR-WS.org, 71-76 pp.
- Becker, C. and Bizer, C.**, 2009, Exploring the geospatial semantic web with {DBpedia} mobile, *Web Semantics: Science, Services and Agents on the World Wide Web* **7(4)**, 278 pp., Semantic Web challenge 2008
- Berners-Lee, T.**, 1998, Semantic web road map, <http://www.w3.org/DesignIssues/Semantic.html> (Erişim tarihi: 20 Kasım 2014)
- Berners-Lee, T.**, 2006, Linked data: Design issues, <http://www.w3.org/DesignIssues/LinkedData.html> (Erişim tarihi: 19 Haziran 2014)
- Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., and Sheets, D.**, 2006, Tabulator: Exploring and analyzing linked data on the semantic web, in *Proceedings of the 3rd International Semantic Web User Interaction Workshop*
- Böhm, C., Naumann, F., Abedjan, Z., Fenz, D., Grütze, T., Hefenbrock, D., Pohl, M., and Sonnabend, D.**, 2010, Profiling linked open data with prolod., in *ICDE Workshops*, IEEE, 175-178 pp.

KAYNAKLAR DİZİNİ (Devam)

- Bizer, C.**, 2007, *Quality-Driven Information Filtering in the Context of Web-Based Information Systems*, Ph.D. thesis, Freie Universitat Berlin
- Bizer, C.**, 2011, Evolving the web into a global data space., in A. A. A. Fernandes, A. J. G. Gray, and K. Belhajjame (eds.), *BNCOD*, Vol. 7051 of *Lecture Notes in Computer Science*, Springer, 1 pp.
- Bizer, C. and Auer, S.**, 2014, Dbpedia, <http://wiki.dbpedia.org> (Erişim tarihi: 15 Mayıs 2015)
- Bizer, C. and Cyganiak, R.**, 2006, *D2R Server – Publishing Relational Databases on the Semantic Web*, Poster at the 5th International Semantic Web Conference (ISWC2006)
- Bizer, C. and Cyganiak, R.**, 2009, Quality-driven information filtering using the wiqa policy framework, *Web Semantics: Science, Services and Agents on the World Wide Web* 7(1), 1 pp.
- Bizer, C., Cyganiak, R., and Hartig, O.**, 2010, Ng4j - named graphs api for jena, <http://wifo5-03.informatik.uni-mannheim.de/bizer/ng4j/#SWP> (Erişim tarihi: 19 Eylül 2014)
- Bizer, C., Cyganiak, R., and Heath, T.**, 2007, How to publish linked data on the web, <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/> (Erişim tarihi: 20 Kasım 2014)
- Bizer, C. and Gauss, T.**, 2007, Disco - hyperdata browser: A simple browser for navigating the semantic web, <http://wifo5-03.informatik.uni-mannheim.de/bizer/ng4j/disco/> (Erişim tarihi: 19 Kasım 2014)
- Bizer, C., Heath, T., and Berners-Lee, T.**, 2009, Linked data - the story so far, *Int. J. Semantic Web Inf. Syst.* 5(3), 1 pp.
- Böhm, C., Lorey, J., and Naumann, F.**, 2011, Creating void descriptions for web-scale data, *Web Semant.* 9(3), 339 pp.
- Carroll, J. J., Bizer, C., Hayes, P., and Stickler, P.**, 2005, Named graphs, provenance and trust, in *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, New York, NY, USA, ACM, 613–622 pp.
- Cheney, J., Gil, Y., Groth, P. T., and Miles, S.**, 2010a, User requirements, http://www.w3.org/2005/Incubator/prov/wiki/User_Requirements (Erişim tarihi: 18 Haziran 2014)

KAYNAKLAR DİZİNİ (Devam)

- Cheney, J., Gil, Y., Groth, P. T., and Miles, S.**, 2010b, W3c provenance incubator group wiki, http://www.w3.org/2005/Incubator/prov/wiki/Main_Page (Erişim tarihi: 20 Haziran 2014)
- Cheng, G. and Qu, Y.**, 2009, Searching linked objects with falcons: Approach, implementation and evaluation., *Int. J. Semantic Web Inf. Syst.* **5(3)**, 49 pp.
- Ciccarese, P. and Soiland-Reyes, S.**, 2015, Pav - provenance, authoring and versioning, <http://pav-ontology.github.io/pav/> (Erişim tarihi: 18 Mart 2015)
- Ciccarese, P., Soiland-Reyes, S., Belhajjame, K., Gray, A. J. G., Goble, C. A., and Clark, T.**, 2013, *PAV ontology: provenance, authoring and versioning*
- Clements, P., Garlan, D., Bass, L., Stafford, J., Nord, R., Ivers, J., and Little, R.**, 2002, *Documenting Software Architectures: Views and Beyond*, Pearson Education
- Consens, Mariano P., H. O. and Teisanu, A. M.**, 2014, Linkedmdb, <http://wiki.dbpedia.org> (Erişim tarihi: 15 Mayıs 2015)
- Cyganiak, R.**, 2010, Void - metadata for rdf datasets, <http://www.slideshare.net/cygri/void-metadata-for-rdf-datasets> (Erişim tarihi: 19 Haziran 2014)
- Cyganiak, R. and Bizer, C.**, 2011, Pubby - a linked data frontend for sparql endpoints, <http://wifo5-03.informatik.uni-mannheim.de/pubby/> (Erişim tarihi: 25 Haziran 2014)
- Debattista, J., Lange, C., and Auer, S.**, 2014, daq, an ontology for dataset quality information, in *Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014)*, Seoul, Korea, April 8, 2014.
- Dividino, R., Gröner, G., Scheglmann, S., and Thimm, M.**, 2012, Ranking rdf with provenance via preference aggregation, in *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW'12)*
- DublinCoreMetadataInitiative**, 2014, The metadata community — supporting innovation in metadata design, implementation, best practices, <http://www.dublincore.org/> (Erişim tarihi: 23 Haziran 2014)
- Erdur, R. C., Alatli, O., Halaç, T. G., and Dikenelli, O.**, 2013, Monitoring the dynamism of the linked data space through environment abstraction, in *Proceedings of the 9th International Conference on Semantic Systems*,

KAYNAKLAR DİZİNİ (Devam)

- I-SEMANTICS '13, New York, NY, USA, ACM, 81–88 pp.
- Flemming, A.**, 2010, *Quality Characteristics of Linked Data Publishing Datasources*, [Online; accessed 13-February-2014]
- Franklin, M., Halevy, A., and Maier, D.**, 2005, From databases to dataspace: A new abstraction for information management, *SIGMOD Rec.* **34(4)**, 27 pp.
- Gil, Y. and Ratnakar, V.**, 2002, Trusting information sources one citizen at a time., in I. Horrocks and J. A. Hendler (eds.), *International Semantic Web Conference*, Vol. 2342 of *Lecture Notes in Computer Science*, Springer, 162-176 pp.
- Görlitz, O. and Staab, S.**, 2011a, Federated data management and query optimization for linked open data, in *New Directions in Web Data Management I*, 109-137 pp.
- Görlitz, O. and Staab, S.**, 2011b, Splendid: Sparql endpoint federation exploiting void descriptions, in *COLD*
- Gray, A. J.**, 2014, Dataset descriptions for linked data systems, *Internet Computing, IEEE* **18(4)**, 66 pp.
- Groth, P. T., Gil, Y., Cheney, J., and Miles, S.**, 2012, Requirements for provenance on the web., *IJDC* **7(1)**, 39 pp.
- Guéret, C., Groth, P., Stadler, C., and Lehmann, J.**, 2012, Assessing linked data mappings using network measures, in *Proceedings of the 9th International Conference on The Semantic Web: Research and Applications*, ESWC'12, Berlin, Heidelberg, Springer-Verlag, 87–102 pp.
- Halevy, A., Franklin, M., and Maier, D.**, 2006, Principles of dataspace systems, in *Proceedings of the Twenty-fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '06, New York, NY, USA, ACM, 1–9 pp.
- Hartig, O.**, 2008, Trustworthiness of data on the web, in *Proceedings of the STI Berlin CSW PhD Workshop*
- Hartig, O., Bizer, C., and Freytag, J.-C.**, 2009, Executing sparql queries over the web of linked data, in *Proceedings of the 8th International Semantic Web Conference*, ISWC '09, Berlin, Heidelberg, Springer-Verlag, 293–309 pp.
- Hartig, O. and Zhao, J.**, 2010, Publishing and consuming provenance metadata on the web of linked data., in D. L. McGuinness, J. Michaelis, and L. Moreau (eds.), *IPAW*, Vol. 6378 of *Lecture Notes in Computer Science*, Springer, 78-90 pp.

KAYNAKLAR DİZİNİ (Devam)

- Hartig, O. and Zhao, J.**, 2012, Provenance vocabulary core ontology specification, <http://trdf.sourceforge.net/provenance/ns.html> (Erişim tarihi: 18 Haziran 2014)
- Heath, T.**, 2007, The linking open data project, <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> (Erişim tarihi: 20 Kasım 2014)
- Heath, T. and Bizer, C.**, 2011, *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, 1st edition
- Hogan, A., Harth, A., Umbrich, J., Kinsella, S., Polleres, A., and Decker, S.**, 2011, Searching and browsing linked data with swse: The semantic web search engine, *Web Semantics: Science, Services and Agents on the World Wide Web* **9(4)**, 365 pp., {JWS} special issue on Semantic Search
- International Press Telecommunication Council**, 2014, The latest generation of multi-media news exchange, http://www.iptc.org/site/News_Exchange_Formats/NewsML-G2/ (Erişim tarihi: 23 Haziran 2014)
- Jacobi, I., Kagal, L., and Khandelwal, A.**, 2011, Rule-based trust assessment on the semantic web., in N. Bassiliades, G. Governatori, and A. Paschke (eds.), *RuleML Europe*, Vol. 6826 of *Lecture Notes in Computer Science*, Springer, 227-241 pp.
- Jonquet, C., LePendu, P., Falconer, S., Coulet, A., Noy, N. F., Musen, M. A., and Shah, N. H.**, 2011, Ncbo resource index: Ontology-based search and mining of biomedical resources, *Web Semantics* **9(3)**, 316 pp., 1st prize of Semantic Web Challenge at the 9th International Semantic Web Conference, ISWC'10, Shanghai, China
- Khandelwal, A., Bao, J., Kagal, L., Jacobi, I., Ding, L., and Hendler, J. A.**, 2010, Analyzing the air language: A semantic web (production) rule language., in P. Hitzler and T. Lukasiewicz (eds.), *RR*, Vol. 6333 of *Lecture Notes in Computer Science*, Springer, 58-72 pp.
- Khandelwal, A., Jacobi, I., and Kagal, L.**, 2011, Linked rules: Principles for rule reuse on the web., in S. Rudolph and C. Gutierrez (eds.), *RR*, Vol. 6902 of *Lecture Notes in Computer Science*, Springer, 108-123 pp.
- Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., and Zhao, J.**, 2013, Prov-o: The prov ontology, *W3C Recommendation*, <http://www.w3.org/TR/prov-o/> (Erişim tarihi: 20 Haziran 2014)

KAYNAKLAR DİZİNİ (Devam)

- Madhavan, J., Cohen, S., Dong, X. L., Halevy, A. Y., Jeffery, S. R., Ko, D., and Yu, C.**, 2007, Web-scale data integration: You can afford to pay as you go., in *CIDR*, www.cidrdb.org, 342-350 pp.
- Markovic, M., Edwards, P., and Corsar, D.**, 2013, Utilising provenance to enhance social computation., in H. Alani, L. Kagal, A. Fokoue, P. T. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. F. Noy, C. Welty, and K. Janowicz (eds.), *International Semantic Web Conference (2)*, Vol. 8219 of *Lecture Notes in Computer Science*, Springer, 440-447 pp.
- Mendes, P. N., Mühleisen, H., and Bizer, C.**, 2012, Sieve: Linked data quality assessment and fusion, in *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, EDBT-ICDT '12, New York, NY, USA, ACM, 116–123 pp.
- Missier, P., Dey, S., Belhajjame, K., Cuevas-Vicenttín, V., and Ludäscher, B.**, 2013, D-prov: Extending the prov provenance model with workflow structure, in *Proceedings of the 5th USENIX Workshop on the Theory and Practice of Provenance*, TaPP '13, Berkeley, CA, USA, USENIX Association, 9:1–9:7 pp.
- Möller, K., Hausenblas, M., Cyganiak, R., and Grimnes, G. A.**, 2010, Learning from linked open data usage: Patterns and metrics, in *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*
- Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., and den Bussche, J. V.**, 2011, The open provenance model core specification (v1.1), *Future Generation Computer Systems* **27**(6), 743 pp.
- Moreau, L., Hartig, O., Simmhan, Y., Myers, J., Lebo, T., Belhajjame, K., and Miles, S.**, 2012, *PROV-AQ: Provenance Access and Query*, Technical report
- Mountantonakis, M., Allocca, C., Fafalios, P., Minadakis, N., Marketakis, Y., Lantzaki, C., and Tzitzikas, Y.**, 2014, Extending void for expressing connectivity metrics of a semantic warehouse, in *Proceedings of the 1st International Workshop on Dataset PROFiling & fEderated Search for Linked Data co-located with the 11th Extended Semantic Web Conference, PROFILES@ESWC 2014, Anissaras, Crete, Greece, May 26, 2014*.
- Omicini, A., Ricci, A., and Viroli, M.**, 2008, Artifacts in the a&a meta-model for multi-agent systems, *Autonomous Agents and Multi-Agent Systems* **17**(3), 432 pp.
- Omitola, T., Gutteridge, C., and Gibbins, N.**, 2011a, voidp: A vocabulary for data and dataset provenance, <http://www.enakting.org/provenance/voidp/> (Erişim

KAYNAKLAR DİZİNİ (Devam)

tarihi: 20 Haziran 2014)

- Omitola, T., Zuo, L., Gutteridge, C., Millard, I. C., Glaser, H., Gibbins, N., and Shadbolt, N.**, 2011b, Tracing the provenance of linked data using void, in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, New York, NY, USA, ACM, 17:1–17:7 pp.
- Polleres, A.**, 2007, From sparql to rules (and back), in *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, New York, NY, USA, ACM, 787–796 pp.
- Polo, L., Mínguez, I., Berrueta, D., Ruiz, C., and Gómez-Pérez, J. M.**, 2014, User preferences in the web of data., *Semantic Web* **5**(1), 67 pp.
- Ramachandran, S., Paulraj, S., Joseph, S., and Ramaraj, V.**, 2009, Enhanced trustworthy and high-quality information retrieval system for web search engines, *CoRR* abs/0911.0914
- Reynolds, D.**, 2014, The organization ontology, *W3C Recommendation*, <http://www.w3.org/TR/vocab-org/> (Erişim tarihi: 12 Ekim 2015)
- Ricci, A., Pianti, M., and Viroli, M.**, 2011, Environment programming in multi-agent systems: an artifact-based perspective, *Autonomous Agents and Multi-Agent Systems* **23**(2), 158 pp.
- Sacco, O. and Passant, A.**, 2011, A privacy preference ontology (ppo) for linked data., in C. Bizer, T. Heath, T. Berners-Lee, and M. Hausenblas (eds.), *LDOW*, Vol. 813 of *CEUR Workshop Proceedings*, CEUR-WS.org
- Tenenboim, L., Shapira, B., and Shoval, P.**, 2008, Ontology-based Classification of News in an Electronic Newspaper, in *Proceedings of INFOS 2008*, Varna, Bulgaria, 89–97 pp.
- TheNewYorkTimesCompany**, 2010, The new york times linked open data, <http://data.nytimes.com/> (Erişim tarihi: 19 Haziran 2014)
- Toniolo, A., Cerutti, F., Oren, N., Norman, Timothy, J., and Sycara, K.**, 2014, Making informed decisions with provenance and argumentation schemes, *Proceedings of the 11th International Workshop on Argumentation in Multi-Agent Systems*
- Toniolo, A., Dropps, T., Cerutti, F., Oren, N., Norman, Timothy, J., and Sycara, K.**, 2015, Reasoning about provenance preferences in cispaces, Annual Fall Meeting of the International Technology Alliance (ITA AFM)

KAYNAKLAR DİZİNİ (Devam)

- Tummarello, G., Cyganiak, R., Catasta, M., Danielczyk, S., Delbru, R., and Decker, S.**, 2010, Sig.ma: Live views on the web of data, *Web Semantics: Science, Services and Agents on the World Wide Web* **8(4)**, 355 pp., Semantic Web Challenge 2009 User Interaction in Semantic Web research
- Vandenbussche, P.-Y.**, 2014, Linked open vocabularies (lov), <http://lov.okfn.org/dataset/lov/> (Erişim tarihi: 5 Aralık 2014)
- Weyns, D., Omicini, A., and Odell, J.**, 2007, Environment as a first class abstraction in multiagent systems, *Autonomous Agents and Multi-Agent Systems* **14(1)**, 5 pp.
- Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., and Auer, S.**, 2015, Quality assessment for linked data: A survey, *Semantic Web Journal*

ÖZGEÇMİŞ

Adı Soyadı: Fatih TEKBACAK
Doğum Tarihi: 1982
Doğum Yeri: İzmir
Uyruğu: T.C.

EĞİTİM

Yüksek Lisans: İzmir Yüksek Teknoloji Enstitüsü
Bilgisayar Yazılımı, 2006
Lisans: İzmir Yüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği Bölümü, 2004
Lise: İzmir Atatürk Lisesi, 2000

İLGİ ALANLARI

Anlamsal Veb, Köken, Bağlı Veri, Güven, Yazılım Mühendisliği

EKLER

EK1: Türkçe-İngilizce Terimler Sözlüğü

EK2: Önek Tanımları

EK3: Senaryodaki İhtiyaçlar Doğrultusunda Eklenen Önek Tanımları

EK4: Köken Boyutları

EK1: Türkçe-İngilizce Terimler Sözlüğü

adlı çizge	<i>named graph</i>
ağ güncesi	<i>blog</i>
bağlı veri	<i>linked data</i>
bağ dolaşma	<i>link traversal</i>
belirtim	<i>annotation</i>
çalışma alanı	<i>workspace</i>
çizge	<i>graph</i>
dağıtık	<i>federated</i>
değerlendirme	<i>assessment</i>
durum çalışması	<i>use case</i>
eşleştirme	<i>mapping</i>
geçişli	<i>transitive</i>
güven ağı	<i>web of trust</i>
ilişkilendirme	<i>association</i>
köken	<i>provenance</i>
köprü	<i>hyperlink</i>
ortam	<i>environment</i>
referanstan ayırma	<i>dereference</i>
sarıcı	<i>wrapper</i>
uçnokta	<i>endpoint</i>
veri tüketimi	<i>data consumption</i>
veri uzayı	<i>dataspace</i>
yığın	<i>dump</i>

EK2: Önek Tanımları

Önek Gösterimi	Aduzayı URI Tanımı
prov:	<http://www.w3.org/ns/prov#>
pav:	<http://purl.org/pav/>
dc:	<http://purl.org/dc/elements/1.1/>
rdfs:	<http://www.w3.org/2000/01/rdf-schema#>
cc:	<http://creativecommons.org/ns#>
owl:	<http://www.w3.org/2002/07/owl#>
rdf:	<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
dcterms:	<http://purl.org/dc/terms/>
foaf:	<http://xmlns.com/foaf/0.1/>
iso:	<http://www.daml.org/2001/09/countries/iso-3166-ont#>
swp:	<http://www.w3.org/2004/03/trix/swp-1#>

EK3: Senaryodaki İhtiyaçlar Doğrultusunda Eklenen Önek Tanımları

Önek Gösterimi	Aduzayı URI Tanımı
prov4org:	<http://ege.edu.tr/seagent/ont/prov4org#>
prov4newsorg:	<http://ege.edu.tr/seagent/ont/prov4newsorg#>
provpref:	<http://ege.edu.tr/seagent/ont/prov/provpref#>
od:	<http://ege.edu.tr/seagent/ont/org/od#>
orgs:	<http://ege.edu.tr/seagent/ont/orgs#>
provreqgraph:	<http://ege.edu.tr/seagent/ont/prov/provreqgraph#>
person:	<http://ege.edu.tr/seagent/ont/person#>

EK4: Köken Boyutları

Kategoriler	Boyutlar	Gereksinim Öneki
İçerik	Amaç	İ-Ama
	Atfetme	İ-Atf
	Süreç	İ-Sür
	Versiyonlama	İ-Ver
	Gerekçe	İ-Gerç
	Gereklilik	İ-Ger
Yönetim	Yayımlama	Y-Yayım
	Erişim	Y-Eriş
	Yayılma	Y-Yayıl
	Ölçek	Y-Ölç
Kullanım	Anlaşılabilirlik	K-Anl
	Birlikte Çalışabilirlik	K-BÇal
	Karşılaştırma	K-Karş
	İzlenebilirlik	K-İzl
	Güven	K-Güv
	Kusurluluk	K-Kus
	Hata Ayıklama	K-HAyık

İçerik/Amaç

Amaç boyutunda, köken ile ilgili kullanılan ifadelerdeki artifakt(veya varlık)ların hangi hedeflerle kullanıldığı ve ne ifade ettiği üzerinde durulmaktadır. Veb ele alındığında artifaktlar, veb dokümanları, verisetleri veya veb servisleri gibi bir URI ile temsil edilen veb kaynaklarıdır.

Haber Birleştiricisi Senaryosunda kökenin amacı kullanıcıyı haber varlığının orijinine ulaştırabilmektir. Bazı durumlarda köken bilgisini elde etmek için bir artifaktın (örneğin haber varlığının) belirli bir kısmına erişmek gerekebilir. Bu koşul altında, bir nesnenin yaşam döngüsü içerisinde etkileşimde bulunduğu farklı sistemler arasındaki geçişleri de gözönüne alarak kökeni hakkında bilgi edinilebilmelidir. Örneğin bir haber, yayımlayıcı kuruluş tarafından dağıtıldıktan sonra haberi içeren gazete kendince değişiklikler yaparak haberi yayımlamış olabilir. Bunun dışında yayımlayıcı kuruluş da güncellemeler yaparak yayımladığında fakat gazete kendi haberini bu güncellemeye tabi tutmadığında vebde birden çok aynı kaynaktan çıkmış tutarsız haberler bulunabilir. Bunun önüne geçebilmek için değişen her bir haber parçasının da kökeninin takibine ihtiyaç duyulmaktadır ve bu

durumda o haber parçası amaç olarak ifade edilmektedir.

İçerik/Atfetme

İçerik kategorisinde köken kayıtlarının yapısı ve anlamı ifade edilir. Tek bir uygulama düşünüldüğünde köken için bir veri modeli oluşturmak kolay gibi görünmesine rağmen farklı uygulama alanlarına uyabilecek bir standart oluşturmak çalışmanın esasını içermektedir.

Dokümanlar, veb kaynakları veya bir kaynakla ilgili URI içeren kaynaklar, insanlar, organizasyonlar ve diğer kimliği saptanabilir gruplar kategorik olarak temsil edilebilir. *Atfetme*, artifakt yaratmak için kullanılan kaynak ve varlıkların kökenini sorgulamaya yarar. Haber Birleştiricisi Senaryosu'na göre kaynaktan kasıt bir veb sitesindeki haber olmakla beraber varlığa örnek olarak ağ güncesi yaratan bir kişi verilebilir.

Atfetme, iki alt başlığa bölünmektedir:

Sorumluluk: Belirli bir parça bilgiyi kimin onayladığını veya desteklediğini saptamak adına kullanılır. Örneğin, fotoğrafçının çektiği bir fotoğrafta gazete tarafından değiştirme yapıp yayımlanırsa fotoğrafın içeriği bu değiştirmeden sonra yasal olmayan bir hal alabilir. Bu nedenle fotoğraf ile ilgili sorumluluğun kime ait olduğuna dair bir köken kuralı tanımlanmalıdır.

Köken iddialarının kaynağı için belirtiler oluşturmak amacıyla atfetme kullanılabilir. Basitçe, kökenle alakalı varlıkların kim tarafından kaydedildiğine, güncellendiğine veya doğrulandığına dair üçüncü şahısların da hesaba katılabildiği durumları tanımlar.

Orijin: Şu an kaydedilmiş veya yeniden düzenlenmiş, doğrulanmış veya doğrulanmamış (dijital imza ile), iddia edilmiş veya iddia sonucuna varılmış bilginin başlangıç noktasına erişebilmek amacını güder. Bir kaynağın orjinal, yeniden düzenlenmiş veya üçüncü şahıslar tarafından iddia edilmiş olmasına dair bulgular orijinin kapsamına girmektedir.

Atfetme, senaryoda anahtar bir kısmı oluşturmaktadır. Üstveri ile oluşturulan ifadeleri kimin veya ne şekilde meydana getirdiği önemlidir. Dublin Çekirdeği (DublinCoreMetadataInitiative, 2014) atfetme bilgisini ifade etmek için kullanılan yaygın bir sözlüktür.

İçerik/Süreç

Artifaktın meydana gelmesi esnasındaki aktivite veya adımları temsil eder. Bu aktivite bir bilgisayar programının çalışarak verinin formatını değiştirmesi veya bir kişinin gerçekleştirdiği bir eylemin meydana getirdiği değişim olabilir. Süreçler çok soyut bir seviyede gösterilebileceği gibi dakikaların detaylarını bile içerecek kadar geniş kapsamlı olabilir. Örneğin, kullanıcı bir maç haberi ile ilgili sadece skor bilgisini öğrenmek isterken, bir diğeri skordan öte maç içinde her dakika gerçekleşen olaylar sonucunda o noktaya nasıl geldiği hakkında bilgi sahibi olmak isteyebilir.

Aktiviteler birbirleri ile zamansal veya nedensel olarak bağlantılı olabilirler. Köken gösteriminin bu bağlantıları ele alması gerekir. Süreçlerin ve bağlantılarının detaylı şekilde ele alınması köken kayıtlarına göre sürecin tekrar çalıştırılarak artifaktların yeniden üretilebilirliğine olanak sağlar. Örneğin, köken kayıtlarına uygun süreç geçişlerine göre bir algoritmanın sürekli güncellenen verisetleri ile çalıştırılabilmesi bu şekilde sağlanır.

Süreç köken bilgisi, atfetme bilgisi ile bağdaştırılmalıdır. Böylece kaynakların (veya varlıkların) hangi aktivitelerde kullanıldığı ve bu aktivitelerde hangi amaca hizmet ettiği hakkında bilgi sahibi olunabilir. Aktivitede bir veri seti algoritmaya girdi olarak verilirken, bu algorithmadaki parametreleri veren kişinin de aktiviteyle ilişkilendirilmesi gerekir.

Süreç kökenini temsil ederken kaynak erişimi de önemlidir. Erişim zamanı, erişilen sunucu veya sunucudan sorumlu olan başka birimler gibi bilgiler köken açısından sorgulanması gereken ve güven gibi birçok kavramı etkileyen faktörlerdir. Haber Birleştiricisi Senaryosu'nda yayımlanan haberlerin erişilme zamanına göre, kullanıcının eriştiği haberde değişiklik yapıp yapılmadığına dair bilgi sahibi olunabilir.

İçerik/Versiyonlama

Artifaktlar zamanla evrim geçirdiğinden dolayı kökeni hakkında bilgi farklı yollarla arttırılmalıdır. Önceki versiyonlar üzerinde yapılan değişiklikler ve hangi süreçlerin bu değişikliklerle ilgili olduğu özellikle vurgulanmalıdır. Veb gibi bir artifakt üzerinde tamamıyla kontrol sahibi olunmasının zor olduğu açık ve dağıtık sistemlerde versiyonlama köken bilgisi daha da önem kazanmaktadır. Bir haberin tekrar yayımlanması, bir projenin tekrar paketlenerek yüklenmesi gibi durumlarda varlıkların yeni versiyonuna gelene kadarki her değişimin ifade edilmesi gerekir. Yeni versiyonun yayımlanması sırasında kaynakla beraber köken üstverisinin

değişikliklerini içeren güncel halinin yayımlanması değişim aşamalarının takibi için önemlidir. Fakat yayımlanan köken verisi ile beraber gizlilik de endişe duyulması gereken kavramlar olarak ortaya çıkar. Örneğin bir yazılımcının bir şirkete ait projede kullandığı kod parçasını başka bir şirkete ait projede kullanıp tekrar kullanabilirliğe dair bir köken tanımı yüklediğinde önceki projeyle ilgili gizliliğe dair bir ihlalde bulunmuş olur.

Kaynak aynı kalmakla beraber yayımlanan veb sayfasının linki veya sayfa mimarisi ile ilgili değişiklikler yapılabilir. Dolayısıyla kaynak üzerinde veb mimarisi ile ilgili ifadeler kullanılarak kullanıcının değişim hakkında detaylı bilgilendirilmesi sağlanmaya çalışılmalıdır.

Aşağıda versiyonlama ile ilgili Haber Toplama Senaryosu'nda karşımıza çıkan durumlar görülmektedir.

- Bir sistem yayımlanan dokümanın veya dokümandaki bir parçanın farklı versiyonlarının izini sürebilmelidir. Doküman veya veri üzerindeki değişiklikler belirlenmeli ve kaydedilmelidir.

- Bir sistem dokümanlar üzerinde kimin değişiklik yaptığının izini sürebilmelidir. Doküman veya verinin farklı versiyon oluşumunun nedeni belirtilmelidir.

- İçerik değiştiği anda bildirebilecek bir mekanizma olmalıdır.

- Köken, verinin farklı versiyonlarına erişebilmek amacıyla sorgulanabilmelidir.

- Versiyonlama bilgisi, farklı sistemler arasında birlikte çalışabilir şekilde oluşturulmalıdır.

İçerik/Gerekçe

Amacı köken üzerinde çıkarsama yapılarak verilen kararların anlaşılabilmesi ve üzerinde tartışılabilmesidir. Örneğin, bir konuda yayımlanan haberle ilgili olarak o konunun uzmanları ilgili kanıtlar göstererek olumlu veya olumsuz yorumlar yaptıysa haber hakkında daha detaylı bir fikir edinilebilir. Ayrıca;

- Bir sistemin, doğru davranışının tanımlanması adına farklı taraflarca kabul edilen bir mekanizma geliştirilmelidir. Bu davranış köken yardımıyla kontrol

edilebilmelidir.

- Sistem tarafından sağlanan köken desteği erişilebilir olmalıdır ve iyi bilinen bir formatta yayımlanabilmelidir.

- Öncül iki maddenin otomatikleştirilmesi ile birlikte düzenlemelere uygunluğun sağlanmasına dair çabanın azaltılması amaçlanmaktadır. Böylece köken bilgisi farklı kullanıcı veya sistemler arasında işbirliğine olanak sunabilir.

İçerik/Gereklilik

Köken bilgisinin bir kısmı veri kaynakları veya bir süreç içerisindeki aktörler tarafından doğrudan elde edilirken, diğer kısmı ise elde edilenler üzerinden türetilir. Bu durum, ardıl gerçeklerin öncül gerçeklerin gerekliliği doğrultusunda oluşmasını ifade etmektedir. Öncül gerçekler arasında bulunan uyuşmayan kayıtlar üzerinde yapılan kalite bazlı çalışmalar sonucu sistemin kalitesini artırıcı sonuçlar ortaya çıkarılabilmekte ve sistemlerin kullanımı kolaylaştırılmaktadır. Örneğin bir haber sitesinin haberleri kategorize etmesi (InternationalPressTelecommunicationCouncil, 2014) ile ilgili olarak eski kayıtlar üzerinden bir sistematik oturtularak bu model üzerinde haberlerin siteye yüklenmesi ve takibi sağlanabilmektedir (Tenenboim et al., 2008).

Yönetim/Yayımlama

İnternet gibi açık sistem bir ağda köken kullanılabilirliği ve erişilebilirliğini sağlayan mekanizmalardır. Bu mekanizmalarda yayımlama önemlidir çünkü bu durum kullanılabilirliği göstermektedir. Köken bilgisinin nasıl keşfedildiği veya dağıtıldığı gibi durumlara ilişkin köken gösterim diline ihtiyaç duyulmaktadır.

Yönetim/Erişim

Köken bilgisi kullanılabilir olduğu anda erişilebilir olması gerekmektedir. Erişilebilmede kastedilen köken bilgisinin bulunması ve sorgulanabilmesidir. Örnek olarak Haber Birleştiricisi Senaryosu incelendiğinde spesifik bir habere ilişkin köken bilgisi elde edilebilmelidir. Bazı durumlarda otoriter köken kaynağının varlık sınıfları ile ilişkileri belirlenmelidir. Örneğin, uygulamayı geliştiren kurumlar kendi kayıtlarını dahili süreçlerde otoriter kaynak olarak gösterebilir veya ihtiyaç duyduklarında güvendikleri otoritelerin kaynaklarına erişebilir. Bu noktada köken bilgisinin gösterimi için sorguların formülasyonu ve işletilme mekanizmaları tanımlanmalıdır.

Yönetim/Yayılma

Gerçekçi koşullar altında köken bilgisinin yayılma kontrolünün sağlanması gerekir. Varlığın yaratıcısı tarafından ne zaman ve nasıl kullanıldığı hakkında kurallar izlenmeli, erişim kontrolü, lisanslama ve kuralların uygulanması gözönüne alınmalıdır. Köken bilgisi, erişim kontrol mekanizmaları ile ilişkilendirilerek kökenin talep eden kişi tarafından hangi durumlarda kullanılabileceğini gösterir. Kullanım şartlarına bağlı olarak artifaktın köken bilgisinin nasıl verileceğine dair bilgiye de ihtiyaç duyulabilir. Bu durumda artifaktın yaratıcıları tarafından sağlanan kullanım hakları lisanslama bilgisi ile verilebilir ve böylece verinin yayılması ile ilgili düzenlemeler gerçekleştirilebilir.

Yönetim/Ölçek

Ağ günceleri veya diğer veb içeriğinin büyük ölçekteki verisi için kökenin izlenmesi sağlanabilmelidir. Artan sayıdaki veri ile ölçeğin üst seviyede tanelilik özelliğine göre köken izleme araçları kullanılabilmelidir. Köken izleme ile ilgilenirken vebden gelen birleştirilmiş verinin mahremiyetinin ihlal edilmesini önleyecek çözümler üzerinde düşünülmelidir. Çok fazla köken kaydı olması durumunda verimli bir şekilde bu kayıtlara erişimin mümkün olması gerekmektedir (Moreau et al., 2012).

Kullanım/Anlaşılrlık

Anlaşılrlık, köken bilgisinin son kullanıcı tarafından tüketilmesini sağlamaktadır. *Soyutlama* ve *sunum* olarak iki başlık altında incelenir. *Soyutlama*, anlaşılrlığı arttırmak için kullanıcılara özet bilgi verme niteliğinde olup tanımların çeşitli seviyelerde yapılmasını sağlar. Çeşitli seviyelerde soyutlama yapılarak artifakta ait köken kayıtlarını çeşitli açılardan incelenme şansını sunar. Anlaşılrlığı arttırmak için çalışmada kullanılan PROV-O ontolojisi gibi bilinir bir model içerisindeki elemanlar üst seviye elemanlar şeklinde ele alınır ve çalışma alanına dair detaylar geliştirilen sisteme eklenebilir. *Sunum* ise köken bilgisinin görselleştirilerek son kullanıcılar tarafından kolaylıkla anlaşılmasını amaçlar.

Kullanım/Birlikte Çalışabilirlik

Birden fazla sisteme ait köken bilgisinin birbirleriyle bütünleştirilmesini sağlar. Heterojen olarak dağılmış farklı gösterimlere sahip köken bilgisinin elde edilerek ortaklaşa çalışır hale getirilmesi önemli bir gereksinimdir. Dönüşümler gerçekleştirilerek bütünleştirilmiş bilgi kümesinin sorgulanması ve gerekli bilginin

bu kümeden çekilmesine ihtiyaç duyulmaktadır. Haber Toplama Senaryosu ele alındığında New York Times'ta bulunan bir kişi (kavram) ile ilgili bir haberin The Guardian gazetesindeki aynı kişinin resmine dair köken bilgisi ile bütünleştirilerek o kişiye ait daha detaylı ve kaliteli bilgi edinilmesi sağlanabilir. Hangi köken bilgisinin hangi kaynaktan desteklendiği ve bu kaynaklara dair güven bilgisi sistemde bütünleştirilmiş güvenilir bir veri topluluğu elde edilmesini sağlar. Eğer farklı sistemler farklı sözlükleri kullanıyorsa da ortak eşlemeler ve standartlaşmış sözlüklerle kullanılabilirlik ve birlikte çalışabilirlik imkanı sunulabilmelidir.

Kullanım/Karşılaştırma

Geçmiş süreçler veya olaylar arasında benzerlik ve farklılıkları belirlemeyi kullanıcıya sunmayı hedeflemektedir. Karşılaştırılabilir iki süreç için bir mekanizma gerçekleştirilmesi ve bu mekanizmanın benzerlik ve farklılıkları kullanıcıya yardımcı olan bir anlatımla ifade etmesi gerekmektedir.

Kullanım/İzlenebilirlik

Bir nesnenin kökeninin beklentilere göre kontrol edilebilirliğini ifade etmektedir. Örneğin kullanıcılara yapılan sözleşmede yer alan işin kararlaştırıldığı süreden önce yapıp yapılmadığını gözlemlemek için faydalıdır. Ayrıca köken kayıtları tutularak hangi gerekliliklerin karşılandığı da izlenebilir. Kaynaklar için yetki verilmesi ve kullanıcıların bu yetkiler dahilinde köken bilgisine bağlı kalması açısından izlenebilirliğe ihtiyaç duyulur. İzlenebilirlik, bırakılan köken izleri ya da birden fazla bağımsız köken kararı birleştirilerek sağlanır.

Kullanım/Güven

Güven, *değerlendirme* ve *ilişkilendirme* olmak üzere farklı iki kapsamda ele alınmaktadır. *Değerlendirme*, veb verisinin güvenilirliğinin değerlendirilmesi için kullanıcılara olanak sağlar (Bizer and Cyganiak, 2009; Ramachandran et al., 2009). Kaynak veriye göre güvenin hesaplanması, verinin farklı versiyonları üzerinden giderek güvenin hesaplanması, eksik köken bilgisi ile karşılaşıldığında bağlı veri sorgulama ile ilişkili düğümlere erişilerek güvenin hesaplanması bu kapsamdadır. *İlişkilendirme* ise üstveri kullanarak veb tabanlı verinin güven değerlendirmesi için kullanıcılara olanak sağlar.

Kullanım/Kusurluluk

Doğru veya tam olmayan köken bilgisi hakkında akıl yürütmeye ilgilendir.

Yayımlanan köken bilgisinin hatalı ya da hileli olarak sunulabileceği gözönüne alınmalıdır. Köken bilgisi yetersiz olabileceği gibi belirsizliğe de sahip olabilir, bu durumda analiz edilerek kusurlu verilerin ayıklanması gerekir. Köken bilgisinin kayıtlanma aşamasında meydana gelebilecek kusurların da üzerinde durulması gerekebilir. Kusurluluk, köken bilgisinin özetlendiği ya da sıkıştırıldığında karşılaşılabileceği gibi detayların silinmesi ya da soyutlanmasından da kaynaklanabilir. Ayrıca kullanıcı bazı köken kayıtlarına erişim iznine sahip değilse de yetersizlik ve dolayısıyla kusurluluk durumuyla karşılaşabilir.

Kullanım/Hata Ayıklama

Köken bilgisindeki başarısızlık durumlarını tespit edebilmeyi amaçlar. Üretilmiş bir artifaktın işleme sürecinde bu başarısızlıkların ayıklanarak iyileştirilmesinin sağlanması amaçlanmaktadır. Bir kaynağın kullanımı sonrasında meydana gelecek birden fazla köken artifaktının analiz edilerek ne kadar kusurlu köken verisi üretildiğine dair bir mekanizma geliştirilmesi gerekmektedir.