Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Bilgi ve Belge Yönetimi Anabilim Dalı

TÜRKİYE'DE KURUMSAL İŞLEYİŞ VE YÖNETSEL YAPI ÇERÇEVESİNDE VERİ GÖLÜ TABANLI VERİ MİMARİLERİNİN GELİŞTİRİLMESİ

Ela ANKARALI

Doktora Tezi

Ankara, 2025

TÜRKİYE'DE KURUMSAL İŞLEYİŞ VE YÖNETSEL YAPI ÇERÇEVESİNDE VERİ GÖLÜ TABANLI VERİ MİMARİLERİNİN GELİŞTİRİLMESİ

Ela ANKARALI

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Bilgi ve Belge Yönetimi Anabilim Dalı

Doktora Tezi

Ankara, 2025

KABUL VE ONAY

Ela Ankaralı tarafından hazırlanan “Türkiye’de Kurumsal İşleyiş ve Yönetsel Yapı Çerçevesinde Veri Gölü Tabanlı Veri Mimarilerinin Geliştirilmesi” başlıklı bu çalışma, 06.01.2025 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.

Prof. Dr. Fahrettin Özdemirci (Başkan)

Prof. Dr. Özgür Külcü (Danışman)

Prof. Dr. Nevzat Özel (Üye)

Doç. Dr. Gökhan Akçapınar (Üye)

Doç. Dr. Tolga Çakmak (Üye)

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Uğur ÖMÜRGÖNÜLŞEN

Enstitü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan ***"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ….. ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

……/……/……

Ela ANKARALI

[1] *"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"*

(1) *Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu*** *iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) *Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulunun*** *gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) *Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı,* ***tezin yapıldığı kurum*** *tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise,* ***ilgili kurum ve kuruluşun önerisi*** *ile* ***enstitü*** *veya* ***fakültenin*** *uygun görüşü üzerine* ***üniversite yönetim kurulu*** *tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.
****** Tez* ***danışmanının*** *önerisi ve* ***enstitü anabilim dalının*** *uygun görüşü üzerine* ***enstitü*** *veya* ***fakülte yönetim kurulu tarafından karar verilir.***

ETİK BEYAN

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, **Prof. Dr. Özgür KÜLCÜ** danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Yazım Yönergesine göre yazıldığını beyan ederim.

Ela ANKARALI

TEŞEKKÜR

Bu tez çalışmasının hazırlanmasında ve tamamlanmasında emeği geçen, bana yol gösteren, destek veren ve sabırla yanımda olan herkese içtenlikle teşekkür ederim.

Öncelikle değerli danışmanım Prof. Dr. Özgür Külcü'ye akademik rehberliği, desteği ve bana olan inancı için minnettarım. Onun yol göstericiliği, teşvik edici katkıları ve desteği, bu çalışmanın şekillenmesinde büyük rol oynamıştır.

Tez izleme komitemde yer alan saygıdeğer hocalarım Prof. Dr. Fahrettin Özdemirci ve Doç. Dr. Gökhan Akçapınar'a süreç boyunca yaptıkları değerli yorumlar ve akademik katkıları için teşekkür ederim. Jüri üyelerim Prof. Dr. Nevzat Özel ve Doç. Dr. Tolga Çakmak'a da değerlendirme sürecinde sundukları kıymetli görüşler ve katkıları için teşekkürlerimi sunarım.

Dr. Meltem Dişli'ye her zaman sorularımı sabırla yanıtladığı ve desteğiyle bana yardımcı olduğu için teşekkür ederim. Ayrıca iş arkadaşlarım Ayşe Merve, Muhlis, Pınar, Sevim ve Setenay'a verdikleri moral ve destek için içtenlikle teşekkür ederim.

Bu uzun ve zaman zaman zorlayıcı süreçte, en büyük desteği ailemden aldım. Sevgili eşim Mert Ankaralı'ya bana gösterdiği sabır, anlayış ve sonsuz desteği için minnettarım. Biricik kızım Ezgi Pırıl'a ise kocaman bir teşekkür borçluyum. Yoğun geçen bu süreçte bana her zaman sevginle güç verdin, kocaman gülümsemen ve sıcacık kucaklamalarınla tüm yorgunluğumu unutturdun. Senin varlığın bu süreci daha güzel hale getirdi.

Canım annem, babam ve Elvan bana her zaman inandığınız, hiçbir koşulda desteğinizi esirgemediğiniz ve her zaman yanımda olduğunuzu hissettirdiğiniz için teşekkür ederim.

ÖZET

ANKARALI, Ela. *Türkiye'de Kurumsal İşleyiş ve Yönetsel Yapı Çerçevesinde Veri Gölü Tabanlı Veri Mimarilerinin Geliştirilmesi*, Doktora Tezi, Ankara, 2025.

Dijital dönüşüm ve veri bilimi alanlarındaki gelişmeler, büyük veri yönetimi ve analizine olan ihtiyacı artırmıştır. Bu tezde, Türkiye'deki kamu kurumlarının büyük veri yönetim süreçlerini daha etkili bir şekilde yönetebilmeleri amacıyla fonksiyonel ve olgunluk tabanlı bir veri gölü mimarisi modeli önerilmiştir. Büyük veri çağında, kamu kurumları tarafından toplanan verilerin analiz edilerek toplumsal faydaya dönüştürülmesi büyük önem taşımaktadır. Mevcut veri yönetim sistemlerinin, büyük veri çağının gereksinimlerini karşılamada yetersiz kaldığı görülmektedir. Önerilen model, veri gölü mimarilerinin esnek yapısını temel alarak, kamu kurumlarının veri yönetim süreçlerini iyileştirmeyi ve büyük veri analitiği kapasitesini artırmayı hedeflemektedir. Tez kapsamında, ulusal ve uluslararası düzeyde başarılı veri yönetim sistemleri, ilgili standartlar ve politikalar incelenmiş; kamu kurumlarının ihtiyaçlarına uygun bir model geliştirilmiştir. Model, verilerin toplanması, temizlenmesi, organize edilmesi, depolanması ve analiz edilmesi süreçlerini sistematik bir yapıya oturtarak verinin yaşam döngüsü süresince veri göllerinin etkin kullanımını sağlamayı hedeflemektedir. Önerilen mimari, veri güvenliği, üst veri yönetimi ve yasal uyumluluk konularının önemini vurgulamaktadır. Veri entegrasyonu, veri yönetişimi ve veri paylaşımı süreçlerini optimize ederek, kamu kurumlarının veri yönetim süreçlerinin verimliliğini artırmayı hedeflemektedir. Bu kapsamda, kurumsal veri süreçleri için standardizasyon, teknik altyapı seçimi, güvenlik önlemleri ve uzun vadeli sürdürülebilirlik konularını içeren bir uygulama modeli sunulmuştur. Önerilen modelin, Türkiye'deki kamu kurumlarının mevcut veri yönetim altyapısını geliştirerek ulusal düzeyde bir veri mimarisi oluşturulmasına temel oluşturabileceği değerlendirilmektedir.

Anahtar Sözcükler

Kurumsal veri, veri mimarisi, veri gölü, veri yönetimi

ABSTRACT

ANKARALI, Ela. *Development of Data Lake Based Data Architectures within the Framework of Institutional Operations and Administrative Structure in Türkiye*, Ph. D. Dissertation, Ankara, 2025.

The advancements in digital transformation and data science have increased the need for big data management and analysis. This thesis proposes a functional and maturity-based data lake architecture model to enable public institutions in Türkiye to manage big data processes more effectively. In the era of big data, it is crucial to analyze the data collected by public institutions and transform it into societal benefits. Existing data management systems have been found inadequate in meeting the requirements of the big data era. The proposed model aims to improve data management processes and enhance big data analytics capacity in public institutions by leveraging the flexible structure of data lake architectures. The thesis examines successful data management systems, relevant standards, and policies at national and international levels, and develops a model tailored to the needs of public institutions. The model aims to ensure the effective use of data lakes throughout the data lifecycle by systematically structuring the processes of data collection, cleaning, organization, storage, and analysis. The proposed architecture emphasizes the importance of data security, metadata management, and legal compliance. It seeks to optimize data integration, data governance, and data sharing processes, thereby enhancing the efficiency of data management processes in public institutions. In this context, an implementation model covering standardization, technical infrastructure selection, security measures, and long-term sustainability for institutional data processes has been presented. The proposed model is considered capable of improving the existing data management infrastructure of public institutions in Türkiye and serving as a foundation for the development of a national-level data architecture.

Keywords

Institutional data, data architecture, data lake, data management

İÇİNDEKİLER

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

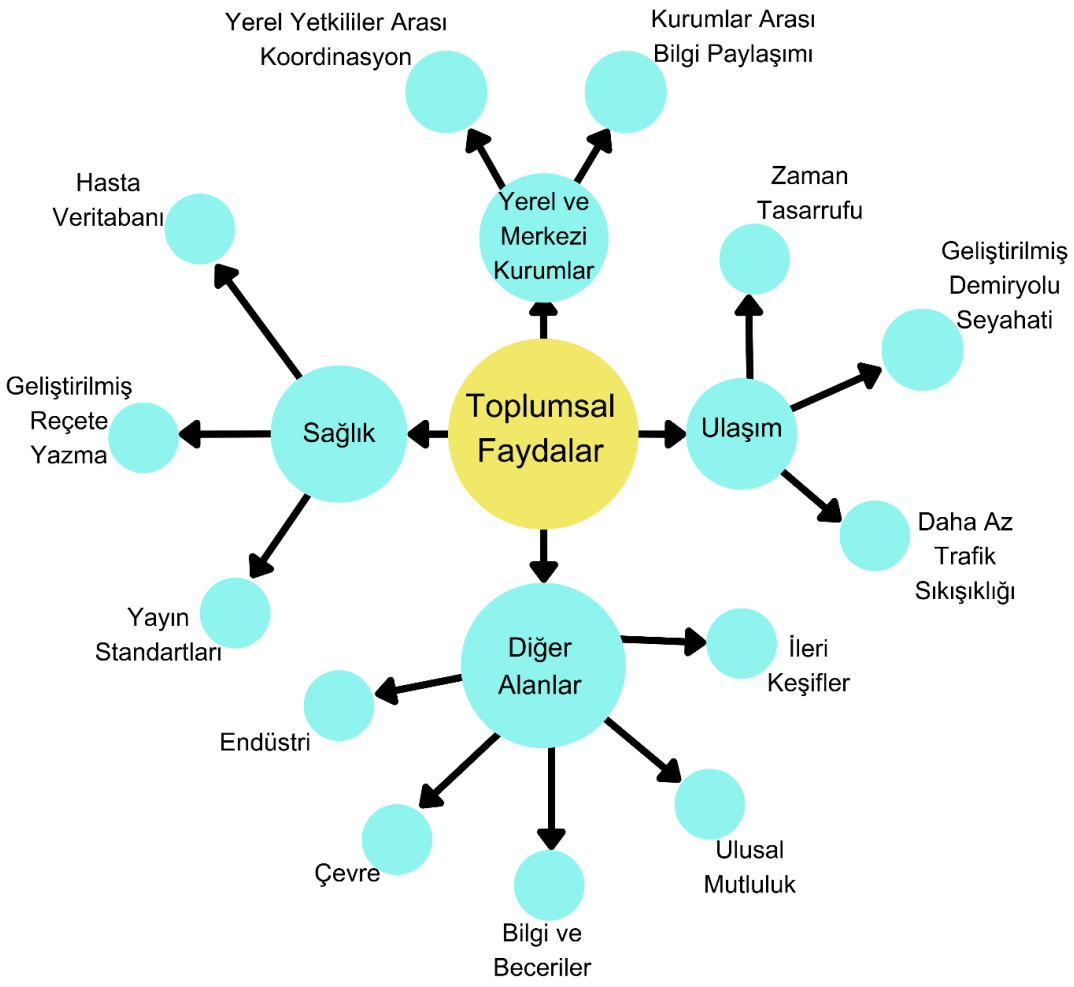
1. BÖLÜM
GİRİŞ

1.1. KONUNUN ÖNEMİ

Günümüzde dijital dönüşüm ve veri bilimi konularında belirgin gelişmeler yaşanmaktadır. Bu gelişmelere paralel olarak ortaya çıkan büyük veri kavramı, tam olarak ne zaman ortaya çıktığı bilinmese de 2011 yılından itibaren yaygın olarak kullanılmaya başlanmıştır (Eyüpoğlu, 2018, s. 4). Dijital dönüşümle birlikte hızla artmaya başlayan büyük veri, dokümanlardan veri toplama ve bu veriyi işleyerek anlamlı sonuçlar ortaya çıkarma ihtiyacını doğurmuştur (Külcü, 2007; Külcü, 2010). Bu bağlamda, yapay zekâ, makine öğrenmesi ve veri madenciliği araçları hemen hemen bütün bilim dallarında aktif olarak kullanılan yöntemler haline gelmiştir (LeCun, Bengio ve Hinton, 2015). Büyük verinin analizinde veriden anlamlı sonuçlar elde etmek için verinin belirli özellikleri taşıması gerekmektedir. Veri ancak makinelerce anlaşılabilir olduğunda yapay zekâ teknolojileri ile işlenebilmesi, anlaşılması ve yorumlanması mümkün olmaktadır (Bhimanpallewar ve diğerleri, 2023). Farklı bilimlerde hızla artan büyük veri, açık veri kavramının ortaya çıkmasını sağlamıştır. Verinin kullanılarak anlamlandırılması ancak verinin kullanıma açık olması ile mümkün olmaktadır. Büyük veri, çeşitli analiz yöntemleri ve teknolojilerle anlamlandırılmakta ve birçok alanda değer yaratmaktadır. Açık veri kavramı ise bu süreçte şeffaflık, inovasyon ve iş birliğini teşvik eden önemli bir unsur olarak öne çıkmaktadır.

Kamu kurumları sundukları hizmetlerle yüksek miktarda veri toplamakta ve oluşturmaktadır (Topçu ve Işık, 2019, s. 834). Kamu verisi kamu kurumlarının ürettiği ve topladığı her tür veri olarak tanımlanabilir ve genellikle büyük hacimli ve farklı türdeki verilerden oluşur (Dekkers, Polman, Velde ve Vries, 2006, s. 25).

Kamu kurumlarının barındırdığı veriden toplumsal fayda elde edilmesi için verinin kurum içinde etkili bir şekilde kullanılması ve uygun koşullarda toplumsal kullanıma açılması gerekmektedir. Kamu kurumları, veriyi daha etkin kullanarak ve paylaşarak toplumsal faydayı artırabilirler. Ayrıca kamu verisinin güvenilir ve sürdürülebilir bir platformda yönetilmesi, verinin topluma faydalı bir şekilde sunulabilmesi için kritik öneme sahiptir (Özdemirci, 2019a, s. 170). Genel çerçevesiyle kamu bilgisinin sağlayacağı toplumsal faydalar Şekil 1'de verilmiştir (Deloitte, 2013, s. 112). Bu süreç, verinin toplanması, işlenmesi, paylaşılması ve analiz edilmesini içeren bir dizi alt süreci gerektirir.



Şekil 1. Kamu bilgisinin toplumsal faydaları (Deloitte, 2013, s. 112)

Kamu bilgisi sağlık alanında hasta veri tabanı, geliştirilmiş reçete yazma, yayın standartları; yerel ve merkezi kurumlar arası bilgi paylaşımı ve koordinasyon;

ulaşım alanında zaman tasarrufu, geliştirilmiş seyahat, daha az trafik sıkışıklığı; ayrıca endüstri, çevre, bilgi ve beceriler, ulusal mutluluk ve ileri keşifler gibi diğer alanlarda toplumsal fayda sağlayacak potansiyele sahiptir. Örneğin Millî Eğitim Bakanlığı'nın öğrenci başarı verileri ile Türkiye İstatistik Kurumu'nun sosyo-ekonomik durum verileri birlikte değerlendirilerek bölgesel eğitim eşitsizlikleri belirlenebilir ve kaynakların daha etkili tahsis edilmesi sağlanabilir. Yine mezun olan öğrencilerin istihdam oranlarının analiz edilmesiyle eğitim-iş gücü uyumunun artırılması ve genç işsizliğiyle mücadele konusunda daha hedefe yönelik politikalar geliştirilebilir. Emniyet Genel Müdürlüğü'nün suç kayıtları ile Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü'nün yerleşim verileri birleştirilerek, suçla mücadelede hangi bölgelere öncelik verileceği belirlenebilir, daha etkili politikalar geliştirilebilir.

Bu potansiyel faydaların sağlanabilmesi için veri depolama ve yönetim sistemlerinin sürekli olarak geliştirilmesi gerekmektedir. Depolama sistemlerine sürekli yeni dosya biçimleri ve yeni nesneler eklenmektedir. Veri biliminde geçmişte küçük ve yapılandırılmış ilişkisel veri tabanları ve dosyalar kullanılmaktayken günümüzde büyük veri öne çıkmaktadır. "Büyük veri" (big data), geleneksel veri işleme yöntemlerinin yetersiz kaldığı ölçüde geniş hacimde, yüksek hızda üretilen ve farklı veri türlerini barındıran veri kümelerini ifade etmektedir (Çiftçi ve Gülsoy, 2020). Ancak bu verinin çoğunluğu hala kaotiktir. Gelecekte ise bu verilerin etiketlenmesi ve işlenmesi ile "Akıllı Kuruluş"lar öne çıkacaktır:

- Geçmiş: İlişkisel veri tabanları ve dosyalar (Küçük ve yapılandırılmış veri)
- Günümüz: Büyük veri (Büyük ve kaotik veri)
- Gelecek: "Akıllı kuruluşlar" (Büyük ve işlenebilen veri)

(Bell, 2018, s. 5).

Bu gelişmeler ışığında, verinin daha iyi anlaşılması ve kullanılabilmesi için veri platformlarının önemi artmaktadır. Veri platformları, üst veri (metadata) aracılığıyla veri kaynağını tanımlayarak ve açıklayarak, verinin anlaşılmasını,

yorumlanmasını ve yönetimini kolaylaştırır. Üst veri, bilgilerin yapılandırılmasını ve makine tarafından işlenmesini, böylece anlamlandırılmasını ve verilerle ilgili bağlamların ortaya çıkarılmasını sağlar. Bu şekilde verilerin kullanılmasına ve yorumlanabilmesine olanak tanır (Atlan, 2024a; Kaisler, Armour, Espinosa ve Money, 2013).

Veri paylaşımı ve depolama amacıyla kullanılan farklı veri platformlarının kullanılabilirlik özellikleri, güçlü ve zayıf yönleri farklılık göstermektedir. Verilerin tam olarak kullanılabilmesi için verilere erişilebilmesine ek olarak verilerin anlaşılabilmesi gerekir. Bunun sağlanabilmesi için veri platformlarının kullanılabilirliğinin ve farklı veri mimarilerinin özelliklerinin değerlendirilmesi önemlidir (Nikiforova ve McBride, 2021, s. 1).

Büyük veri çağında, kuruluşlar artan veri hacimlerini yönetmek ve analiz etmek için çeşitli veri mimarilerine yönelmişlerdir. Bu mimariler arasında en yaygın olarak kullanılanlardan biri veri ambarlarıdır. Kimball ve Ross (2013) veri ambarlarını, farklı kaynaklardan gelen verileri entegre eden, tarihsel ve tutarlı bir veri deposu sağlayan merkezi sistemler olarak tanımlamaktadır. Bu sistemler, işletmelerin karar verme süreçlerini desteklemek için tasarlanmış olup genellikle yapılandırılmış verileri saklar ve analiz eder. Inmon (2005) veri ambarlarının, verileri önceden tanımlanmış bir şemaya göre düzenlediğini ve depoladığını belirtmektedir. Bu "schema-on-write" (şemayı yazarken tanımlama) yaklaşımı, veri kalitesini ve tutarlılığını sağlamada etkili olsa da yeni veri türlerinin hızla entegre edilmesini zorlaştırabilmektedir. Ayrıca veri ambarları genellikle yüksek maliyetli altyapı yatırımları gerektirmek ve büyük ölçekli, yapılandırılmamış verileri işlemekte zorlanabilmektedir. Bu sınırlamalar, özellikle büyük veri çağında, alternatif veri mimarilerine olan ihtiyacı artırmıştır. Miloslavskaya ve Tolstoy (2016), veri göllerinin, veri ambarlarının bu kısıtlamalarına bir yanıt olarak ortaya çıktığını vurgulamaktadır. Veri gölleri, veri ambarlarının aksine, ham verileri orijinal formatlarında saklayabilmeyi mümkün kılan daha esnek bir veri yönetimi yaklaşımı sunmaktadır.

Veri göllerinin esnekliği ve sunduğu potansiyel, Türkiye'deki veri yönetim sistemleri için büyük bir fırsat oluşturmaktadır. Tüm dünyada olduğu gibi Türkiye'de de kamu ve özel sektör kurumları giderek daha fazla veri üretmekte ve bu verilerin etkin bir şekilde yönetilmesi ve analiz edilmesi büyük bir gereklilik haline gelmektedir. Ancak Türkiye'de kullanılan mevcut veri yönetim sistemleri, büyük veriyi yönetmek için yeterli kapasiteye sahip değildir ve bu durum büyük veri çağında rekabetçi kalmayı zorlaştırmaktadır. Türkiye İstatistik Kurumu'nun (TÜİK) büyük veri ve ileri analitik konularında yeni projeler geliştirdiği görülmektedir ("B3LAB", t.y.). Bu durum, TÜİK'in büyük veri yönetimi konusundaki çabalarını ve mevcut altyapıyı geliştirme ihtiyacını göstermektedir. Özellikle kamu kurumları, sağlık, eğitim, finans gibi kritik sektörlerde veri göllerinin sunduğu esneklikten ve maliyet etkinliğinden faydalanarak daha verimli veri yönetim süreçleri geliştirebilirler (Fang, 2015). Veri göllerinin bu potansiyeli, yalnızca verilerin saklanması değil, aynı zamanda gerçek zamanlı analizler ve çapraz veri analizleri yapma imkânı sunmasıyla Türkiye'deki veri mimarisi gelişimine büyük katkı sağlayabilecek potansiyele sahiptir (Fang, 2015; John ve Misra, 2017).

Dünyada kamu ve özel sektör kurumlarında veri gölü kullanımı oldukça yaygındır. ABD Nüfus Bürosu, veri gölü altyapısı sayesinde büyük hacimli nüfus verilerini gerçek zamanlı olarak işleyerek ülke genelinde daha hızlı ve doğru karar alma süreçlerini desteklemektedir (U.S. Census Bureau, 2022). Benzer şekilde, Netflix ve Uber gibi küresel ölçekte hizmet veren şirketler, farklı kaynaklardan gelen veri akışlarını veri gölü altyapısında toplayarak kişiselleştirme, dinamik fiyatlandırma ve operasyonel iyileştirmeler için gerçek zamanlı analizler gerçekleştirebilmektedir. Nestlé, geniş tedarik zinciri ve satış noktalarından oluşan veri ekosistemini tek noktada yönetmek amacıyla veri gölü teknolojileri kullanırken Accenture ve Capital One gibi firmalar, artan veri hacimlerini daha etkin şekilde işleyip finans ve danışmanlık süreçlerini zenginleştirmeyi hedeflemektedir (Monte Carlo, t.y.). Ayrıca Amazon Web Services (AWS), Microsoft Azure ve Google Cloud gibi bulut hizmeti veren şirketler, küresel ölçekte veri gölü hizmeti sunarak finans, sağlık, perakende, eğitim ve araştırma gibi pek

çok alanda veri analitiği, iş zekâsı ve yapay zekâ çözümlerinin geliştirilmesine katkı sağlamaktadır.

Türkiye'de kamu kurumlarının büyük veriyi daha etkin kullanarak toplumsal fayda yaratmaları, yapay zekâ, robotik süreç otomasyonu ve kuantum teknolojileri gibi yenilikçi çözümlere geçişle mümkün olacaktır (Özdemirci, 2019a). Veri gölleri, bu teknolojilere geçiş ve uygulama süreçlerinde iyi bir veri yönetimi altyapısı sağlamaktadır ve kurumlar tarafından dikkatlice araştırılmalı ve değerlendirilmelidir.

Bu bağlamda, “Kamu Veri Alanı Projesi” (Türkiye Yüzyılı, 2022) de kamu kurumlarının verilerini tek çatı altında toplama ve yönetme hedefiyle hayata geçirilmiştir. Projenin 2025 yılında tamamlanması hedeflenmiş olmakla birlikte, sağladığı yenilikler ve kaydedilen aşamalara ilişkin bilgiler henüz sınırlıdır. Ancak projenin varlığı, Türkiye’de kamu veri yönetimi alanında yenilikçi yaklaşımlara duyulan ihtiyacı göstermesi bakımından önemlidir. Kamu Veri Alanı Projesi, kamu kurumları arasında güvenilir ve “veri egemenliğini” garanti edecek şekilde veri depolama, paylaşma ve işleme olanağı tanıyan gelişmiş standart ve yönergelerle uyumlu bir çerçeveye sahiptir. Bu sayede, farklı kamu kurumlarının elindeki veri kümeleri, mahremiyet ve güvenlik esasları doğrultusunda bir araya getirilebilecek ve birlikte analiz edilebilecektir. Proje, Avrupa Birliği’nin “veri uzayları” (*data spaces*) yaklaşımından ilham alarak, mahremiyet ihlali ve siber güvenlik risklerini en aza indiren kontrollü ve izlenebilir bir veri paylaşım altyapısı oluşturmayı hedeflemektedir. Bunun da ötesinde Kamu Veri Alanı, ileri analitik ve yapay zekâ uygulamalarının geliştirilmesi için gerekli yüksek kaliteli veriye erişimi kolaylaştırarak, kamunun “veri temelli organizasyona” dönüşüm sürecini hızlandırmayı amaçlamaktadır. Proje kapsamında referans mimari, gerekli mevzuat düzenlemeleri ve insan kaynağı planlamaları ele alınmakta olup, Cumhurbaşkanlığı Dijital Dönüşüm Ofisi koordinasyonunda yürütülmektedir. TÜİK, TÜBİTAK ULAKBİM, TÜBİTAK BİLGEM ve Çevre, Şehircilik ve İklim Değişikliği Bakanlığı ile ilgili çalışma gruplarının iş birliği, projeye kurumsal ve teknik altyapı desteği sağlamaktadır.

Tüm bu yaklaşımlar göz önüne alındığında, veri gölleri, “Kamu Veri Alanı Projesi” gibi ulusal ölçekli girişimlerde farklı veri kaynaklarının esnek, ölçeklenebilir ve güvenli biçimde yönetilmesi için önemli bir dayanak noktası oluşturmaktadır. Veri gölleri, büyük verinin çok yönlü ihtiyaçlarına cevap verirken, projenin hedeflediği veri egemenliği, mahremiyet ve siber güvenlik ilkeleriyle de uyumlu bir mimari sunmaktadır. Dolayısıyla, Kamu Veri Alanı kapsamında planlanan referans mimaride veri gölü teknolojilerinin değerlendirilmesi hem kısa vadede ileri analitik uygulamalarının geliştirilmesine hem de uzun vadede sürdürülebilir bir “veri temelli kamu” ekosisteminin inşasına değerli bir katkı sağlayacaktır.

Veri göllerinin Türkiye'deki uygulamaları henüz sınırlı olmakla birlikte, büyük veri teknolojilerine olan ilgi artmaktadır. Ayvaz ve Salman (2020), Türkiye'deki kurumların büyük veri teknolojilerini kullanma olgunluğunu inceledikleri çalışmada, kurumların bu alanda ilerleme kaydettiğini, ancak olgunluk seviyelerinin henüz yeterince yüksek olmadığını belirtmektedir. Bu bağlamda kurumlar, akıllı veri gölü sistemlerini değerlendirip kendi sistemlerin uyarlayarak, veri yönetimi ve analiz süreçlerinde daha etkin ve verimli sonuçlar elde edebilirler (Hai, Geisler ve Quix, 2016).

1.2. ARAŞTIRMANIN AMACI

Kamu sektöründe dijital dönüşümle birlikte giderek artan kamusal verinin toplumsal fayda sağlaması için analiz edilerek sonuçların yorumlanması gerekmektedir. Geleneksel yöntemlerle analiz edilmesi mümkün olmayan büyük verinin analiz edilmesi farklı beceriler ve politikalar gerektirmektedir. Ancak ulusal veri politikaları doğrultusunda geliştirilen ve uygulanan yöntemler sayesinde kamu verisinin anlam kazanması ve işlevsel hale gelmesi mümkün olur. Türkiye’de bu konuda çeşitli çalışmalar bulunmakla birlikte kamusal veri mimarisi üzerine politika düzeyinde bir strateji belirlenmediği görülmüştür. Bu doğrultuda araştırmanın amacı, kamusal veri yönetiminin ve veri mimarisinin özellikleri ve gereklilikleri hakkında bilgi vermek; kamu verisine yönelik olarak dünyada ve

ülkemizde gerçekleştirilen bilimsel çalışmaları analiz etmek; büyük verinin işlenmesi ve son kullanıcı için anlamlı sonuçlar elde edilmesi süreçlerini açıklamak; veri gölü kavramını ve farklı veri gölü mimarilerinin özelliklerini ve avantajlarını değerlendirmek; veri gölü tabanlı mimarilerin geleneksel veri mimarileriyle karşılaştırmalı analizini yapmak ve Veri Politikalarını inceleyerek Türkiye'nin kamusal veri politikalarına ve veri mimarisine yönelik bir model önerisi ortaya koymak olarak belirlenmiştir.

Böylece Türkiye'deki kamu kurumlarının elinde bulunan veriden üst düzeyde faydalanılmasına katkı sağlanması amaçlanmaktadır. Önerilen model ile kamu kurumlarının değerli verilerini saklama, işleme, verilere ulaşma ve verileri analiz etme süreçlerini verimli bir şekilde yönetmesi, kurumsal bilgi sistemlerinin ve veri arşivlerinin daha etkin bir şekilde kullanılması, böylece verinin değerlendirilmesi ve anlamlandırılması süreçlerine katkı sağlanması amaçlanmaktadır. Model, verinin yaşam döngüsünü bütüncül bir şekilde ele alarak, verilerin oluşturulmasından arşivlenmesine kadarki tüm aşamaları kapsayacak şekilde tasarlanmıştır. Veri gölü mimarileri temel alınarak önerilen veri mimarisi modelinin kullanılmasının Türkiye'deki kurumların veri yönetimi kabiliyetlerini artırarak, büyük veri analitiği ve karar alma süreçlerine önemli katkılar sağlayabileceği öngörülmektedir. Bu tezdeki ana odak ve çerçeve, kamu kurumlarına yönelik veri mimarisi çözümleri ve modellerinin geliştirilmesi olmakla birlikte, önerilen mimari ve yöntemlerin diğer kurumlar tarafından da benimsenmesi durumunda benzer faydaların sağlanabileceği değerlendirilmektedir.

Ulusal düzeyde veri yönetimi ve entegrasyonunun önemine işaret eden bu çalışma, aynı zamanda ulusal veri arşivlerinin oluşturulmasına yönelik bir çerçeve sunmaktadır. Bu bağlamda, ulusal veri arşivi oluşturma süreci, verilerin standardizasyonu, güvenliği, teknik altyapıların seçimi ve verilerin işlenmesi gibi birçok bileşeni içermektedir. Böyle bir arşivin oluşturulması verilerin analiz edilmesi, farklı kurumlar arasında güvenli bir şekilde paylaşılması ve uzun vadeli olarak sürdürülebilir bir yapı kurulması açısından kritik öneme sahiptir. Verilerin

ulusal ölçekte etkili bir şekilde kullanılması ve kamu hizmetlerinde iyileştirmeler sağlanması için bu konunun çalışılması önerilmektedir.

Bu kapsamda, tezde kurum düzeyinde geliştirilen modelin, ulusal düzeyde bir veri arşivi modeline temel oluşturabileceği değerlendirilmektedir. Tez kapsamında, bu doğrultuda ilerlemek için bir altyapı sunulmuş ve ulusal düzeyde bir veri arşivi modelinin geliştirilmesini destekleyecek öneriler sunulmuştur. Bu öneriler, farklı kurumların ürettiği arşivsel verilerin sistematik veri entegrasyonu ve veri yönetişim stratejilerinin geliştirilmesi ile ulusal düzeyde bir araya getirilmesini kapsamaktadır.

1.3. ARAŞTIRMANIN PROBLEMİ VE ARAŞTIRMA SORULARI

Araştırmanın problemi “Türkiye’de kamu kurumlarının veri süreçlerini etkin bir şekilde yönetememesi, veri yapılarında ve üst veri alanlarında standardizasyon, veri akışları, erişim ve yetkilendirme ile uzun süreli koruma ve arşivleme sorunlarının yaşanmasına yol açmaktadır. Ayrıca bu durum kamu verisinin analiz edilmesiyle elde edilmesi mümkün olan potansiyel faydaların yeterince sağlanamamasına neden olmaktadır” olarak belirlenmiştir. Çalışmanın temel araştırma sorusu “Türkiye’de kamu kurumlarının sahip olduğu verinin nasıl yönetildiğine ve kullanıldığına yönelik veri mimarisi modeli ve standart bir veri mimarisi politikası bulunmakta mıdır?” olarak belirlenmiştir. Temel araştırma sorusuna bağlı olarak belirlenen çalışmanın alt araştırma soruları aşağıda verilmiştir:

- **S1:** Kamu kurumlarının yürüttükleri iş süreçlerinde oluşan verileri etkin bir şekilde kullanamamasının sebepleri nelerdir?
- **S2:** Mevcut veri mimarileri ve politikaları kamu verisinin analiz edilmesine nasıl katkı sağlamaktadır?
- **S3:** Kamu kurumlarının veri akışlarını optimize etmek için hangi teknikler ve stratejiler önerilebilir?

- **S4:** Üst veri yönetimi ve veri sınıflandırma süreçleri kamu verisinin etkin kullanımını nasıl artırabilir?
- **S5:** Literatürde bulunan veri mimarisi çözümleri incelendiğinde Türkiye için hangi çıkarımlar yapılabilir?
- **S6:** Veri gölü mimarileri, kamu verisinin yönetilmesi ve analiz edilmesi sürecinde hangi yenilikçi çözümler sunabilir?

1.4. ARAŞTIRMA HİPOTEZLERİ

Araştırmanın temel hipotezi "Türkiye'de kamu kurumlarının sahip olduğu verinin nasıl yönetildiğine ve kullanıldığına yönelik kamuya açık veri mimarisi modeli ve standart bir veri mimarisi politikası bulunmamaktadır." şeklinde oluşturulmuştur. Ana hipoteze bağlı olarak belirlenen çalışmanın alt hipotezleri aşağıda verilmiştir:

- **H1:** Kamu kurumlarının mevcut veri süreçlerinde kullanılan teknolojiler, büyük veri çağının gereksinimlerini karşılamamaktadır.
- **H2:** Kamu kurumlarının sistematik bir veri mimarisi politikası dahilinde veri yönetimi süreçlerini uygulamaması, kurumsal verilerin etkili kullanımını engellemektedir.
- **H3:** Veri paylaşımı ve erişilebilirlik mekanizmalarındaki eksiklikler, kamu verisinin potansiyel toplumsal faydalarını sınırlamaktadır.
- **H4:** Veri gölü mimarilerinin uygulanması, kamu kurumlarının veri yönetim süreçlerini iyileştirebilir.
- **H5:** Üst veri yönetimi ve veri yönetişimindeki eksiklikler, veri mimarilerinin etkinliğini sınırlandırmaktadır.
- **H6:** Veri gölü mimarilerinin uygulanması, kamu kurumlarının veri analiz kapasitesini artırabilir.

Bu hipotezler, önerilen modelin hem mevcut sorunlara çözüm sunabileceğini hem de kamu kurumlarının veri yönetim süreçlerinde yenilikçi yaklaşımlar geliştirebileceğini göstermeyi amaçlamaktadır.

1.5. ARAŞTIRMANIN KAPSAMI

Bu çalışma kapsamında Türkiye'deki kamu kurumlarının veri yönetim süreçlerinin mevcut durumu dikkate alınarak, verinin etkili bir şekilde yönetilmesi, analiz edilmesi ve anlamlandırılması süreçlerine ilişkin kapsamlı bir araştırma yapılmıştır.

Ülkemizdeki kamu verisinin büyük bir kısmını bulunduran Cumhurbaşkanlığı, bakanlıklar ve bağlı kamu kurumlarının web sayfaları, veri yönetimi ve arşivlenmesinde kullanılan yöntemler ve veri mimarisi modellerini belirlemek amacıyla incelenmiştir. Ayrıca bu kapsamda kamusal veri konusunda hazırlanmış strateji, politika, yol haritası vb. içeren dokümanlar değerlendirilmiştir.

Dünya genelinde başarılı olarak kabul edilen veri yönetimi sistemleri, açık veri platformları ve veri mimarisi çözümleri ve uygulamaları incelenmiştir. Bu analizler, Türkiye'deki kamu kurumlarına yönelik uluslararası standartlarla ve yenilikçi teknolojilere uyumlu model önerilerinin geliştirilmesine katkı sağlamıştır. Çalışma, veri gölü mimarilerinin kamu kurumlarında uygulanabilirliğini ve bu mimarilerin sağlayabileceği potansiyel faydaları değerlendirmeyi hedeflemiş, aynı zamanda karşılaşılabilecek uygulama zorluklarına yönelik çözüm önerileri geliştirmiştir.

Farklı ülkelerde kullanılan ve literatürde var olan veri platformları incelenmiş ve bu yaklaşımların ve uygulamaların Türkiye'deki kamu kurumları için uygulanabilirlikleri değerlendirilmiştir. Uluslararası düzeyde başarılı veri politikaları ve stratejileri temel alınarak, büyük veri çağının gerekliliklerine uygun, sürdürülebilir ve esnek bir veri mimarisi modeli geliştirilmiştir.

Aşağıda verilen kurum, kuruluş ve kütüphane veri tabanları konuya yönelik literatüre ulaşmak amacıyla taranmıştır:

- Dergipark
- Directory of Open Access Journals (DOAJ)
- EBSCOhost Research Platform (search.ebscohost.com)
- EMERALD

- Google Akademik
- Google Kitaplar
- IEEE Xplore
- Kütüphaneler ve Yayımlar Genel Müdürlüğü
- Proquest
- ScienceDirect
- Scopus
- ULAKBİM (Ulusal Akademik Ağ ve Bilgi Merkezi)
- Üniversite Kütüphaneleri (Orta Doğu Teknik Üniversitesi, Hacettepe Üniversitesi)
- Web of Knowledge
- YÖK Tez Merkezi

1.6. ARAŞTIRMANIN YÖNTEMİ, VERİ TOPLAMA VE DEĞERLENDİRME TEKNİKLERİ

Bu çalışmada yöntem olarak, Türkiye'de kamu kurumlarının sahip olduğu verinin nasıl yönetildiğini ve kullanıldığını belirlemek, ayrıca dünyada ve ülkemizde kamu verisi konusunda gerçekleştirilen bilimsel çalışmaları analiz etmek amaçlarıyla betimleme, içerik analizi ve belgesel tarama (doküman analizi) yöntemleri kullanılmıştır. Betimleme, araştırmaya konu olan olay, grup veya varlığı tanımlayan ve açıklayan bir araştırma yöntemidir. Mevcut olayları, geçmişteki olay ve koşullarla ilişkilerini de inceleyerek aradaki etkileşimiyle açıklamayı amaçlar (Kaptan, 1995'ten aktaran Erbaş, 2018, s. 14). İçerik analizi, nesnel, ölçülebilir ve doğrulanabilir bilgilere ulaşmayı hedefleyen, doküman, metin ve evrak gibi çeşitli materyalleri örnekleme, kodlama, kategorileştirme vb. belirli kurallar çerçevesinde inceleyen ve nitel araştırma yöntemleri kapsamında kullanılan bir tekniktir (Metin ve Ünal, 2022, s. 1). Belgesel tarama yöntemi ise araştırma verilerinin temel kaynağı olarak çeşitli belgelerin toplanmasını, detaylı bir şekilde incelenmesini, sorgulanmasını ve analiz edilmesini içeren bilimsel bir araştırma yöntemidir (Sak,

Şahin Sak, Şendil ve Nas, 2021, s. 1). Çalışmada ayrıca kurumsal bilgi sistemleri sistem analizi yöntemiyle değerlendirilmiştir. Sistem analizi, bir sistemin mevcut durumunu anlamak ve gereksinimlerini tanımlamak amacıyla yürütülen kapsamlı bir inceleme sürecidir; mevcut sistemin iyileştirilmesi veya yeni bir sistemin tasarlanması ve geliştirilmesi için gerekli bilgi ve verileri sağlamayı hedefler (Dennis, Wixom ve Tegarden, 2015).

Çalışma kapsamında veri, literatür taraması tekniğiyle elde edilmiştir. Öncelikle veri kavramı detaylarıyla değerlendirilmiştir. Büyük verinin işlenmesi ve son kullanıcı için anlamlı sonuçlar elde edilmesi süreçleri ve Ulusal Veri Politikaları incelenmiştir. Kamu verisine yönelik gerçekleştirilen bilimsel çalışmalar analiz edilmiştir. Dünyadaki ve ülkemizdeki farklı kamu kurumlarının web sayfaları taranarak kurumlarda kullanılan veri mimarilerine ilişkin bilgi paylaşımı yapılıp yapılmadığı incelenmiş, kamuya açık veri paylaşımında bulunan kurumların bu paylaşımları hangi platformlar aracılığıyla gerçekleştirdiği, bu amaçlarla kamu kurumlarının standart bir yöntem kullanıp kullanmadığı belirlenmiştir. Ülkemizdeki kamu verisinin önemli bir kısmını elinde bulunduran Cumhurbaşkanlığı, Bakanlıklar ve bağlı kamu kurumlarının web sayfaları kamuya açık verinin paylaşılmasında kullanılan yöntem ve kamusal veri konusunda hazırlanmış strateji, politika, yol haritası vb. içeren dokümanların belirlenmesi amacıyla taranmıştır. Yine dünya genelinde veri paylaşımı konusunda önde gelen kamu kurumlarının web sayfaları aynı çerçevede değerlendirilmiştir.

Elde edilen sonuçlar doğrultusunda Türkiye'de kamu kurumları için veri mimarisinin nasıl olması gerektiği konusunda bir model önerisi sunulmuştur.

Kısaca aşağıdaki adımlar izlenmiştir:

- Veri Toplama: Literatür taraması yöntemiyle ulusal ve uluslararası kaynaklardan veri toplanmıştır.
- Analiz: Veri gölü mimarileri ve kamu veri yönetim sistemleri, dünyadaki başarılı uygulamalar ışığında karşılaştırmalı olarak incelenmiştir.

- Model Önerisi: Veri gölü mimarisi temel alınarak Türkiye'deki kamu kurumları için uygulanabilir bir model geliştirilmiştir.

Çalışmada ele alınan araştırma soruları, ilgili hipotezlerle ilişkilendirilmiş ve her bir sorunun yanıtlanmasında kullanılan yöntemler Tablo 1'de sunulmuştur.

Tablo 1. Araştırma soruları, ilgili hipotezler ve kullanılan analiz yöntemleri

Araştırma Sorusu	**Hipotez**	**Analiz Yöntemi**
Ana Araştırma Sorusu	Ana Hipotez	Literatür değerlendirmesi Uygulama örneklerinin incelenmesi Belgesel tarama Sistem analizi
S1	H1	Literatür değerlendirmesi Uygulama örneklerinin incelenmesi Sistem analizi
S2	H2	Literatür değerlendirmesi Belgesel tarama Sistem analizi
S3	H3	Literatür değerlendirmesi Belgesel tarama
S4	H4	Literatür değerlendirmesi
S5	H5	Literatür değerlendirmesi
S6	H6	Literatür değerlendirmesi Uygulama örneklerinin incelenmesi

2. BÖLÜM
VERİ BİLİMİ VE VERİ YÖNETİMİ UYGULAMALARI

“Bilgi çağı” kavramı yaygın olarak kullanılmakla birlikte “veri çağı” kavramının günümüz koşullarını daha doğru tanımladığı ifade edilmektedir. Her gün farklı disiplinlerden petabaytlarca veri bilgisayar ağlarına akmaktadır. ‘Bilgisayarlaşma’, veri toplama ve depolama araçlarının yaygın kullanımı veri hacminde ve çeşitliliğinde çok hızlı artışların meydana gelmesine yol açmaktadır. Kurumsal işlemler, bilimsel uygulamalar, deneyler, mühendislik uygulamaları, kişisel deneyimler vb. süreçlerde sürekli olarak veri üretilmekte ve üretilen veriler küresel telekomünikasyon ağlarıyla sürekli olarak bir yerlere taşınmaktadır. Tıbbi kayıtlar, arama motorları aracılığıyla yapılan Web aramaları, sosyal medya, bloglar büyük miktarlarda veri üreten kaynakların bir kısmıdır. Bu veriler bir taraftan iş ve süreçleri anlatırken analitik olarak da değerlendirilerek ikincil amaçlarla kullanılabilmektedir. Örneğin veri madenciliği çalışmaları ile farklı disiplinlerden elde edilen veriler değerlendirilerek sistematik karar süreçleri desteklenebilmektedir (Han, Kamber ve Pei, 2012, s. 1, 2). Veriye dayalı karar verme ve bilgiye erişimin önemi giderek artmakta, bu süreçlerin etkin yönetimi işletmelerden bilim dünyasına kadar geniş bir yelpazede büyük gelişmeler yaşanmasını sağlamaktadır.

Veriye dayalı karar verme ve bilgiye erişimin önemi giderek artmakta, bu süreçlerin etkin yönetimi işletmelerden bilim dünyasına kadar geniş bir yelpazede büyük gelişmeler yaşanmasını sağlamaktadır (Hossain, Yasmin, Biswas ve Asha, 2024). Örneğin 179 firma üzerinde anket yöntemiyle yapılan bir çalışmada “veri odaklı karar vermeyi” benimseyen firmaların, diğer yatırımları ve bilgi teknolojilerini kullanımları da göz önüne alındığında, üretkenliklerinin beklenenden %5-6 daha yüksek olabildiği gösterilmiştir (Brynjolfsson, Hitt ve Kim, 2011). Veri kullanılabilirliğindeki %10 artışın kurum verimliliğini %17-%49 oranında arttırdığı, büyük verinin veri bilimi süreçleriyle bütünleşmesinin ise

yatırım karlılığını %241 oranında arttırabildiği ortaya konulmuştur (Brynjolfsson, Hitt ve Kim, 2011'den aktaran Gökalp, Kayabay, Çoban, Yandık ve Eren, 2018b, s. 95). Veri bilimi ile avantaj sağlamayı hedefleyen şirketlerin tüm birimlerinde veriye dayalı karar verme mekanizmalarını benimsemeleri ve uygulamaları önemlidir. Aksi takdirde rekabet avantajı elde etmeleri zordur (Gökalp, Kayabay, Çoban, Yandık ve Eren, 2018b, s. 95; Borges, Bernardino ve Pedrosa, 2021). Bu bulgular, veri bilimi ve analitiğin işletmeler ve kurumlar için kritik bir rekabet avantajı sağladığını göstermektedir. Veriye dayalı karar verme stratejik planlama, operasyonel etkinlikler, müşteri ilişkileri yönetimi gibi farklı aşamalarda kullanılabilir (Michael, Ipede, Adejumo, Adenekan, Adebayo, Ojo ve Ayodele, 2024). Günümüzde veri odaklı karar verme, sadece işletmelerin ve kurumların etkinliğini ve verimliliğini artırmakla kalmayıp, aynı zamanda uzun vadeli başarılarını da destekleyen kritik bir stratejidir. Veri bilimi ve analitiğin doğru uygulanması, şirketlerin ve kurumların değişen koşullara uyum sağlamalarını ve sürekli olarak değer yaratmalarını sağlar (Adeyeye ve Akanbi, 2024). Tüm bu süreçler bir Veri Yönetim Planı çerçevesinde yürütülmelidir. TÜBİTAK Açık Bilim Politikasında Veri Yönetim Planı, "Yeni bir araştırmaya başlamadan önce araştırmacı tarafından çalışmada kullanılacak verilerin çalışma süresince ve çalışma tamamlandıktan sonra nasıl yönetileceğinin, diğer bir ifadeyle veri yaşam döngüsündeki her bir adımın nasıl gerçekleştirileceğinin anlatıldığı doküman" olarak tanımlanmaktadır (*TÜBİTAK Açık Bilim Politikası*, t.y., s. 3).

Veri yönetimi araştırmayla birlikte başlar ve verilerin nasıl elde edileceği, toplanacağı, saklanacağı ve arşivleneceği ile ilgili zamansal planlamaları içerir ("OpenAIRE", t.y.). Bu süreçler, bilimsel araştırmalarda ve genel iş süreçlerinde verilerin etkin ve güvenilir bir şekilde yönetilmesini sağlar. Bu süreçlerin etkili bir şekilde yönetilmesi, bilimsel araştırmalarda veri güvenliğini sağlamak, veriye dayalı karar alma süreçlerini desteklemek ve kurumsal veri yönetimi standartlarını karşılamak için kritik öneme sahiptir. Veri yönetimi uygulamaları, veri güvenliği, veri erişimi ve veri paylaşımı gibi konularda da yönetmeliklere uyumu sağlamak için gereklidir.

Veri bilimi alanı yapılandırılmış veya yapılandırılmamış verinin işlenmesinde matematik, istatistik ve bilgisayar bilimleri gibi alanlara ilişkin yöntemleri kullanan bir disiplindir (Altun, Şahin ve Öztaş, 2017, s. 2025). Bu yöntemlerin kullanılmasıyla verilerin keşfedilmesi ve karar destek sistemlerinin geliştirilmesi sağlanır.

2.1. VERİ VE BÜYÜK VERİ

Veri, bilgi kaynağı olarak toplanan ölçüm, gözlem veya kayıt işlemleri sonucunda elde edilen ölçütleri ifade eder. Çeşitli veri türleri ve veriyi temsil etmenin farklı yolları mevcuttur; örneğin, belirli bir ülkedeki insan sayısı, kişilerin doğdukları ülkeler, belirli bir ürünün satış değeri veya hava sıcaklık ortalamaları gibi örnekler bu kapsama girer (“Australian Bureau of Statistics”, t.y.). Günümüz bilgi ve iletişim teknolojilerinde, verinin işlenmesi, saklanması ve analiz edilmesi süreçleri, veriye yapılandırılmış biçimde ulaşılmasını sağlayarak bilgi üretiminde ve karar alma mekanizmalarında kritik roller üstlenmektedir

Bilgi ve iletişim teknolojilerinde, özellikle internet teknolojilerinde yaşanan gelişmelerle sürekli olarak veri toplanmaktadır ve geçmişe oranla hız ve hacim açısından elde edilen veri çok daha fazladır. Sosyal medya ve çevrimiçi hizmetler sayesinde internet çok büyük bir veri elde etme platformu haline gelmiştir ve elde edilen veri her geçen gün hızla artmaktadır. Öyle ki veri artık gigabayt yerine eksabayt, zettabayt ve yottabayt olarak ölçülmektedir. Bu veriler çok farklı alanlarda kullanılarak değer elde edilebilecek potansiyele sahiptir (Ardagna, Barbierato, Gianniti, Gribaudo, Pinto, da Silva ve Almeida, 2020). Yaşanan bu gelişmeler “büyük veri” kavramını doğurmuştur (Doğan ve Arslantekin, 2016, s. 15,16).

Büyük veri kavramı sadece verinin hacminin büyüklüğünü ifade etmemektedir. Veri yapısı ve hızı da büyük veri kavramının önemli boyutlarıdır. Verinin yapısı depolanan verinin yapılandırılmış olup olmamasıyla ilgilidir. Yapılandırılmış veriyi

analiz etmek daha kolaydır. Bu nedenle verinin etiketlenmiş olması, sınıflandırma ve kümelendirme işlemlerinin yapılması önemlidir. Hız ise, verinin üretilme hızı ve güncelliğini ne kadar koruduğuyla ilgilidir (Kayabay, Gökalp, Akyol, Eren ve Koçyiğit, 2016, s. 4; Yaseen ve Obaid, 2020). Veri, doğru ve uygun şekilde işlendiğinde, stratejik kararlar almak ve çeşitli alanlarda iyileştirmeler yapmak için değerli bilgiler sağlar.

Büyük verinin değerinin ortaya konabilmesi için çeşitli aşamalardan geçirilmesi gerekmektedir. Büyük verinin işlenmesine yatırım yapan kurumlar iş ve uygulamalarında büyük avantaja sahip olmaktadır (Doğan ve Arslantekin, 2016; Prabhugouda ve Asra, 2024). Çeşitli kurumlar, kendi kurumsal verilerine ek olarak dış kaynaklardan veri alarak veriyi analiz edip bilgi elde etmeye ihtiyaç duymaktadır. Ancak kurum dışından elde edilen veri kurumun veri tabanlarına aktarılabilir nitelikte olmayabilir veya yapılandırılmamış olabilir. Büyük teknoloji şirketleri büyük veri ile ilgili büyük yatırımlar yaparak verinin anlamlandırılması için çalışmaktadır (Doğan ve Arslantekin, 2016, s. 21). Veri madenciliği, makine öğrenmesi ve yapay zekâ gibi ileri analiz teknikleri, büyük veriden anlamlı bilgiler çıkarma sürecinde birbirini tamamlayıcı rollere sahiptir. Bu yöntemler sayesinde işletmeler, kendi ihtiyaçları doğrultusunda stratejik kararlarını daha sağlam temellere dayandırabilirler.

Scopus'ta yer alan veriye göre "Büyük Veri" terimi ilk kez 1970 tarihli atmosferik sondajlar ve okyanus sondajları üzerine bir makalede kullanılmıştır. İlgili makalede, 1969 yılında yedi Amerika Birleşik Devletleri departmanı ve ajansının ortak projesi olan, Barbados Oşinografik ve Meteorolojik Deneyi'nin sonuçları tartışılmıştır (Halevi ve Moed, 2012, s. 3). 2008 yılından itibaren ise Büyük Veri üzerine yapılan yayın sayısında çok hızlı bir artış görülmüştür. Bilgisayar dünyasında ise "Büyük Veri" terimi ilk kez 2005 yılında Roger Magoulas (O'Reilly Medya) tarafından, "karmaşıklığı ve boyutu nedeniyle geleneksel veri yönetimi teknikleriyle yönetilmesi ve işlenmesi mümkün olmayan büyük miktarda veriyi tanımlamak için kullanılmıştır." (Ularu, Puican, Apostu ve Velicanu, 2012, s. 3). Terimin daha çok bilgisayar bilimi ile ilgili olduğu düşünülse de sağlık, beşerî

bilimler, mühendislik vb. birçok disiplinde Büyük Veri kavramı kullanılmaktadır (Halevi ve Moed, 2012, s. 5).

Amerika Birleşik Devletleri'nde (ABD) 2011 yılında Büyük Veri Yönlendirme Grubu'nun kurulmasıyla bu alandaki çalışmaların önemi vurgulanmıştır. Büyük verinin değerine yönelik olarak 2014 yılında Amerika Birleşik Devletleri (ABD) Başkanı'na sunulan "Big Data: Seizing Opportunities, Preserving Values (Büyük Veri: Fırsatların Yakalanması, Değerlerin Korunması)" adlı rapor dikkat çeken çalışmalardan biri olmuştur. Rapor genel hatlarıyla kamu ve özel sektörün büyük veriden maksimum fayda sağlarken riskleri nasıl azaltabileceğine yöneliktir. Yine raporda büyük veri ile ABD ekonomisinde, sağlık ve eğitim hizmetlerinde, toplumun güvenliği ve enerji verimliliği konularında sağlayabileceği fırsatlara değinilmiştir (Aktaran Altun, Şahin ve Öztaş, 2017, s. 2029). 2016 yılında ise "The Federal Big Data Research and Development Strategic Plan (Büyük Veri Araştırma ve Geliştirme Federal Stratejik Planı)" isimli planın hazırlanmasıyla birlikte büyük veri araştırmalarına yönelik yedi strateji belirlenmiştir (*The Federal Big Data*, 2016). Bu stratejiler:

1. Büyük veri teknolojilerinden faydalanarak yeni nesil yetenekler yaratma
2. Veri ve ulaşılan bilginin güvenilirliğini anlamak ve daha iyi kararlar alarak keşifleri mümkün kılmak için Ar-Ge'yi destekleme
3. Büyük veri inovasyonuna olanak sağlayan araştırma siber altyapısı oluşturma
4. Veri paylaşımı ve yönetimini teşvik eden politikalarla verinin değerini artırma
5. Büyük veri toplama, paylaşma ve kullanmayla ilgili gizlilik, güvenlik ve etik durumları anlama
6. Büyük veriye yönelik işgücü talebini karşılamak amacıyla ulusal eğitim ve öğretimi geliştirme

7. Ulusal büyük veri ekosisteminde bağlantılar geliştirme (*The Federal Big Data*, 2016).

ABD'nin büyük veri stratejileri, büyük verinin sunduğu fırsatları değerlendirirken karşılaşılan riskleri yönetmeyi ve bu alandaki yeniliklerin keşfini teşvik etmeyi amaçlarken aynı zamanda bu alandaki gelişmelerin sürdürülebilirliğini sağlamaya da yöneliktir.

Büyük veri uygulamaları daha çok müşterilere daha iyi ürünleri, daha doğru şekilde ve doğru zamanda sunma amacıyla pazarlama alanında kullanılsa da (Şeker, 2015, s. 15), bütün kurumların ellerindeki verinin değerini ortaya çıkaracak süreçleri yönetmesi, kurumlara da değer katacaktır. Büyük verinin temel önemi farklı türdeki büyük miktarda verinin kullanımında verimliliği artırma potansiyeli olarak ifade edilmektedir. Kurumların veriyi doğru şekilde tanımlaması ve kullanması, işlerini daha iyi yaparak farklı alanlarda verimliliklerini artırmalarına katkı sağlayabilir. Büyük veri bilgi teknolojisinde mevcut kayıtlardaki paternleri analiz ederek güvenliği iyileştirmek ve sorunları gidermek, çağrı merkezleri aracılığıyla elde edilen bilgileri kullanarak hizmetleri özelleştirip müşteri memnuniyetini artırmak, yine sosyal medya aracılığıyla potansiyel müşterilerin tercihlerini belirleyerek hizmet ve ürünlerin iyileştirilmesi, çevrim içi dolandırıcılık tespiti, finansal işlemlerden elde edilen bilgilerin kullanılarak risk değerlendirmesi yapılması vb. amaçlarla kullanılmaktadır (Ularu ve diğerleri, 2012, s. 4; Jia, 2023). Bu uygulamalar, büyük verinin kurum ve kuruluşlar için ne kadar değerli olduğunu göstermektedir. Verinin doğru kullanımı yalnızca operasyonel verimliliği artırmakla kalmaz, aynı zamanda stratejik kararlar alınmasına da katkı sağlar.

Büyük veri çalışmalarına örnek olarak Güney Afrika ve Avustralya'da 2024 yılında tamamlanması planlanan Kilometrekare Dizisi (Square Kilometre Array [SKA]) projesi verilebilir. SKA projesi tamamlandığında günde 1 eksabayttan fazla ham veri üretecektir. SKA kapsamında 3000'den fazla alıcı çanak ile bir kilometrekarelik birleşik bir bilgi toplama alanı üretilmesi hedeflenmektedir. Proje kapsamında Dünya'nın çevresini iki kez sarmak için yeterli optik fiber kullanacağı

ve proje bütçesinin 1,5 milyar Euro olduğu belirtilmektedir. Büyük veri çalışmalarına bir diğer örnek olarak ise Büyük Hadron Çarpıştırıcısı (Large Hadron Collider) verilebilir. Bu proje kapsamında ise 2012 yılında 150 milyon sensör ile 22 petabayt veri üretildiği belirtilmiştir. İnsan Genomu Projesi (Human Genome Project) ve dünyanın gözlemi için kullanılan uydular büyük veri çalışmalarına örnek olarak gösterilebilecek diğer çalışmalardır (Harris, 2012, s. 11).

Bu projeler, büyük veri kullanımının farklı alanlardaki potansiyelini göstermektedir. Büyük veri teknolojileri, bilimsel araştırmaların yanı sıra günlük yaşamda birçok farklı alanda verimliliği artırma ve problemlere yenilikçi çözümler sunma potansiyeline sahiptir.

2.1.1. Yapısına göre Veri

2.1.1.1. Yapılandırılmış Veri

Yapılandırılmış veri, büyük ölçüde sayısal olan, genellikle düzenli bir tablo formatında depolanan ve işletim sistemlerinden, veri tabanlarından ve teknoloji araçlarından vb. gelen iyi tanımlanmış bir yapıya sahip olan veriyi ifade eder (Losee, 2006). Çoğu işletme, genellikle satır-sütun formatına ve ay/gün/yıl gibi açık üst veri öğelerine sahip olan yapılandırılmış veriyi kullanır. İşletmeler genellikle büyük miktardaki yapılandırılmış veriyi yönetme ve analiz etme konusunda başarılıdır, çünkü veri çok iyi tanımlanmış bir şekilde bulunmaktadır. Yapılandırılmış veri makineler tarafından kolayca işlenebilir (“Hitachi Vantara”, t.y.). Yapılandırılmış verilerin tanımlanmış formatı, otomasyon ve verimlilik açısından önemli avantajlar sunar.

2.1.1.2. Yapılandırılmamış Veri

Yapılandırılmamış veri kurallara tabi olmayan, resim, taranmış doküman, ses/video kayıtları ve ofis dosyaları gibi heterojen verilerdir (Mishra ve Misra, 2017). Bu veriler farklı kaynaklardan gelen ve farklı formatlara sahip olan, değişen karmaşıklık düzeylerine sahip verilerdir ve işlenmesi zordur. Yapılandırılmamış veri belirli kurallara sahip olmadığı için işletmeler tarafından yapılandırılmamış verileri anlamak ve bunlardan değer çıkarmak zordur. Bir müşterinin bıraktığı sesli mesaj, bir çalışanın el yazısıyla aldığı notlar, bir sosyal medya paylaşımı vb. veriler yapılandırılmamış verilere örnektir. Bu verilerin işlenmesi için akıllı bir veri yönetim stratejisi gerekir (“Hitachi Vantara”, t.y.). Yapılandırılmamış veriler, işletmeler için büyük bir potansiyel taşır ancak bu verilerin işlenmesi ve analiz edilmesi, yapılandırılmış verilere kıyasla daha karmaşıktır. Verilerin çeşitli formatlarda olması ve farklı kaynaklardan gelmesi, ileri analiz teknikleri ve veri yönetim stratejileri gerektirir.

2.1.1.3. Yarı Yapılandırılmış Veri

Yarı yapılandırılmış veri, yapılandırılmış ve yapılandırılmamış veri türlerinin bir karışımıdır. Genellikle sayısal veri içerir ancak bu sayısal verilerin çıkarılması zordur ve akıllı bir veri yönetim stratejisi gerektirir. Bu verilerin kendi içlerinde bir yapısı olabilir ancak bu yapı standart veri yönetim süreçleri için gerekli yapıya uygun değildir. Yarı yapılandırılmış veri, XML veya JSON gibi kendi kendini tanımlayan bir şema kullanan her türlü veridir (Truică, Apostol, Darmont, ve Pedersen, 2021). Bu tür veriler açık uçlu şemaya sahiptir. Açık uçlu şema, yarı yapılandırılmış verilerin gömülü yapıyı tanımlamak için onu oluşturan uygulamaya bağlı olmadığı anlamına gelir. Tanımlar ve kısıtlamalar veri kümelerini oluşturan uygulamadan bağımsız olarak dosyanın içine gömülür. Bu tür veriler, yapılandırılmış bir veri deposundaki belirli kayıt türleri için ek özellikleri kaydetmek için yapılandırılmış verilerle birleştirilebilir (“Hitachi Vantara”, t.y.).

Kurumlar için önemli veri kaynağı olan Elektronik Belge Yönetim Sistemleri (EBYS) yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış veri üretimi açısından son derece zengin bir kaynaktır. Örneğin, yapılandırılmış veri kategorisinde belgelerin üst verileri (ör. belge numarası, tarih, gönderen ve alan kişiler), işlem kayıtları (ör. belge düzenleme ve onay süreçleri) gibi veriler üretilir. Yarı yapılandırılmış veri kategorisinde XML veya JSON formatında belge süreçlerine dair ayrıntılar, belge ile ilişkili diğer yazışma ve dosya bilgileri yer alır. Yapılandırılmamış veri kategorisinde ise EBYS sistemlerinden elde edilen PDF, taranmış doküman, video veya ses kayıtları gibi içerikler bulunur.

2.1.2. Türüne Göre Veri

2.1.2.1. Analog Veri

Analog veriler herhangi bir makine veya otomatik cihaz tarafından, internete bağlı olmadan dahi oluşturulan veridir. Bir nükleer reaktörden, bir cihazın CPU kullanımına kadar farklı türdeki verileri içerir. Bir üretim olayı gibi bir tetikleyici ile ilişkilidir; “bir parça oluşturuldu”, “bir kutu taşındı” gibi olaylar analog kaydın oluşturulmasına sebep olurlar. Analog ölçüm herhangi bir kullanıcı girişi veya ekstra işlem olmadan çoğunlukla mekanik olarak yapılır. Genel olarak uzun bir sayı listesinden oluşan analog veriler büyük hacimli ve çok tekrarlıdır (Inmon ve Linstedt, 2014). Analog cihazların oluşturduğu veriler genel olarak ısı, ağırlık, boyut, kimyasal bileşim vb. Fiziksel değer ölçümleridir ve bu ölçümler genel olarak küçük farklılıklar içermektedir ancak önemli olan genellikle bu farklılıklar olmaktadır. Farklı ölçümlerin nedenini anlamak gerekir, bu nedenle analog verilerle ilgili üst işlem bilgileri verinin kendisinden çok daha önemlidir. Üst işlem ölçüm zamanı, konumu, hızı, veri yükleyenin kim olduğu vb. bilgileri içerir (Inmon, 2016, s. 35).

Analog ölçüm sürecinde yakalanan ham verilere eşlik eden yardımcı verilere üst işlem verisi denir. Üst işlem verisi, verinin kaydedildiği tarih, yer, kim tarafından

ve ne ile kaydedildiği, vb. destekleyici verileri içerir. Çoğu zaman analog ölçümler log adı verilen veri “defterlerinde” saklanır. Veri logu, bir analog ölçüm oluşturan olaylar sırasında tespit edilen bir veya daha fazla değişkenin sıralı ölçümüdür. Bir veri logu çok detaylıdır ve analog veri noktaları çok küçük aralıklarla oluşturulur. Veri logunun yapısı genel olarak oldukça karmaşıktır. Bu karmaşıklık, veri yönetim süreçlerinde bir dizi zorluk yaratır ve genellikle üst işlem bilgilerini doğru bir şekilde yorumlamak için özel veri mimarisi çözümlerine ihtiyaç duyulur (Li, Kennedy, Ngoran, Wu ve Hunter, 2013). Bu nedenle, çoğu zaman veri logunu okumak ve yorumlamak için yardımcı yazılımlar ve programlar kullanılır. Çoğunlukla, veri logu yalnızca ilgilenilen olayları değil, meydana gelen tüm olayları yakalar (Inmon, 2016, s. 36-38).

2.1.2.2. Uygulama Verisi

Uygulama verileri bir işlem veya uygulamanın yürütülmesi ile oluşur. Örneğin satış verileri, sevkiyat verileri, ödeme verileri, fatura verileri, bankacılık verileri, üretim süreci kontrol verileri gibi veriler uygulama verileridir. Bir işlem meydana geldiğinde, bir uygulama tarafından ölçülür ve veri oluşur. Tipik olarak veriler ortak ve tekrar eden tek tip bir yapıya sahiptir. Uygulama verilerinin farklı formları olmakla birlikte en yaygın olanı uygulamada kaydetme işlemi ile oluşturulan verilerdir. Bu kayıt verileri uygulamanın ait olduğu veri tabanı yönetim sistemi ile önceden şekillendirilmiş olarak veri göletine gönderilebilir veya şekillendirilmemiş olabilir. Genel olarak uygulama verileri ortak ve tekrar eden bir yapıya sahiptir (Inmon, 2016, s. 39). Uygulama verisi kayıtlarının yapısı analog verilerde kullanılan yardımcı verilere göre çok daha kapsamlıdır ve çeşitli özellikler taşıyan veri modüllerine sahip olabilmektedir. Örnek vermek gerekirse, bu modüllerden bir (veya daha fazlası) kayıt verisine ait anahtar olarak atanabilir ve benzer bir şekilde kayıt verisine ait dizin (verinin konumu) bilgisini içerebilir. Bu özniteliklerden biri veya daha fazlası bir anahtar olarak atanabilir. Niteliklerden biri veya daha fazlası bağımsız bir dizine sahip olabilir (Inmon, 2016, s. 40).

Uygulama verilerinin etkin yönetimi, kurumların operasyonel verimliliğini artırmak ve veri odaklı kararlar almak için kritik öneme sahiptir (Davenport ve Bean, 2018).

2.1.2.3 Metin Verisi

Metinsel veriler de genel olarak bir uygulama ile ilişkilendirilir ancak biçimsel olarak çok farklıdır. Uygulama verileri daha önce açıklandığı gibi tek tip kayıtlardan oluşurken, metin verileri tek tipte olmaz. Bu nedenle metinsel verilere "yapılandırılmamış veri" denir (Inmon, 2016, s. 40). Yapılandırılmamış verileri otomatik araçlarla kolayca işlemek mümkün değildir. Sözleşmeler, e-posta yazışmaları, mahkeme kararları, sosyal medya yazışmaları vb. veriler metin verileridir. Metin verilerine ancak yüzeysel analiz yapılabilir, metin verilerinin analitik olarak kullanılabilmesi için metinsel belirsizlik giderme (textual disambiguation) ile dönüştürülmesi gerekir (textual disambiguation). Bu süreç, metinsel belirsizlikleri çözerek verileri anlamlı ve analiz edilebilir hale getirir ve doğal dil işleme ve metin madenciliği gibi teknikler kullanılarak gerçekleştirilir (Gharehchopogh ve Khalifelu, 2011). Metin kapsamlı olarak analiz edilecekse öncelikle yapısal olmayan veri biçiminin metinsel belirsizlik giderme yoluyla dönüştürülmesi ve bilgisayar tarafından analiz edilebilecek biçime dönüştürülmesi gerekir (Inmon, 2015). Metinsel belirsizlik giderme ile ayrıca metnin bağlamı belirlenmiş ve metin ile ilişkilendirilmiş olur. Analog veriler ve uygulama verilerini ise verilerin yakalanmasındaki tekdüzelik nedeniyle nadiren dönüştürmek gerekir (Inmon, 2016, s. 41).

Analog veri, uygulama ve metin verisi ana gruplandırmasına ek olarak veri gölündeki verileri tekrarlayan ve tekrarlamayan veriler olarak da ayırmak mümkündür. Genel olarak analog veri ve uygulama verileri tekrarlayan veri, metin verisi ise tekrarlamayan veridir (Inmon, 2016, s. 42). Metin verilerinin yapılandırılmamış ve tekrarlamayan doğası ve karmaşıklığı, onu diğer iki ana veri türünden ayıran en önemli özellikleridir (Gharehchopogh ve Khalifelu, 2011).

Tekrarlamayan verinin ticari değeri yüksek iken, tekrarlayan verinin ticari değeri düşüktür (Inmon, 2016, s. 43).

Kurumlarda EBYS özellikle uygulama verisi ve metin verisi kategorilerinde önemli bir veri kaynağıdır. EBYS belge gönderim ve onay süreçlerinden kaynaklanan uygulama verilerini yönetirken, kurumsal yazışmalar, sözleşmeler gibi metin verilerini depolamak ve düzenlemek için kritik bir rol oynar.

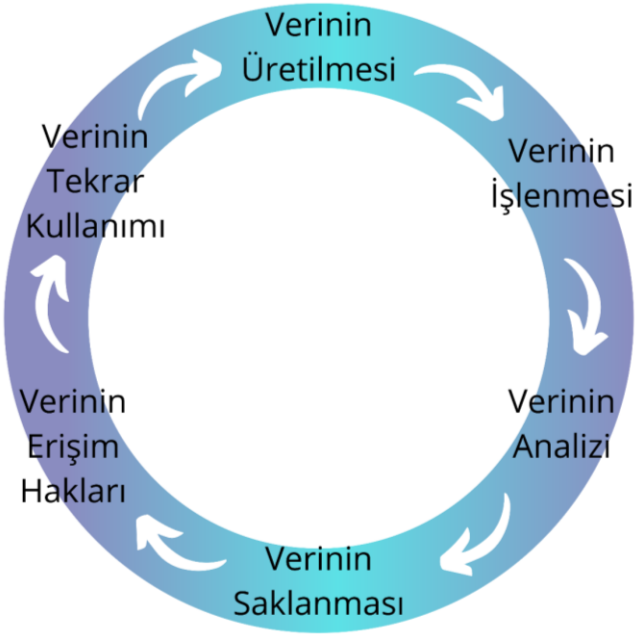
2.1.3. Büyük Verinin Özellikleri

Büyük verinin özellikleri 5V olarak adlandırılan veri tanımında yer almaktadır. Bunlar volume (hacim, büyüklük) yani verinin kapladığı yer, velocity (hız) yani verinin yakalanabilmesi için gereken işlem hızı ve verinin değişim hızı, variety (çeşitlilik) yani verinin kaynağının ve özelliklerinin çeşitliliği, veracity (güvenilirlik) yani veri kaynağının ne kadar güvenilir olduğu ve value (değer) yani verinin büyüklüğünün, değişim hızının, çeşitliliğinin ve güvenilirliğinin ifade ettiği değerdir (Younas, 2019; Ghasemaghaei, 2019). Büyük verinin hacmi büyüktür, değişim hızı yüksektir, çeşitliliği fazladır ve veri kaynakları aynı ölçüde güvenilir değildir. Büyük veri çalışmalarının ilk 4V ile ilgili problemleri çözerek verinin daha değerli olmasına odaklandığı söylenebilir. Büyük veriye yönelik çalışmalarda çok yüksek hacimdeki, çok hızlı değişen, çok farklı kaynaklardan gelen ve bu nedenle çok farklı güvenilirlik derecelerine sahip olan verileri yönetmek ön plandadır. Bu süreç ne kadar iyi yönetilebilirse verinin değeri de o kadar çok artacaktır (Şeker, 2015, s. 10,11).

2.1.4. Verinin Yaşam Döngüsü

Büyük veri için bir diğer önemli kavram ise “Verinin Yaşam Döngüsü”dür. Verinin yaşam döngüsüne yönelik olarak farklı kurumların farklı yaklaşımları bulunmaktadır. Essex Üniversitesi’nin literatüre kazandırdığı Data Life Cycle

(Verinin Yaşam Döngüsü) kavramına göre veri üretilir, işlenir, analiz edilir, güvenilirliği ve bozulmaması sağlanarak saklanır, veriye erişilir ve veri tekrar kullanılır. Bu süreci tekrar verinin üretilmesi izleyerek verinin yaşam döngüsü yeniden başlamış olur (Şeker, 2015, s. 12). Verinin yaşam döngüsü kavramı, büyük veri yönetiminin temel taşlarından biridir. Şekil 2'de verinin yaşam döngüsü şeması verilmiştir (Rahul ve Banyal, 2020; Şeker, 2015, s. 11).



Şekil 2. Verinin yaşam döngüsü (Şeker, 2015, s. 11)

Büyük Veri Yaşam Döngüsü (Big Data Life Cycle) ise Veri Yaşam Döngüsü kavramına göre yenidir ve farklı anlamlarda kullanılabilmektedir. EMC firmasının geliştirdiği Büyük Veri Yaşam Döngüsü, gerekler ve ihtiyaçların belirlendiği iş dünyası, verinin alınıp entegre edildiği veri ambarı, analitik modeli geliştiren ve iyileştiren veri bilim insanı, yeni bakış açıları geliştiren iş zekâsı ve bakış açılarını kullanan ve verimi ölçen iş dünyası süreçlerini içeren bir yaşam döngüsünden bahsetmektedir (Rahul ve Banyal, 2020; Şeker, 2015, s. 13).

Büyük verinin yaşam döngüsüne yönelik bir başka yaklaşım ise SAS firmasına ait "Analiz Yaşam Döngüsü (The Analytics Life Cycle)" dür. Analiz Yaşam Döngüsü, problemin tanımlanması, verinin hazırlanması, verinin keşfi, dönüşüm ve seçim,

modelin inşası, modelin doğrulanması, modelin çalıştırılması, sonuçların değerlendirilmesi ve gözlenmesi süreçlerinden oluşmaktadır (Koo, Kang ve Kim, 2020). Sonuçların değerlendirilmesi sürecini tekrar problemin tanımlanması süreci takip eder. Analiz yaşam döngüsü süreçlerinde verinin değerlendirilmesi ve görselleştirilmesi aşamalarında veri bilim insanı, analizlerin değerlendirilmesi ve tahmin modellerinin oluşturulması aşamasında veri madencisi/istatistikçi, model doğrulama, çalıştırma ve gözlemi aşamalarında Bilgi Teknolojileri sistem yöneticisi, karar verme aşamasında ise yönetim/iş dünyası gibi farklı uzmanlık alanlarının etkisi olmaktadır (Şeker, 2015, s. 14). SAS firmasının Analiz Yaşam Döngüsü, veri analizi sürecinin aşamalarını detaylı bir şekilde ele almaktadır.

2.1.5. Ölçeklenebilir Büyük Veri Mimarisi

Ölçeklenebilir mimarinin uygulanabilmesi, verinin özelliklerine ve ne şekilde kullanılacağına yönelik olarak değişen özel teknoloji ve yöntemleri gerektirir (Gökalp, Kayabay, Çoban, Yandık ve Eren, 2018b, s. 95). Bu bağlamda toplanan verilerin toplu olarak işlendiği “yığın veri işleme”, akan veri üzerinde anlık işlem yapılan “eş zamanlı akan veri işleme” gibi farklı uygulama mimarileri bulunmaktadır (Kayabay ve diğerleri, 2016, s. 6).

Yığın veri işleme, toplanan verilerin toplu olarak işlenmesidir. Bu yönteme örnek olarak, bir şirketin gün içinde topladığı verileri gün sonunda işleyerek günlük rapor oluşturması veya belirli bir dönemde topladığı verileri işleyerek aylık veya yıllık raporlar oluşturması gösterilebilir. Bu yöntem verimli ve güvenilir olmakla birlikte, toplanan verinin hacmine bağlı olarak işlem süresi uzun olabilmektedir (Benjelloun ve diğerleri, 2020). Yığın veri işlemede en fazla kullanılan teknolojilerden biri Hadoop’tur (Kayabay ve diğerleri, 2016, s. 6). Hadoop 2006 yılında Yahoo tarafından geliştirilen açık kaynak kodlu bir büyük veri işleme ekosistemidir (“10 Years of Hadoop”, 2016). HDFS (Hadoop Distributed File System - Hadoop Dağıtık Dosya sistemi) ve MapReduce teknolojilerini kullanmaktadır. HDFS, verilerin farklı makinelere (cluster) dağıtılarak saklandığı

bir yapıyı ifade etmektedir. Veri kaybını engellemek amacıyla dosyalar kopyalanarak farklı makinelerde saklanmaktadır. MapReduce dağıtık sistemler üzerindeki verilerin toplu olarak işlenmesini sağlayan yöntemdir ancak zor ve maliyeti yüksektir. YARN ise Hadoop'un kaynak yönetimi aracıdır. Hangi makinanın ne kadar kaynak kullanacağını belirler ("Apache Hadoop", t.y.)

Hadoop Java, Hadoop Mapreduce, Apache Pig, Apache Hive gibi farklı MapReduce geliştirme yöntemleri kullanmaktadır ("MapReduce vs. Pig vs. Hive", 2022). Spark, Flink ve Beam yığın veri işlemede kullanılan diğer teknolojilerdir (Kayabay ve diğerleri, 2016, s. 9).

Eş zamanlı akan veri işleme ise, işlem süresini kısaltmaya yönelik bir teknolojidir. Bu yöntemde saklanan veriler yerine sisteme akan veriler kullanılarak işlem yapılır. Bu amaçla kullanılan büyük veri işleme teknolojilerinden biri Spark'tır. Spark da Hadoop'ta olduğu gibi verileri dağıtık işlemle analiz etmektedir ancak Spark'ta Hadoop'tan farklı olarak dosya depolama sistemi bulunmamaktadır ve veriler RAM üzerinde tutulmaktadır. Bu özelliği ile Spark belirli işlemlerde, Hadoop'tan yaklaşık olarak 40 kat daha hızlı olabilmektedir. (Zaharia ve diğerleri, 2012). Spark ile de yığın veri işleme mümkün olmakla birlikte hızlı veri işleme özelliği nedeniyle nesnelerin interneti uygulamaları için çok uygundur (Kayabay ve diğerleri, 2016, s. 6; Kini ve Pai, 2024). Storm, Samza, S4, Flink Streaming ve Beam akan veri işlemeye olanak sağlayan diğer teknolojilerdir (Kayabay ve diğerleri, 2016, s. 9; Islam, Borovica-Gajic ve Karunasekera, 2022). Büyük veri işleme mimarileri, verinin özelliklerine ve işlenme gereksinimlerine göre farklılık gösterir. Yığın veri işleme, topluca veri işlemek için kullanılırken, eş zamanlı akan veri işleme anlık ve gerçek zamanlı verileri işlemek için kullanılır. Her iki veri işleme şekli de kendi kullanım alanlarına ve gereksinimlerine göre farklı teknolojiler ve yöntemler kullanarak veri işleme süreçlerini optimize eder (Pishgoo, Azirani ve Raahemi, 2021).

2.1.6. Üst Veri Yönetimi

Veri mimarileri, büyük miktarda veri ile çalışırken verilerin anlamlandırılması ve yönetilmesi için doğru üst veri sistemlerine ihtiyaç duyar. Üst veri, verinin nereden geldiği, hangi işlemlerden geçtiği, nasıl işleneceği ve hangi süreçlerde kullanılacağı gibi kritik bilgileri sağlar. Etkin bir üst veri yönetimi olmadan, veri depoları hızla düzensiz veri yığınlarına dönüşme riski taşır. Bu durum, veri keşfi, analiz ve işleme aşamalarında ciddi zorluklara yol açabilir (Suriarachchi ve Plale, 2016). Üst veri, veri gölleri içinde farklı fonksiyonlar ve işlevler sağlayarak, veri göllerinin düzenli ve organize bir şekilde tutulmasını sağlar. Her veri mimarisi modeli için, üst veri yönetimi, verilerin kullanılabilirliğini artırmak, verilerin doğru amaçlar için kullanılmasını sağlamak ve kullanıcılar arasında etkin erişim kontrolü uygulamak açısından kritik bir süreçtir.

Üst veri, verilerin anlamlandırılmasını sağlayarak verinin kullanılabilirliğini artıran önemli bir bileşendir. Üst veri sayesinde veri keşfi ve organizasyonu daha düzenli hale gelir ve milyonlarca veri arasından doğru veriye hızlı erişim sağlanır. Özellikle büyük veri setlerinde, üst veri yönetimi, verilerin keşfini ve analizini kolaylaştırarak daha verimli bir veri altyapısı oluşturur. Ayrıca üst veri, verilerin yönetimi ve erişim kontrollerini düzenlemek için önemli bir altyapı sunar. Üst veri yönetimi, veri yönetişimi süreçlerinde büyük bir zorunluluk haline gelmiştir ve düzenleyici gerekliliklere uyum sağlamada kritik bir rol oynamaktadır. Sawadogo, Kibata ve Darmont (2019) üst veri yönetiminin, veri keşfi, entegrasyonu ve analizi için önemini vurgulamaktadır. Etkili bir üst veri stratejisi, organizasyonların veri varlıklarını daha iyi anlamalarına, yönetmelerine ve bu verilerden değer elde etmelerine olanak tanır. Aynı zamanda üst veri, verilerin kalitesini ve doğruluğunu kontrol etmek için de kullanılır. Veri kaynakları hakkında doğru üst veri bilgisine sahip olunmadığında, veri kalitesi düşebilir ve bu durum analizlerin doğruluğunu olumsuz etkileyebilir (Miloslavskaya ve Tolstoy, 2016). Bu nedenle hem veri yönetişimi hem de veri kalitesi açısından üst veri yönetimi, verilerin etkin kullanımı için kritik bir rol oynamaktadır.

Bununla birlikte, veri mimarilerindeki verilerin miktarı ve çeşitliliği arttıkça üst veri yönetimi daha karmaşık hale gelir. Özellikle yapılandırılmamış ve yarı yapılandırılmış verilerle çalışırken, veri bataklığı riski daha belirgin hale gelir. Etkili bir üst veri yönetimi yapılmadığında, bu veriler hızla kontrolsüz bir yığın haline gelebilir ve verilerin verimli kullanımı mümkün olmayabilir. Ayrıca üst verinin sürekli olarak güncellenmesi ve izlenmesi gereklidir; veriler sürekli eklenir, değiştirilir ve işlenir, bu durum ise sürekli bir yönetim ve izleme sürecini zorunlu kılar. Büyük veri kümelerinde bu süreçler, genellikle otomasyon araçları ve veri yönetim platformları ile yürütülmektedir. Örneğin, otomatik veri kataloğu araçları, veri kaynaklarını düzenli olarak tarayarak yeni eklenen veya değiştirilen veriler için üst veriyi güncelleyebilir (Quix, Hai ve Vatov, 2016). Ayrıca, yapay zekâ ve makine öğrenimi gibi teknolojiler üst veri analizi ve sınıflandırmasında etkin bir şekilde kullanılabilir. Veriler göllere sürekli olarak eklenmekte ve değiştirilip işlenmektedir. Bu durum ise sürekli bir yönetim ve izleme sürecini zorunlu kılmaktadır. Bu süreçler, özellikle büyük ölçekli veri mimarilerinde önemli bir iş yükü ve maliyet yaratabilir (Sawadogo ve diğerleri, 2019).

Üst veri yönetiminin kritik bir unsuru da dijital nesnelere kalıcı tanımlayıcılar (Persistent Identifier – PID) atamaktır (Kahn ve Wilensky, 2006; CNRI, 2023). Bu tanımlayıcılar, verinin fiziksel konumu veya barındırıldığı platform değişse bile aynı kaynağa atıfta bulunabilmeyi mümkün kılar. Böylece veri keşfi, veri yönetişimi ve veri kalitesi gibi alanlarda tutarlılık sağlanır. Özellikle akademik yayıncılıkta en yaygın kullanılan PID sistemlerinden biri, DOI (Digital Object Identifier) olup bilimsel makale ve veri kümelerine kalıcı bir adres atar (International DOI Foundation, 2022). Handle Sistemi de yaygın olarak kullanılan bir diğer örnektir. Sistem, veriye “prefix/suffix” yapısında bir kimlik (örneğin hdl:20.500.12345/ABCDE) tanımlar ve bu kimlik, “handle constant” (tutma sabiti) olarak adlandırılan değişmez bir göstergedir. Bu sayede büyük veri ortamlarında veri yığınları arasında güvenilir, kalıcı ve doğrulanabilir bir bağlantı kurmak kolaylaşır.

2.1.6.1. Üst Veri Türleri

Üst veri, farklı işlevlere sahip çeşitli türlerde olup, veri yönetim sistemlerinde kritik bir rol oynar. Veri mimarilerinde etkin bir yönetim sağlamak için bu üst veri çeşitliliği oldukça önemlidir.

Yapısal üst veri, verilerin formatı, yapısı ve teknik özellikleri hakkında bilgi sunar. Bu tür üst veri, verinin boyutu, yapısı (yapılandırılmış, yarı yapılandırılmış ya da yapılandırılmamış), saklama formatı ve depolama yeri gibi teknik bilgileri içerir. Özellikle veri yöneticileri ve bilim insanları için, verinin nasıl işleneceği konusunda rehberlik sağlar (Karkosková, 2023).

İşlevsel üst veri, verilerin nasıl işlendiği, hangi işlemlerden geçtiği ve kullanım amacı hakkında bilgi verir. Bu üst veri, örneğin verilerin hangi temizleme süreçlerinden geçtiğini veya hangi algoritmalarla analiz edileceğini gösterir.

Veri güvenliği açısından ise güvenlik üst verisi, verilerin kimler tarafından erişilebileceği, nasıl korunacağı ve hangi güvenlik protokollerinin uygulanacağı konularında bilgi sağlar (Holom, Rafetseder, Kritzinger ve Sehrschön, 2020). Türkiye'deki Kişisel Verilerin Korunması Kanunu (KVKK) gibi regülasyonlar doğrultusunda, güvenlik üst verisi kişisel verilerin korunmasında kritik öneme sahiptir ve doğru kullanıcıların belirlenen amaçlar için verilere erişimini denetler.

Küresel üst veri ise veri kümeleri arasındaki ilişkileri ve veri bütünlüğünü sağlayarak, farklı veri kaynaklarının birbiriyle nasıl ilişkilendirileceği konusunda bilgi sunar. Bu üst veri türü, veri mimarilerindeki çeşitli veri kümelerinin uyumluluğunu ve bütünlüğünü sağlayarak analiz süreçlerinin verimliliğine katkı sağlar (Milić, Veljković ve Stoimenov, 2021b).

2.1.7. Üst Veri (Metadata) Modelleri

Üst veri (metadata) modelleri, verilerin tanımlanması ve yönetilmesi için kullanılan standartlaştırılmış yapılardır. Farklı üst veri modelleri, veri setlerinin

kullanılabilirliğini ve entegrasyonunu sağlar. Bu modeller, verilerin tanımlanması, düzenlenmesi ve bağlamlarının anlaşılmasına ve veri kümeleri arasındaki ilişkilerin ortaya çıkarılmasına yardımcı olur (Atlan, 2024b).

Sınırlılıkları ve kabiliyetleri farklılık gösteren çeşitli üst veri modelleri bulunmaktadır. Kamusal verilerin kullanılmasına yönelik olarak üst veri modellerinin değerlendirilmesi önemlidir. Üst veriler farklı veri kümeleri arasındaki ilişkilerin ortaya çıkarılmasını da sağlar (Milić, Veljković ve Stoimenov, 2021a, s. 119). Verilerin çeşitli özelliklerini tanımlayan yayıncı, yayın tarihi, format vb. üst veri bilgileri, verinin daha iyi tanımlanmasını ve kullanılmasını sağlar. “Üst veri bir veri kaynağını tanımlayan ve açıklayan, veri almayı, kullanmayı veya yönetmeyi kolaylaştıran yapılandırılmış bilgilerdir.” (Milić ve diğerleri, 2021a, s. 120). Üst veri kısaca “veri hakkında veri” olarak tanımlanmaktadır (Duval, 2001, s. 591). Üst veri, veri platformlarının temel bileşenlerinden biridir ve verilerin etkin bir şekilde kullanılması için kritik öneme sahiptir.

Aşağıda yaygın olarak kullanılan üst veri modelleri açıklanmaktadır:

2.1.7.1. DCAT (Data Catalog Vocabulary)

DCAT (Data Catalog Vocabulary), veri kümelerini tanımlamak, paylaşmak ve kataloglamak için kullanılan, W3C (World Wide Web Consortium) tarafından geliştirilen bir standarttır. DCAT, veri setlerinin daha keşfedilebilir ve erişilebilir hale getirilmelerini sağlar. Bu standart, veri kümelerinin üst verilerini tanımlamak için RDF (Resource Description Framework) üzerine kurulmuştur. Dcat’in dcat:Catalog, dcat:Resource, dcat:Dataset, dcat:Distribution, dcat:DataService, dcat:DatasetSeries, dcat:CatalogRecord olmak üzere 7 temel sınıfı bulunmaktadır. DCAT, web üzerinde yayınlanan veri katalogları arasında birlikte çalışabilirliği kolaylaştırmak için tasarlanmış bir sözlüktür (“Data Catalog Vocabulary”, 2024).

2.1.7.2. Schema.org

Schema.org, web sayfalarının içeriklerini arama motorlarının daha iyi anlamasını sağlamak için kullanılan bir yapılandırılmış veri işaretleme sözlüğüdür. Google, Microsoft, Yahoo ve Yandex iş birliğiyle oluşturulmuştur ve web sitelerinde yer alan çeşitli veri türlerinin (makaleler, etkinlikler, ürünler, tarifler gibi) tanımlanmasını sağlar. Schema.org verileri, HTML'e eklenen Microdata, RDFa ve JSON-LD gibi biçimlerde uygulanabilir ve böylece arama motorları, zengin sonuçlar sunarak kullanıcılara daha yararlı bilgiler sağlayabilir (“About schema.org”, t.y.).

2.1.7.3. CKAN Üst Veri Şeması

CKAN (Comprehensive Knowledge Archive Network), açık veri portalları için geliştirilmiş açık kaynaklı bir yazılımdır. Açık veri platformları arasında en yaygın olarak kullanılanlardan biri olup, veri setleri için başlık, tanımlayıcı, açıklama, revizyon geçmişi, lisans vb. üst verileri içeren zengin bir üst veri seti sunar. Özellikle büyük veri projelerinde, veri kümelerinin bulunabilirliği ve erişilebilirliğinin artırılması amacıyla hükümetler, araştırma kurumları ve büyük miktarda veri toplayan diğer kuruluşlar tarafından sıkça kullanılır (“CKAN Overview”, t.y.; “CKAN Metadata”, t.y.). CKAN üst veri şeması ise CKAN platformu üzerinde yayımlanan veri setlerinin tanımlanmasını ve yönetilmesini sağlar. CKAN üst veri şeması, esnek yapısı sayesinde kullanıcıların veri setlerine özel ek alanlar tanımlamasına ve üst verileri yönetmesine olanak tanır. Kanada, Singapur ve Avustralya hükümetleri tarafından çeşitli açık veri platformlarında bu yapı kullanılmaktadır (“CKAN the World’s”, t.y.).

2.1.7.4. ISO 19115 - Coğrafi Bilgi Üst Veri Standardı

ISO 19115, Coğrafi Bilgi Üst Veri Standardı, coğrafi kapsamı olabilecek bilgi ve kaynakları tanımlamak ve kataloglamak, öğeler arasındaki ilişkileri tanımlamak için bir model sağlamayı amaçlayan uluslararası bir standarttır (“GeoNetwork Opensource”, t.y.). Bu standart, coğrafi veri kümelerinin kimliği, kapsamı, mekansal ve zamansal yönleri gibi üst verilerinin belirlenmesini sağlar. Ayrıca veri setlerinin daha geniş kapsamlı tanımlanması için isteğe bağlı ek üst veri elemanları içerir (“ISO Online”, 2014).

2.1.7.5. Dublin Core Üst Veri Element Seti

Dublin Core Üst Veri Element Seti, kaynakları tanımlamak ve bulmak için kullanılan 15 üst veri elemanını içeren uluslararası bir standarttır. İlk olarak 1995 yılında Ohio'da düzenlenen bir çalıştayda oluşturulmuş, zamanla gelişerek birçok farklı sektör ve uygulama tarafından benimsenmiştir. "Core" (çekirdek) terimi, elemanların geniş ve esnek kullanımlarını ifade etmek için kullanılmıştır. Bu 15 eleman, kaynakların başlık, yaratıcı, konu, tarih, format gibi temel bilgilerini tanımlar. Dublin Core, web üzerindeki dijital kaynakların organizasyonu ve keşfi için geniş çapta benimsenmiş bir üst veri sözlüğüdür. ISO 15836-1:2017 ve ANSI/NISO Z39.85-2012 gibi uluslararası standartlar tarafından da kabul edilmiştir (Weibel, 2009; Dublin Core Metadata Initiative, 2020).

2.1.7.6. CERIF (Common European Research Information Format)

CERIF, Avrupa Birliği'nin üye ülkelere araştırma faaliyetleriyle ilgili bilgilerin toplanması, yönetimi ve paylaşımı için önerdiği standarttır. Araştırma bilgi sistemleri için yapılandırılmış bir ilişkisel veri tabanı modeli sağlar. Araştırma faaliyetlerine ait verilerin düzenli bir şekilde saklanmasını ve raporlanmasını

hedefler. Bu standardın amacı, üye ülkeler arasında araştırma verilerinin uyumlu hale gelmesi ve veri paylaşımının kolaylaşmasıdır (Jörg, t.y.).

2.1.7.7. MARC (Machine-Readable Cataloging)

MARC, kütüphaneler tarafından kullanılan bir üst veri standardıdır. Kitaplar, dergiler vb. materyallerin kataloglanmasında yaygın olarak kullanılır. Kütüphaneler tarafından kataloglanan materyalerin makine tarafından okunabilmesi için geliştirilmiştir. Kongre Kütüphanesi tarafından başlatılan bir girişim tarafından geliştirilen bir veri formatıdır. Bu format günümüzde kullanılan çoğu kütüphane kataloğunun temelini oluşturmaktadır. MARC 1990'ların sonlarında MARC 21 olarak adlandırılmıştır ve bu isimle kullanılmaktadır (“MARC Standards”, 2006).

2.1.7.8. MODS Üst Veri Nesne Tanımlama Şeması

MODS (Metadata Object Description Schema) 2002 yılında Kongre Kütüphanesi'nin Ağ Geliştirme ve MARC Standartları Ofisi tarafından, özellikle kütüphane uygulamaları için kullanılabilecek, bibliyografik veri tanımlamaları için geliştirilmiştir. MARC 21 kayıtlarından seçilen verileri taşımak ve orijinal kaynak tanım kayıtları oluşturmak amacıyla bir XML şeması olarak tasarlanmıştır. MODS, MARC alanlarının bir alt kümesine ek olarak ve MODS'a özgü üst veri alanları içermektedir ve dil tabanlı etiketler kullanmaktadır (“The Library of Congress”, 2022).

2.1.7.9. FGDC (Federal Geographic Data Committee) Üst Verisi

FGDC üst verisi, Coğrafi Bilgi Sistemleri (GIS) dosyaları, haritalar ve diğer konum tabanlı veri kaynaklarını tanımlayan bir üst veri standardıdır. Coğrafi verilerin

keşfi ve yönetimi için gerekli verileri sağlar. ABD federal hükümeti tarafından coğrafi verilerin paylaşımı ve standardizasyonu için kullanılmaktadır (“FGDC.gov”, t.y.).

Üst veri modelleri veri kümelerinin kullanım amacına ve ihtiyaçlarına göre değerlendirilmelidir. Değerlendirme modelin mevcut veri altyapısı ve sistemleri ile ne kadar uyumlu olduğu, veri türleri ve yapılarına uygunluk esnekliği, veri setlerini ne kadar kapsamlı bir şekilde tanımlayabildiği, modelin uygulanmasının ve kullanılmasının ne kadar kolay olduğu, uluslararası standartlara uygunluk ve geniş kabul görmüşlük durumu ve veri yönetimi ve keşfinde ne kadar etkili olduğu dikkate alınarak yapılmalıdır.

2.2. ARAŞTIRMA VE KAMU VERİSİ

2.2.1. Araştırma Verisi

Araştırma verisi, “Bilimsel çalışmalarda birincil kaynak olan ve araştırma sonuçlarını doğrulamak için kullanılan sayısal çıktılar, metinsel kayıtlar, görseller ya da sesler gibi maddi kayıtlar” olarak tanımlanmaktadır (*TÜBİTAK Açık Bilim Politikası*, t.y., s. 2).

Daha önce belirtildiği gibi verinin değeri verinin araştırmalara kattığı değerle ölçülebilir (Doğan ve Arslantekin, 2016, s. 19). “Araştırma verileri, bir argümanın, teorinin, testin, hipotezin veya diğer araştırma çıktılarının dayandığı gerçekler, gözlemler, görüntüler, bilgisayar programı sonuçları, kayıtlar, ölçümler veya deneyimler biçimindeki veriler anlamına gelir. Araştırma projeleri sırasında üretilen, toplanan veya kullanılan verilerle ilgilidir ve bazı durumlarda araştırma çıktısının kendisini içerebilir. Veriler sayısal, tanımlayıcı, görsel veya dokunsal olabilir. Ham, temizlenmiş veya işlenmiş olabilir ve herhangi bir formatta veya ortamda tutulabilir. Araştırma verileri, birçok disiplinde, yazılımın, algoritmanın veya modelin uygulandığı ham verilere ek olarak, araştırma sonucuna ulaşmak için kullanılan yazılım, algoritma, model ve/veya parametreleri zorunlu olarak

içerebilir.” (“QUT Manual of Policies”, 2013). Araştırma verilerinin etkili bir şekilde yönetilmesi, yalnızca araştırma sonuçlarının doğruluğunu değil, aynı zamanda veri paylaşımını ve yeniden kullanılabilirliği de artırır (Briney, Coates ve Goben, 2020).

Araştırma projeleri çok farklı çeşitte veri üretir ve toplar. Bir veri yönetim planı oluşturmak için veriler kaynağına, biçimine, kararlılığına ve hacmine göre sınıflandırılabilir. Araştırma verilerinin farklı disiplinlerde sınıflandırılması, araştırma süreçlerini daha verimli hale getirebilir ve disiplinler arası iş birliğini artırabilir (Weber, Kranzlmüller, Fromm ve Tavares de Sousa, 2020)

Kaynağına göre veri gözlemsel veri, deneysel veri, simülasyon verisi ve türetilmiş/derlenmiş veri olarak 4 türe ayrılabilir:

Gözlemsel Veri: Genellikle laboratuvar dışında, gerçek zamanlı olarak yakalanan bu nedenle çoğunlukla yeniden elde edilmesi mümkün olmayan ve özenle korunması gereken sensor okuması, görüntüler, anket vb. veridir.

Deneysel Veri: Genellikle laboratuvar ortamında veya farklı bir kontrollü ortamda elde edilen, bu nedenle tekrarlanması mümkün olan simülasyon, gen dizilimi, kromatogram vb. Veridir.

Simülasyon Verisi: Oluşturulan test modellerinden makine tarafından oluşturulan, tekrar üretilmesi mümkün olan iklim modelleri, ekonomik modeller vb. Veridir

Türetilmiş/Derlenmiş Veri: Mevcut veri kümeleri kullanılarak üretilen, tekrarlanabilir ancak çok pahalı ve zaman alıcı olabilen, veri madenciliği, üç boyutlu modeller vb. veridir.

Biçimine göre veri, laboratuvar notları, anket yanıtları vb. metin verisi; tablolar, ölçümler vb. sayısal veri; resimler, ses kayıtları, video vb. görsel-işitsel veri; modeller, bilgisayar kodları; kimyada CIF, astronomide FITS vb. disipline özgü veri; ekipman çıkışı vb. cihaza özel veri olarak türlere ayrılabilir. (“DMP tool”, 2023)

Bir proje süresince veya proje sona erdiğinde bu projeye ait veri sabit kalabilir veya değişebilir. Zamana karşı kararlılığına göre veri sabit veri kümesi, büyüyen veri kümesi ve revize edilebilir veri kümesi olarak üçe ayrılır:

Sabit Veri Kümesi: Oluşturulduktan veya toplandıktan sonra değişmeyen veri kümesidir.

Büyüyen Veri Kümesi: Bulundurduğu eski veri değiştirilemeyen veya silinemeyen, ancak yeni veri eklenebilen veri kümesidir.

Revize Edilebilir Veri Kümesi: Bulundurduğu eski veri değişebilen veya silinebilen, ayrıca yeni veri eklenebilen veri kümesidir. (“DMP tool”, 2023)

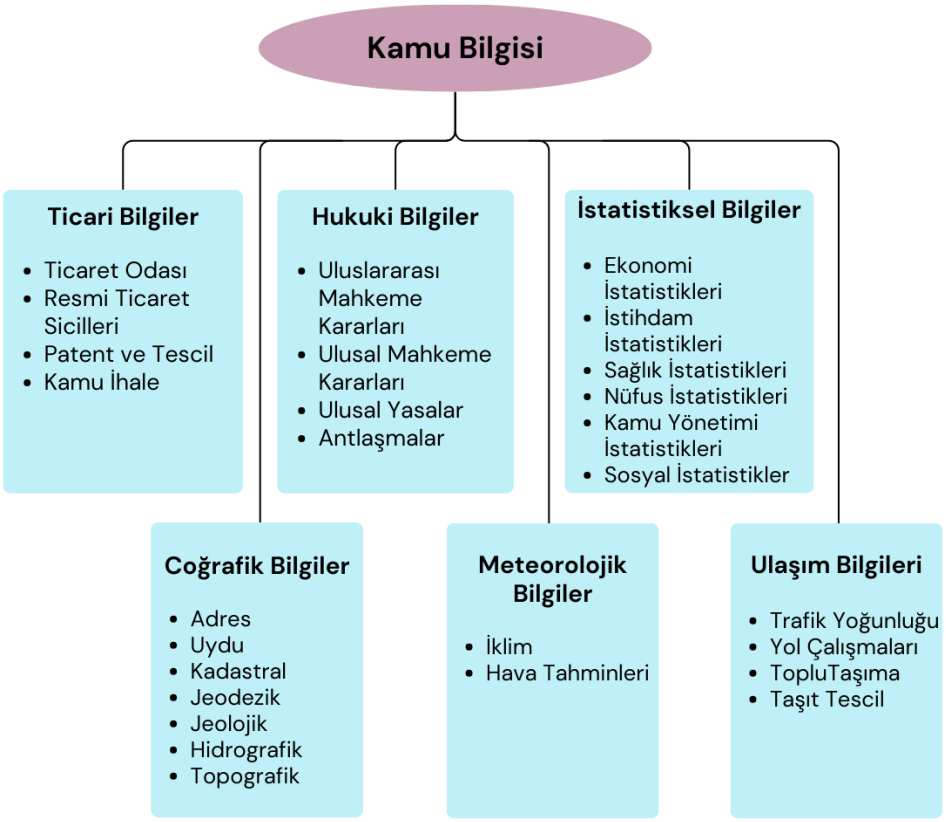
Veride yaşanması muhtemel değişime göre veri yönetim sürecinin nasıl yürütüleceği konusunda bir plan yapılması gerekir. Veri yönetimi, veri kümesinin sabit, büyüyen veya revize edilebilir olmasına bağlı olarak farklı stratejiler ve yöntemler gerektirir. Yine bir projede ne kadar veri üretileceği önemlidir. Verinin manuel veya makine ile toplanacak olması, projede ortaya çıkacak verinin nasıl saklanacağı, saklanmayacak veri varsa buna nasıl karar verileceği, verinin büyüme oranının tahmin edilmesi, ortaya çıkacak verinin boyutuna göre depolama ve yedekleme kapasitesinin önceden planlanması önemlidir. (“DMP tool”, 2023).

Araştırma verileri yönetimi, araştırmalarda kullanılacak verilerin yaşam döngüsü boyunca gerektirdiği tüm süreç ve eylemleri kapsamaktadır. Araştırma verilerinin yaşam döngüsü ise verilerin oluşturulmasından kullanılmasına kadar olan bütün aşamaları kapsar. Yaşam döngüsü çoğunlukla verinin oluşturulduğu proje sonuçlandıktan sonra da devam eder (“QUT Manual of Policies”, 2013).

2.2.2. Kamu Verisi

Kamu verisi kamu kurumlarının ürettiği ve topladığı her tür veri olarak tanımlanabilir ve genellikle büyük hacimli ve farklı türdeki verilerden oluşur

(Dekkers, Polman, Velde ve Vries, 2006, s. 25). Kamu kurumları sundukları hizmetlerle yüksek miktarda veri toplamakta ve oluşturmaktadır (Topçu ve Işık, 2019, s. 834). Şekil 3'te genel hatlarıyla kamu kurumlarından elde edilebilecek bilgilerin sınıflandırması verilmiştir (Dekkers ve diğerleri, 2006, s. 25).



Şekil 3. Kamu bilgisi sınıflandırması (Dekkers ve diğerleri, 2006, s. 25)

Kamu bilgisi, ticaret odası, resmi ticaret sicilleri, patent ve tescil, kamu ihaleleri gibi ticari bilgiler; adres, uydu, bina, kadastral, jeodezik, jeoloji, hidrografik, topografik gibi coğrafik bilgiler; uluslararası ve ulusal mahkeme kararları, ulusal yasalar ve antlaşmalar gibi hukuki bilgiler; iklim ve hava tahminleri gibi meteorolojik bilgiler; ekonomi, istihdam, sağlık, nüfus, kamu yönetimi ve sosyal istatistikler gibi istatistiksel bilgiler; trafik yoğunluğu, yol çalışması, toplu taşıma ve taşıt tescil gibi ulaşım bilgileri olarak sınıflanabilir (Dekkers ve diğerleri, 2006, s. 25).

Bu çalışmada kullanılan kamu kurumları kavramı üniversiteler, TÜBİTAK vb. araştırma kurumlarını kapsadığı için, kamu verisi kavramı araştırma verilerini de içerecek şekilde ele alınmıştır.

2.3. AÇIK VERİ

Budapeşte Açık Erişim İnisiyatifi bildirisine göre açık erişim; “Hakem değerlendirmesinden geçmiş bilimsel literatürün, internet aracılığıyla; finansal, yasal ve teknik engeller olmaksızın, serbestçe erişilebilir, okunabilir, indirilebilir, kopyalanabilir, dağıtılabilir, basılabilir, taranabilir, tam metinlere bağlantı verilebilir, dizinlenebilir, yazılıma veri olarak aktarılabilir ve her türlü yasal amaç için kullanılabilir olması” şeklinde tanımlanmıştır (“Budapest Open Access Initiative”, 2002).

Açık bilim, “Bilimin; yayınların, araştırma verilerinin, laboratuvar notlarının ve diğer araştırma süreçlerinin ücretsiz erişilebildiği, araştırmanın yeniden kullanımı, dağıtımı ve üretilmesine izin veren koşullarla, diğer araştırmacıların birlikte çalışabileceği ve katkıda bulunabileceği şekilde yapılması.” (*TÜBİTAK Açık Bilim Politikası*, t.y., s. 2) olarak tanımlanmıştır.

Açık veri ise, “herhangi bir telif hakkı, patent ya da diğer kontrol mekanizmalarına tabi olmaksızın herkes tarafından ücretsiz ve özgürce kullanılabilen, tekrar kullanılabilen ve dağıtılabilen veri olarak tanımlanmaktadır.” (*TÜBİTAK Açık Bilim Politikası*, t.y., s. 2). Açık veri ile sağlanan kolay erişim, kamu, özel sektör ve vatandaşların oluşturduğu sistem ile farklı sektörlere yeni fırsatlar doğurmaktadır. Örneğin açık veri platformları, kullanıcıların veriye anonim ve güvenli bir şekilde erişmesine olanak sağlayarak inovasyonu destekler (Laufs, Peters ve Schultz, 2022). Devletin topladığı veriyi, açık veri standartlarına uygun olarak kullanıma sunduğu açık veri platformları bu fırsatlardan en fazla etkiye sahip olanıdır (Özkan, 2019, s. 10).

Açık verinin anonimleştirilerek kullanılması verideki kişisel ve hassas verinin açığa çıkmasını önleme bakımından kritik öneme sahiptir. Verinin biçimi korunarak kişisel ve hassas verilere erişim engellenir. Bu konuda “Federe Öğrenme” ve “Diferansiyel Mahremiyet” gibi farklı mahremiyet modelleri geliştirilmektedir (“Türkiye Cumhuriyeti”, t.y.a). Son zamanlarda, açık veri platformlarının güvenliğini artırmak ve veri bütünlüğü ile doğruluğunu sağlamak amacıyla blokzincir tabanlı modeller kullanılmaya başlanmıştır. Bu modeller, verinin izlenebilirliğini ve değişmezliğini sağlayarak açık veri platformlarının güvenilirliğini artırmaktadır (Dang ve Anh, 2021).

2.3.1. Açık Devlet Verisi

Açık veri herhangi bir kısıtlama olmadan işlenebilir, kullanılabilir ve üçüncü kişiler tarafından değişiklik yapılabilir veridir. Açık veri kurum ve kuruluşlara birlikte çalışabilirlik imkânı sağlayarak daha fazla verim elde edilmesine katkı sağlar (“Türkiye Cumhuriyeti”, t.y.a).

Açık devlet verisi, kamu kurumları tarafından üretilen ve toplanan verilerin, belirli standartlara uygun olarak ve herkesin erişimine açık olacak şekilde yayımlanmasını ifade eder. Bu veriler, şeffaflığı artırmak, vatandaşların katılımın ve kullanımını sağlamak, yenilikçi çözümler geliştirmek ve ekonomik değeri artırmak amacıyla paylaşılır. Böylece kamu verilerin daha etkin bir şekilde kullanılması hedeflenir.

Dünyada kullanılan farklı açık devlet verisi platformlarına örnek olarak CKAN, DKAN, Socrata, OpenDataSoft, OGDI, OGPL verilebilir. CKAN, ABD'de data.gov, İngiltere'de data.gov.uk tarafından; DKAN, İtalya'da data.gov.it, ABD'de healthdata.gov tarafından; Socrata, Kanada'da data.edmonton.ca, ABD'de data.cityofchicago.org tarafından; OpenDataSoft, Fransa'da data.sncf.com tarafından; OGDI Fransa'da opendata.bordeaux.fr, Kolombiya'da datos.gov.co

tarafından; OGPL ise Hindistan'da data.gov.in tarafından kullanılmaktadır (Milćı, Veljković ve Stoimenov, 2021, s. 119).

Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi bünyesinde başlatılan AçıkVeri projesi ile veriden değer üretmek amaçlanmaktadır. Veriye erişimin sınırlı olması yapay zekâ ve yenilikçi teknolojilerde gelişmeye engel oluşturur. Bu proje kapsamında oluşturulacak açık veri portalı aracılığıyla, anonimleştirilerek mahremiyeti sağlanmış verinin paylaşılması, böylece yapay zekâ ve yenilikçi teknolojilerin ülkemizde geliştirilmesi hedeflenmektedir (“Türkiye Cumhuriyeti”, t.y.a).

“Akıllı kuruluş” olabilmek için öncelikle verinin etkin şekilde kullanılması gerekmektedir. Bu hedefe ulaşmak için kişisel ve hassas verilerin anonimleştirilmesi ve bu verilerin kontrollü bir şekilde kullanıma açılması büyük önem taşımaktadır. Son yıllarda, verinin artan stratejik önemi doğrultusunda birçok ülke, açık veri politikaları çerçevesinde devlet verilerini paylaşmaya başlamıştır (Attard, Orlandi, Scerri ve Auer, 2015, s. 399). Açık devlet verisi, kamu çalışmalarının şeffaflığını artırmanın yanı sıra, hesap verebilirliği güçlendirmektedir. Ayrıca açık devlet veri platformları aracılığıyla, devlet verilerinin paylaşılarak ilgili paydaşlar tarafından yeniden kullanılması mümkün hale gelmektedir. Pek çok hükümet, verilerini kamuya açık hale getirmek için açık devlet verisi portalları geliştirmiştir (Kassen, 2019; Bates, 2014). Bu platformlar, yalnızca kamuya açık verilerin paylaşımını değil, aynı zamanda kurum içi verilerin düzenli, güvenli ve erişilebilir şekilde saklanmasını da sağlamaktadır. Böylece, kurum içi verilerin analiz edilmesi yöneticilere daha bilinçli ve etkili karar alma imkânı tanırken, hizmet kalitesinin artırılmasına da katkıda bulunur. Kamu kurumlarının şeffaf, erişilebilir ve kullanışlı bir veri yönetim anlayışını benimsemesi, araştırma, inovasyon ve gelişim için önemli fırsatlar yaratmaktadır.

2.3.2. Açık Verinin Faydaları

Açık veri hem akademik çalışmalarda hem de politika geliştirme, yeni iş birlikleri kurma, bürokratik süreçleri azaltma, teşvik mekanizmaları oluşturma ve ekonomik değer yaratma amaçlarıyla kullanılabilir. Açık veri kamu ve özel sektör için çeşitli faydalar sağlamaktadır. Açık devlet verisi, kamu hizmetlerinin şeffaf ve hesap verilebilir olmasına katkı sağlar, ayrıca vatandaşların toplumsal katılımını olanaklı hale getirir. Açık verinin faydaları aşağıdaki şekilde gruplandırılabilir.

- Verinin veriyi üretenler ve toplayanlar dışında farklı paydaşlar tarafından da kullanılabilmesi ve yenilikçi çözümler geliştirilmesi (Haberer ve Schnurr, 2020),
- Açık verinin kullanımının artmasıyla, sektörler arası çapraz veri kullanımının mümkün olması ve böylece etkinliğin artarak maliyetlerin düşmesi,
- Verinin haritalanması ve yayınlanmasına yönelik çalışmalar ile veri kullanımı için zemin oluşturulması ve kalitenin arttırılması,
- İhtiyaç duyulan verinin açık olarak kullanıma sunulmasıyla birlikte yeni iş modellerinin ortaya çıkması (Laufs, Peters ve Schultz, 2022),
- Hizmetlerin şeffaf ve hesap verilebilir olmasıyla güven ortamının oluşması (Özkan, 2019, s.14).

2.3.2.1. Açık Verinin Kullanıcı Açısından Faydaları

Açık verinin belirli alanlardaki faydalarına aşağıdaki örnekler verilebilir:

- Trafik: Açık verinin sağlayacağı en verimli kazanım özellikle büyük şehirlerde yaşanan trafik sıkışıklığının yarattığı zaman kaybının engellenmesidir (Özkan, 2019, s. 14). Örneğin trafikte yaşanan

problemlerin ve sıkışıklıkların anlık olarak takip edilebilmesini sağlayan böylece alternatif yolların kullanılmasına olanak sağlayan Polonya'da geliştirilen "Warszawski Ninja" gibi uygulamalar mevcuttur ("European Data", 2016). Ülkemizde otobüslerin hangi durağa hangi saatte ulaşacağını gerçek zamanlı olarak gösteren uygulamalar kullanılmaktadır. Yine trafik kazalarının nerede ve ne zaman gerçekleşebileceğini tahmin etmeye yönelik uygulamalar mevcuttur (Özkan, 2019, s. 16).

- Çevre: Açık verinin enerji verimliliğinin gözlemlenmesi, çevre kirliliği ile ilgili bilgilere etkin erişim sağlanması, çevre ile ilgili bilinç oluşturulması ve ilgili kampanyalar için temel oluşturması gibi etkileri bulunmaktadır (*International Open Data*, 2017, s. 30). Örneğin "Plume Labs" isimli uygulama ile dünya üzerindeki farklı bölgelerin hava kirlilik seviyeleri anlık olarak takip edilebilmektedir ("Plume Labs", t.y.).

- Sağlık: Sağlık verileri için bütün Avrupa ülkelerinin açık veri portallarında ayrı bir bölüm bulunmaktadır ve en fazla indirilen veriler sağlık verileridir (Özkan, 2019, s. 16). Amerika Birleşik Devletleri'nde sağlık veri tabanlarının açılmasıyla 25000 ölümün engellenebileceği belirtilmektedir (Ferenstein, 2013). Yine sağlık verileri ile ilk yardım noktalarının nerede olması gerektiği belirlenebilmektedir. Yine farklı sağlık kuruluşlarının belirli hastalıklardaki tedavi oranlarının açıklanmasıyla daha başarılı hastanelerin nasıl tedavi uyguladıklarına yönelik paylaşım yapması sağlanabilir (*Creating Value*, 2015, s. 89).

- İtfaiye: İtfaiye ekiplerinin olaylara müdahale için kullanacakları yolun uygunluğu ve olayın gerçekleştiği binada tehlikeli maddeler bulunup bulunmadığını öğrenmeleri olaya daha az riskle ve daha hızlı şekilde müdahale etmelerini sağlayacaktır (Özkan, 2019, s. 17).

- Enerji: Hane halkı veya kurum ve kuruluşların enerji tüketim miktarlarının paylaşılmasıyla bireylerin bu konudaki farkındalıklarının artarak enerji tüketimlerini azaltmaya yöneldikleri belirtilmektedir (*Creating Value*, 2015,

s. 96). Avrupa Veri Portalı'nın 9. Analitik raporunda açık veri kullanımının enerji tüketimini %16 azaltabileceği belirtilmektedir (Berends, Carrara ve Radu, 2020).

- Güvenlik: Bireylerin bulundukları ortamda saldırıya maruz kalma riskini puanlayan uygulamalar bulunmaktadır. "LuckyMe" uygulaması bunlardan biridir. Kişinin bulunduğu ortama ait ambulans verisi, polis verisi ve kullanıcı raporları kullanılarak, örneğin bir sokağın, risk oranı belirlenebilmektedir. Kişiler sokak hakkında puanlama yapabilmekte ve yaşadıkları kavga, saldırı vb. olayları raporlayabilmektedir (Özkan, 2019, s. 17).

- Bilim: Bilimsel çalışmaların ve çalışma verilerinin kamuya açılmasıyla birlikte aynı verinin farklı çalışmalarda kullanılabilmesi veya benzer konuda çalışan araştırmacıların araştırmalarına yön vermesi mümkün olmaktadır. Ülkemizde de açık arşivler konusunda olumlu gelişmeler yaşanmakla birlikte araştırmacıların bu konudaki farkındalık düzeyleri yetersiz bulunmaktadır (Tonta, 2015, s. 247).

2.3.2.2. Açık Verinin Kamuya Faydaları

Avrupa Açık Veri Portalı'nın 2015 yılında kullanıma açılmasıyla birlikte açık veri platformları yaygınlaşmaya başlamış, verinin vatandaşların kullanımına açık hale gelme süreci hız kazanmıştır. Açık veri politikalarıyla girişimcilerin, araştırmacıların hatta isteyen her vatandaşın istediği sebeple veriye erişebilmesi, veriyi kullanması ve paylaşması, yani verinin tekrar kullanılabilmesi hedeflenmektedir. Böylece aynı verinin farklı şekilde yorumlanması, farklı amaçlarla tekrar analiz edilmesi ve farklı kazanımlar elde edilmesine olanak sağlanır, veriden elde edilecek bilginin sadece veriyi sağlayan kurumun tekeline bırakılmasının önüne geçilir. Ayrıca veri "küresel ekonominin yeni sermayesidir". Değerli olan verinin varlığı değil açık olması, böylece inovasyon faaliyetlerinde

kullanılabilmesidir. Kendi alanlarındaki veriye ulaşabilen firmalar veriyi kullanarak sundukları ürün ve hizmetleri geliştirebilirler, bu durum sadece firmalara değil bu ürün ve hizmetleri kullanan tüm vatandaşlara ve böylece devletlere de fayda sağlar (Khurshid ve diğerleri, 2020). Açık veri politikaları şeffaflık bakımından da önemli olup vatandaşların ülke genelindeki uygulamalar hakkında fikir sahibi olmasına ve geri bildirim vermesine böylece uygulamaların iyileştirilmesine olanak sağlar (Kawashita, Baptista ve Soares, 2022). Açık verinin mevcut kullanımına ek olarak yakın gelecekte farklı amaçlarla kullanılması ve henüz tahmin edilemeyen farklı faydalar sağlanması beklenmektedir. Kamunun sunduğu veriden birçok farklı alanda değer elde edilmesi mümkün olmakla birlikte bunlardan en önemlisi kamunun sağladığı verinin kamu yararı için kullanılmasıdır (Özkan, 2019, s. 18). Açık veri politikalarının kamu kurumlarınca uygulanmasının çok yönlü faydaları bulunmaktadır:

- Performans ölçümü
- Kamu kurumlarının sunduğu hizmetlerin denetimi
- Kamu kurumlarındaki iyi uygulama örnekleri
- Stratejik kararlar verme
- Üretkenliğin artması
- Kamu hizmetlerinin daha verimli yürütülmesi
- Veri toplama ve yönetim maliyetlerinin azalması
- Farklı verilerin anlaşılmasını kolaylaştırma ve ekonomic riskleri azaltma
- Yasal ve zorunlu raporlama gerekliliklerini karşılama (*Open Data,* 2012, s. 12).

Kamu kurumlarının uygulayacağı açık veri politikaları veri güvenliği, gelir kaybı, inceleme geçirme kaygısı, açık veri süreçlerinin yaratacağı giderler gibi

çekincelere de neden olmaktadır. Yine açık veri konusunda yeterli bilincin olmaması da açık veri uygulamaları için engel oluşturabilmektedir. Açık veri devlet ve özel sektör arasında daha güçlü ortaklıklar kurulmasını ve sunulan hizmetlerin kalitesinin artmasına katkı sağlamaktadır (Özkan, 2019, s.19). Bu bölümde literatür değerlendirmesi ve belge tarama yöntemleri kullanılarak elde edilen bilgiler ve sonuçlar, tezde yer alan "Veri paylaşımı ve erişilebilirlik mekanizmalarındaki eksiklikler, kamu verisinin potansiyel toplumsal faydalarını sınırlamaktadır" hipotezini destekler niteliktedir. Bu hipotez, kamu verisinin etkin bir şekilde kullanılabilmesi ve toplumsal faydalarının artırılabilmesi için mevcut paylaşım ve erişim süreçlerinde iyileştirme gerekliliğine işaret etmektedir.

2.3.2.3. Açık Verinin Özel Sektör Açısından Faydaları

Açık veri hem kamu kurumlarının hem de özel sektöre ait verilerin açılmasını kapsar. Örneğin kriz anlarında akıllı telefonlardaki lokasyon verisinin kullanılması veya uydulardan yol durumu, elektrik kullanımı gibi verilerin kullanılması kriz yönetimini kolaylaştıracak verilerdir. Yine Twitter platformu aracılığıyla atılan tweet verileri problemlerin belirlenmesi amacıyla kullanılabilir (Özkan, 2019, s.20). Açık verinin özel sektör için çok yönlü faydaları bulunmaktadır:

- Hukuki yükümlülüklerin sağlanması
- Veri ile ilgili hizmetlerin satılması
- Hükümet ile iş birliği yapılması
- Yatırım ve yeniliklere olanak sağlaması
- İş ortaklarının müşterileri daha iyi anlaması
- Endüstrinin ve yerel işletmelerin desteklenmesi (*Open Data,* 2012, s. 18).

2.4. VERİ ANALİTİĞİ

Bilgisayar teknolojilerindeki gelişmeler basılı kayıtlardan dijital arşivlere geçişi sağlamış, bu geçiş arşivlerin doğasını değiştirmiştir. Bu gelişmelerden en önemlisi daha hızlı, erişebilir ve ucuz hesaplama gücüdür. Her geçen gün hesaplama işlemlerinin hızı artmakta, ancak hız artışı maliyeti arttırmamaktadır. Bu gelişimin ne kadar ileri seviyelere ulaşacağı ise tahmin edilememektedir. Aynı zamanda bilgisayar hafızalarının maliyeti azalmaktadır ve maliyetin de azalmaya devam edeceği düşünülmektedir. Bu gelişmelerden daha da önemlisi, bilgisayar ağlarının genişlemesi, dijital transfer ve bilginin uzaktan işlenmesi yeteneklerinin artmasıdır. (Mitchell, 1996, s. 201). Dijital arşivleme sistemlerinde yapay zekâ ve otomasyonun entegrasyonu, arşivlerin hem yönetimini hem de erişimini daha verimli hale getirmiştir (Colavizza, Blanke, Jeurgens ve Noordegraaf, 2021). Bilgisayar teknolojilerindeki bu ilerlemeler, bilgi yönetimi ve arşivleme süreçlerini köklü bir şekilde değiştirmektedir. Daha hızlı ve ucuz hesaplama gücü, azalan hafıza maliyetleri ve genişleyen bilgisayar ağları sayesinde dijital arşivler, daha verimli, erişilebilir ve sürdürülebilir hale gelmiştir. Gelişen veri analitiği ve yapay zekâ tabanlı optimizasyon teknikleri, dijital arşivlerin büyük veri kümelerini daha etkin bir şekilde yönetmesine olanak tanımaktadır (Li ve Zhang, 2022). Bu süreç, sadece arşivleme yöntemlerinin dijitalleşmesini değil, aynı zamanda bilginin saklanma, düzenlenme ve işlenme biçimlerinde de köklü değişiklikler yaratmıştır. Gelişen teknolojiler sayesinde büyük veri kümeleri kolaylıkla analiz edilebilmekte, veriye dayalı karar destek sistemleri geliştirilmektedir. Böylece geçmişte fiziksel sınırlamalardan dolayı erişimi zor olan arşivler, artık dünyanın her yerinden anında erişilebilir hale gelmeye başlamıştır.

Dijitalleşmeyle birlikte, makine öğrenmesi ve veri madenciliği araçları hemen hemen bütün bilim dallarında aktif olarak kullanılan yöntemler haline gelmiştir (LeCun ve diğerleri, 2015). Dijitalleşme; resim, yazı, video ve ses gibi bütün medya araçlarının kullanımı için uygun format sunmaktadır ve bütün bu araçların benzer yöntemler ile paylaşılabilmesine olanak sağlamaktadır (Weller, 2011, s. 8). Bütün bu gelişmeler sonucunda veri her geçen gün daha büyük bir hızla

artmaya devam etmektedir. Ancak veriden enformasyon elde edilebilmesi için verinin derlenmesi, bunun için de verinin belirli kurallar dahilinde kaydedilmesi gerekir (Doğan ve Arslantekin, 2016, s. 17). Verinin değeri verinin araştırmalara kattığı değerle ölçülebilir (Doğan ve Arslantekin, 2016, s. 19). Sadece 2017 yılında önceki 5000 yılda üretilenden daha fazla veri üretilmiştir. Enerji, imalat, sağlık, uzay, nesnelerin interneti, siber güvenlik, sosyal medya veri üretilen alanların bir kısmıdır. Bu verinin ancak %0,5'i operasyonel olarak analiz edilmektedir. Kuruluşların, büyük miktarda veriyi doğru karar vermelerine yardımcı olacak şekilde sunma yeteneğine sahip platformları kullanması kritik olacaktır. Bu süreç ise işletmeler arasında büyük rekabete neden olmaktadır. “Eğer veri yeni petrolse, onu kim işleyecek ve rafine edecektir?” (Bell, 2018, s. 2). Dijitalleşme ile verilere hızlı ve kolay erişim sağlanması bilgi yönetiminde büyük bir dönüşüm yaratmıştır. Veriyi doğru bir şekilde işleyip analiz eden kurumlar, bu süreçten en büyük faydayı sağlayacaklardır.

Verinin işlenerek bilgi elde edilmesi süreci öncelikle verinin seçilerek örneklem kümesinin belirlenmesi, önişleme, verinin indirgenmesi, veri madenciliği teknikleriyle verideki örüntülerin ortaya çıkarılması ve yorumlanıp doğrulanarak bilgi elde edilmesi aşamalarından oluşur (Fayyad, Piatetsky-Shapiro ve Uthurusamy, 1996’dan aktaran Sever ve Oğuz, 2002, s. 177)

2.4.1. Veri İşleme ve Metodolojiler

Veriyi işlemeye yönelik çeşitli araçlar bulunmaktadır. Ancak bu araçlar çeşitli nedenlerle kurum ve kuruluşlar tarafından yeterince kullanılamamaktadır. Veri işlemenin karmaşıklığı ve bu süreçleri yönetebilmek için gerekli olan teknik bilgi eksikliği bu nedenlerin başlıcalarıdır. Bu problemi ortadan kaldırmak amacıyla son yıllarda kütüphaneler ve veri akışı prensibine dayanan çeşitli modeller geliştirilmektedir (Kayabay ve diğerleri, 2016, s. 2).

Verinin kullanılabilmesi için öncelikle belirli kurallar dahilinde kaydedilmesi gerekir. Ancak bu şekilde veriden enformasyon ve bilgi edinilmesi sağlanabilir. Dijital bir belgeyi okumak, o belgeyi oluştururken kullanılan veya onun eşdeğeri bir yazılım gerektirir. Metin dosyaları için bu genellikle bir problem oluşturmamakla birlikte, bilgisayar destekli tasarım dosyaları gibi büyük modelleri içeren belgeleri okumak çoğunlukla belgenin oluşturulduğu yazılımla mümkün olmaktadır. Ancak yazılımların ömrü çok kısadır, bu nedenle arşivlenen belgelerin ileride okunmasını sağlayan teknolojilerin sağlanması dijital arşivler için önemli bir problemdir. Yazılım tekrar oluşturulsa dahi çoğu zaman ilgili yazılım hangi donanım için tasarlandıysa o donanıma da ihtiyaç duyulmaktadır ve donanımların ömrü de çok kısadır. Örneğin Boston'daki Bilgisayar Müzesi, bilgisayarların ilk dönemlerinden kalma bir koleksiyona sahip olmasına rağmen, bunların çok azı çalışabilir durumdadır. (Mitchell, 1996, s. 203). Dijital arşivlerin uzun vadeli korunması ve erişimi, yazılım ve donanım bağımlılığını azaltacak stratejiler gerektirir. Bu stratejiler ile bilgi kaybının önlenmesi sağlanabilir (Baldin ve Eliseev, 2020).

Verinin analiz edilebilmesi için öncelikle işlenerek analiz araçlarıyla okunabilecek formata dönüştürülmesi gerekir. Koleksiyonlar genellikle disipline özgü veya oluşturulduğu araca özel özellikler içermektedir; bu nedenle verilerin ilk önce belirli standartlar doğrultusunda yeniden biçimlendirilmesi gerekir. Örneğin bir dijital kitap arşivi her bir kitabın her bir sayfası için ayrı ASCII dosyası içerir ve her sistem veya yazılım bu kadar sayıdaki küçük dosyayı işleyebilecek yeteneğe sahip değildir. Analiz yazılımları için genellikle her bir kitabın tek bir dosya olarak görülmesi gerekir, bu amaçla verinin işlenerek her bir sayfa için ayrı ayrı olan ASCII dosyalarının her kitap için tek bir dosya olarak biçimlendirilmesi gerekir. Yine PDF, EPUB, DjVu formatındaki belgelerin metinlerini ortaya çıkarmak için de ön işleme gerekir. Metin içeriğinin ortaya çıkarılmasında XML standardı yaygın olarak kullanılmakla birlikte tek bir XML dosyası olarak indirilebilen arşivlerde dahi ilgilenilen kısımları ayırmak için yine ön işleme aşaması gerekli olmaktadır. Ön işleme sonrasında verinin analiz için kullanılacak yazılıma uygun formatta saklanması gerekir; veri çok fazla yazılımda kullanılacaksa farklı formatlara kolayca dönüşümü mümkün kılan ara formatta kaydedilmesi daha verimli

olacaktır (Leetaru, 2012, s. 21). Verinin analiz için uygun formatta saklanması ve ön işleme süreci, verinin analiz araçlarıyla uyumlu hale gelmesini sağlar ve böylece verinin kullanılabilirliği artar ve analiz süreci daha etkili yürütülebilir.

Belgeler aslında "büyük sözcük koleksiyonları"dır. Bir arşivde yer alan bilgiyi sadece sözcük sıklıklarıyla anlamaya çalışmak, arşivde keşfedilebilecekleri sınırlar. Bu nedenle veri çalışmalarında "bilginin farklı boyutlarını yakalayan, uzayı, zamanı ve duyguyu belirten kelimeleri tanıyan daha yüksek dereceli temsillerin oluşturulması" analizlerle belgelerin "gerçek anlamı"nı ortaya çıkarmaya fayda sağlar. Veri analizi çalışmalarının ilk dönemlerinde belgelere ait farklı göstergeler ayrı ayrı incelenmiş, bu göstergeler arasındaki ilişkiler yeterince incelenmemiştir. Yeni nesil Büyük Veri araştırmalarında arşivlerde yer alan bağlantıları ve ilişkileri anlamaya daha fazla çalışılmaktadır. Ancak bu bağlantıları keşfetmek hesaplamalarda birtakım zorlukları beraberinde getirmektedir (Leetaru, 2012, s. 22). Bu zorlukları aşmak için gelişmiş analitik teknikler, doğal dil işleme, veri görselleştirme ve yapay zekâ gibi çözümler kullanılabilir.

Veri arşivleri oluşturulduktan sonra genellikle ilk önce çeşitli veri madenciliği algoritmalarıyla isim, yer adı (konum) gibi önemli veriler çıkarılarak veya okunabilirlik, duygu gibi metne ait özellikler belirlenmeye çalışılarak verilere ait üst veri katmanları oluşturulur (Pepper, Senin, Jebbia, Breen ve Greenberg, 2022). Bu üst veri daha sonra veriye erişim ve veri analizi için kullanılır. Her çalışmanın gerektirdiği veri madenciliği uygulamaları ve üst veri gereklilikleri farklı olmakla birlikte, "konum" bilgisinin gelişmekte olan ve kullanımı yaygınlaşan bir üst veri bilgisi olduğu belirtilmektedir. Yine üst veri olarak "zaman" bilgisi de yaygın olarak kullanılmaktadır (Leetaru, 2012, s. 22). Hızla artan büyük verinin analizinde verinin uzamsal ve zamansal boyutlarının işlenmesi, veriden anlamlı sonuçlar elde etmek için etkin olarak kullanılmaktadır. Zaman, bilgi alanında merkezi bir rol oynar; bilgi çıkarma, konu belirleme, sorgu analizi ve özetleme gibi çeşitli alanlarda çalışılmıştır (Alonso, Strötgen, Baeza-Yates ve Gertz, 2011). İlgili çalışmalarda genellikle makine öğrenmesi metotlarını kullanarak dokümanları birbirleriyle ilişkilendirme yöntemleri kullanılmaktadır (Martins ve Calado, 2010).

Bir dokümanın üst verileri yoluyla doğrudan erişilebilen zaman bilgisi, zamana duyarlı arama veya zamansal kümeleme gibi çeşitli görevler için kullanılabilir (Alonso ve diğerleri, 2011). Metindeki duygu ise duygu madenciliği araçlarıyla belirlenir. Bu araçlarla metinde geçen ve belirli duyguları temsil eden sözcük sayıları belirlenerek metinde hâkim olan duygu hakkında yüzeysel de olsa bir tahmin elde edilir (Leetaru, 2012, s. 23). Konum, zaman ve duygu gibi üst veri bilgileri, verinin daha derinlemesine anlaşılmasına yardımcı olur ve veri analizi sürecinde önemli bir rol oynar.

2.4.1.1. Veri Türüne Göre Veri İşleme

Verilerin analize hazırlanması için ön işleme süreçlerinden geçirilmesi gerekir. Bu işlemler farklı veri türleri için farklıdır ve aşağıda bu işlemler açıklanmıştır.

2.4.1.1.1. Analog Verilerin Koşullandırılması

Dönüşüm veya şekilsel değişim olarak adlandırılan koşullandırma işlemleriyle analog veriler analiz için uygun bir forma dönüştürülür. Bu işlemlere geçmişte veri azaltma veya sıkıştırma adı verilmiştir. Veri azaltmanın amacı depolanan veri hacmini ve verilerin işlenmesi için gereken iş miktarını azaltmaktır. Her veri kümesi için veri azaltma miktarı farklıdır. Veri azaltma için farklı teknikler kullanılmaktadır. Bu tekniklerin bir kısmı aşağıda kısaca açıklanmıştır:

Veri Çıkarma: İhtiyaç duyulmayan verinin kaldırılması işlemi olup oldukça yaygın olarak kullanılan kullanışlı bir veri azaltma yöntemidir. Örneğin çok hassas olmayan bir ölçüm değeri için 23,6867789 olarak küsuratlı şekilde verilen bir değerin uygulamada yalnızca ilk iki basamağı anlamlıdır. Yani bu ölçümü 23,69'a yuvarlamak ve bu değeri saklamak verinin anlamı bakımından yeterlidir ve büyük ölçüde yer tasarrufu sağlar. Bir diğer veri çıkarma yöntemi ise eşik değerdir. Örneğin elektronik gözle kontrol edilen bir üretim hattında üretilen parçanın belirli

ölçü aralığında olması gerekir ve sadece belirlenen eşik değerlerin dışında kalan verileri kaydetmek yeterlidir. Boyu 3,347 cm'den uzun, 3,340 cm'den kısa olmadığı sürece parça uyumludur ve bu aralığın dışında kalan parçaları belirlemek yeterli olacaktır (Inmon, 2016, s. 69-70).

Veri Kümeleme: Veri azaltmada kullanılan bir diğer yöntem kümelemedir. Bu yöntemde belirli özellikleri ortak olan, örneğin belirli sayılar arasında kalan sayılar, aynı küme altında gruplandırılır.

Örneğin aşağıda verilen ölçüm verilerini ele alalım:

4,12 - 4,15 - 4,45 - 4,24 - 4,37 - 4,46 - 4,14 - 4,22 - 4,43 - 4,48

Bu değerleri aşağıdaki şekilde gruplamak mümkündür: Değeri 4,10 ile 4,19 aralığında olan 3, 4,20 ile 4,29 aralığında olan 2, 4,30 ile 4,39 aralığında olan 1, 4,40 ile 4,49 aralığında olan ise 4 sayı bulunmaktadır.

4,1 - 3

4,2 - 2

4,3 - 1

4,4 - 4

Bu kümelemeye aşağıdaki gibi verinin sıra numarası da dahil edilebilir:

4,1 - (1), (2), (7)

4,2 - (4), (8)

4,3 - (5)

4,4 - (3), (6), (9), (10)

Örnekte kullanılan veri sayısı çok az olduğu için kümeleme işleminin veri azaltma potansiyeli tam olarak yansıtılamamakla birlikte, kümeleme işlemiyle sayıları

temsil etmek için gereken alan miktarının büyük ölçüde azaltılması potansiyeli bulunmaktadır. Yine verilen örnekten çok daha karmaşık ve çeşitli kümeleme yöntemleri bulunmaktadır (Inmon, 2016, s. 71-72).

Veri İlişkileri: Veri ilişkileri, ölçüm verileri arasında ilişki kurulmasıdır. Veri ilişkileri, çeşitli ölçüm verileri arasında anlamlı bağlantılar kurularak daha derin analizlerin yapılmasını sağlar. Örneğin bir yerel yönetimin bütçe harcamalarına ait veriler ele alındığında, bu harcamalar hizmet alanlarına (ör. altyapı, sağlık, eğitim) göre sınıflandırılıp, nüfus yoğunluğu ve kişi başına düşen harcama gibi verilerle ilişkilendirilebilir.

- **Altyapı**: Harcama: 50 milyon TL, Nüfus yoğunluğu: 300 kişi/km², Kişi başına harcama: 167 TL
- **Sağlık**: Harcama: 30 milyon TL, Nüfus yoğunluğu: 500 kişi/km², Kişi başına harcama: 60 TL
- **Eğitim**: Harcama: 40 milyon TL, Nüfus yoğunluğu: 400 kişi/km², Kişi Başına harcama: 100 TL
- **Çevre**: Harcama: 20 milyon TL, Nüfus yoğunluğu: 200 kişi/km², Kişi başına harcama: 100 TL
- **Ulaşım**: Harcama: 60 milyon TL, Nüfus yoğunluğu: 600 kişi/km², Kişi başına harcama: 100 TL

Farklı verilerle ilişkilendirilmiş veriler çok daha fazla kullanılabilir olmakta, bu nedenle daha fazla talep edilmektedir (Inmon, 2016, s. 73-75).

Gelecekte Kullanım Olasılığı: Analog veri göletleriyle ilgili tasarım süreçlerinde gölete yerleştirilecek verinin gelecekteki kullanım olasılığı dikkate alınmalıdır. Bir verinin gelecekte kullanım olasılığı yoksa bu veriyi gölette tutmaya gerek yoktur, veri göletten çıkarılabilir. Verinin gelecekte kullanım olasılığı çok düşükse bu veri gölete daha az göze çarpan bir şekilde yerleştirilir. Gelecekte kullanım olasılığı yüksek olan veriler gölette daha belirgin bir yerde tutulur. Verinin gelecekte kullanım olasılığını her zaman doğru tahmin etmek mümkün olmayacağı için

kullanılma ihtimalinin çok az olduğu düşünülen verileri silmek yerine daha az göze çarpan yere taşımak daha güvenli olacaktır (Inmon, 2016, s. 76).

Aykırı Değerler: Veri grubunun modeline uymayan verilerdir. Genel olarak modelden küçük farklılıklar bulunabilir ancak çoğu ölçüm tanımlanabilir bir modele uyar ve tahmin edilebilir. Bu modele ve verinin doğal değişkenliğine uymayan değerler aykırı değer olarak adlandırılır. Aykırı değerler üzerinde özel olarak çalışılması gereken değerlerdir.

Örneğin belirli bir hat üzerindeki telefon görüşmelerinin uzunlukları analiz edildiğinde bu görüşmelerin çoğunluğunun 5-6 dakika sürdüğü ancak 3 aramanın 24 saatten uzun olduğu tespit edilmiştir. Uzun aramalar araştırıldığında bir aramanın verileri aktaran bilgisayarla çalışan bir bilgisayar olduğu, bir aramanın ekipman arızası nedeniyle uzun görünen kısa bir görüşme olduğu, bir aramanın ise film indirmek için yanlış bir hattı kullanan bir müşteri olduğu görülmüştür.

Kuruluşlar aykırı değerleri inceleyerek bu değerleri veri kümesinden çıkarmaya, veri kümesini bu değerleri de içerecek şekilde yeniden tanımlamaya veya ölçümlerin dağılımını tanımlayan yeni bir algoritma ile farklı bir veri seti oluşturmaya karar verebilirler.

Tekilleştirme (Deduplication): Yinelenen verilerin, veri yığınlarının kaldırılmasıdır.

Eksizyon (Excision): Analiz için ihtiyaç duyulmayan ve duyulma ihtimali olmayan verilerin kaldırılmasıdır.

Sıkıştırma (Compression): Verinin çok sıkı bir şekilde paketlenmesidir. Ancak sıkıştırılmış verilerin değiştirilmesine ihtiyaç duyulduğunda değişiklik yapılması oldukça zordur.

Düzleştirme (Smoothing): Verideki aykırı değerlerin kaldırılması veya düzenlenmesi işlemidir.

İnterpolasyon (Interpolation): Verinin kendisine yakın değerler kullanılarak olası değerinin bulunmasıdır.

Örnekleme (Sampling): Veriyi temsil eden alt bir veri kümesinin belirlenmesidir. Analitik işlemler için kullanılabilir ancak detaylı güncelleme işlemleri için kullanılamaz.

Yuvarlama (Rounding): Küsuratların yuvarlanarak kaldırılması işlemidir.

Kodlama (Encoding): Uzun dizilerin daha kısa dizilerle temsil edilmesidir (Inmon, 2016, s. 67).

Tokenizasyon (Tokenization): Verilerde yüksek miktarda tekrar olduğu durumlarda etkili olarak kullanılan bir kodlama biçimidir.

Eşik (Threshholding): Veri için bir sınır değerin belirlenmesi, bu değerin altındaki veya üzerindeki değerlerin saklanmasıdır.

Kümeleme (Clustering): Bir veri tekilleştirme yöntemi olan kümeleme işlemi ile verilerdeki benzer ve kesin değerler gruplanır.

Veri, yukarıda açıklanan uygun işlemlerden geçirilerek koşullandırılır ve analistin kullanımına sunulur (Inmon, 2016, s. 78-80). Veriye uygun olarak veri azaltma tekniklerinden biri veya birkaçı uygulanabilir.

2.4.1.1.2. Uygulama Verilerinin Koşullandırılması

2.4.1.1.2.1. Standart Verit Tabanı Formatı

Genellikle uygulamalar satır ve sütun biçiminde verileri depolarlar ve veri mimarilerine veri bu standart formda aktarılır. Ancak verinin veri tabanına ait standart formatta olması bu düzenin taşınması anlamına gelmez. Veri, ilgili veri mimarisine aktarıldıktan sonra mimarinin teknolojisine göre yönetilir ve bu teknoloji genellikle standart bir veri tabanı yönetim sisteminden farklı olmaktadır (Inmon, 2016, s. 89).

Uygulama verilerinin kaynaklarının genellikle farklı olması nedeniyle kayıtlar ayrı ayrı gruplandırılarak organize edilir. Bu kayıtların öznitelikleri vardır ve bu öznitelikler anahtar veya indeks olabilmektedir (Inmon, 2016, s. 89).

2.4.1.1.2.2. Veri Entegrasyonu

Uygulama verilerinin saklandığı birime ulaşan veriler eğer entegre bir yapıda gelmiyorsa, bu veriler sisteme aktarıldıktan sonra ilk olarak entegre edilmeleri gerekir. Entegre iş yapısına sahip olması, verilerin ana konu alanları doğrultusunda düzenlenmesidir. Verinin analiz edilebilmesi için bu zorunludur. Genel olarak kurumsal konular müşteri, ürün, sipariş, teslimat vb. olabilmektedir. Uygulama veri göletindeki verinin entegrasyonunun sağlanması için gölet içinde bir veri modelinin olması gerekir. Genellikle uygulama veri göletlerinde kurumsal veri modelleri bulunur, eğer kurumsal veri modeli bulunmuyorsa genel iş modelleri bulunur. Çoğunlukla kurumsal veri ambarının veri modeli, uygulama veri göleti için de uygun model olmaktadır (Inmon, 2016, s. 90-91).

Veri modelinin birçok avantajı vardır. Avantajlardan biri, veri modelinin verilerin nasıl ilişkilendirilmesi gerektiği konusunda üst düzey rehberlik sağlamasıdır. Bu çerçevede veri modeli aslında uygulama veri göleti için bir üst veri oluşturarak gölet içindeki önemli öğeler için bir kılavuz sağlar. Uygulama veri göletindeki üst veriler, kayıtları ve bunların anlamlarını, niteliklerini ve anlamlarını, anahtarları, dizinleri, veri ilişkilerini vb. verilerin ayrıntılı bir tanımını verir (Inmon, 2016, s. 92).

Uygulama veri göletleri verileri uzun bir süre boyunca tutar ancak veri modelinin kendisi zamanla değişir. Bu nedenle uygulama veri göletlerinin veri modelinin esnek bir yapıda olması gerekir. Uygulama veri göletindeki veriler analiz edilirken üst veri üzerinde zaman içinde yapılan değişikliklerin bilinmesi gerekir. Bu nedenle uygulama veri göletlerinin veri modeli çok karmaşıktır.

Veriler uygulama veri göletine geldiğinde entegre edilmemiş durumda ise dönüştürülmesi gerekir. Bu dönüştürme işlemi, analog veri göletlerindeki verilere uygulanan koşullandırma işlemine çok benzerdir.

Aşağıda bu veri dönüştürme sürecinde yapılan işlemlere örnekler verilmiştir:

Aşağıda farklı uygulamalardan gelen cinsiyet verisi görülmektedir.

Uygulama A - cinsiyet - x, y

Uygulama B - cinsiyet - 0,1

Uygulama C - cinsiyet - F, M

Uygulama D - cinsiyet - kadın, erken

Bu farklılığın analizlerde sorun oluşturmaması için entegre edilerek ortak bir yapıya dönüştürülmesi gerekir. Bütün cinsiyet verileri aşağıdaki gibi aynı biçime dönüştürülmelidir:

Cinsiyet - F, M

Yine farklı birim veya standartlara göre bulunan uzunluk ölçüleri ortak tek bir birime dönüştürülerek saklanmalıdır: Örneğin, santimetre, metre, inç, mil vb. birimlerdeki ölçüler metre birimine dönüştürülerek standartlaşma sağlanmalıdır. Farklı para birimleri, farklı sıcaklık ölçüm birimleri vb. farklılıklar entegre edilerek veri göleti için belirlenen standart biçime dönüştürülmelidir.

Yukarıda verilen örnekler basitçe entegrasyon dönüşümlerini açıklamak amacıyla verilmiştir, uygulama veri göletlerinde çok daha kapsamlı dönüştürme işlemleri yapılmalıdır. Aksi takdirde verinin anlamlı olarak analiz edilmesi mümkün değildir (Inmon, 2016, s. 93-95).

İki uygulamanın birlikte yer aldığı işlemlerde bir uygulamadan diğerine bir işaretçi (pointer) kullanılır. Örneğin bir müşteri bir konser bileti satın aldığında hem müşteri uygulamasının veri tabanı hem de bilet satış uygulamasının veri tabanı

bulunabilir. Müşteri veri tabanı müşterilerin isimlerini, bilet veri tabanı ise konsere ait tarih, saat ve satılan bilet numaralarını tutabilir. Bu iki uygulamanın verisi entegre edildiğinde hangi müşterinin hangi tarih ve saatteki konser için hangi bilet numarasına sahip olduğu elde edilebilir.

Kesişen uygulamalar olduğunda ise durum biraz daha karmaşıklaşır. Bu durumda iki uygulamanın kesişiminden bağımsız veriler oluşur. Örneğin, bir valilik ve bir sosyal hizmetler müdürlüğü ele alındığında, iki kurumun gerçekleştirdiği işlemler sonucunda oluşan verilerin kesiştiği noktalar bulunabilir.

Valilik tarafından sağlanan bir hizmet kaydı aşağıdakine benzer bir veri içerebilir:

Valilik: Ankara Valiliği

Hizmet Türü: İhtiyaç sahiplerine gıda yardımı

Sosyal Hizmetler Müdürlüğü ise aynı süreçle ilgili şu veriyi oluşturabilir:

Kurum: Çankaya Sosyal Hizmetler Müdürlüğü

Yardım Alan Kişi: Ahmet Yılmaz

Adres: Üniversiteler Mahallesi, 32. Cadde, No:25, Çankaya, Ankara

Yardım Türü: Gıda Paketi

Miktar: 20 kg

Bu iki veri setinin kesişim noktaları (örneğin adres veya yardım türü) bulunmaktadır.

Bazı durumlarda analist bir uygulamanın bütün verilerini veri göletine göndermek yerine sadece uygulama veri tabanındaki verilerin belirli bir alt kümesini belirleyerek veri göletine gönderebilir. Örneğin bir müşteri hizmetleri telefon kayıtları verilerinin sadece 4 dakika ve üzerinde olan kısmı veri göletine analiz için

aktarılabilir. Bu seçim sistemin yükünü ve yapması gereken işi oldukça azaltır (Inmon, 2016, s. 100).

2.4.1.1.3. Metin Verilerinin Koşullandırılması

2.4.1.1.3.1. Metinsel Belirsizlik Giderme

Metinsel belirsizlik giderme (Textual Disambiguation), metnin standart bir veri tabanı formatına dönüştürülmesi için kullanılan bir veri koşullandırma yöntemidir. Bu yöntem, metin verilerinin kurumsal karar verme süreçlerinde kullanılma yeteneğine büyük katkı sağlamıştır. Metni okumak, analiz etmek ve metnin bağlamına uygun olarak standart bir formata dönüştürmek için kullanılır. Ancak çoğu şirket bu teknolojiyi kullanmamakta, metinsel veri metinsel veri göletine ham metin olarak ulaşmaktadır. Kurumlar sahip oldukları metinsel veriyi analiz edebilmek için, metinsel belirsizliği gidermeli, metni standart bir formata dönüştürmelidir. Bu işlem yapılmadan ham metin verisi analiz edildiğinde yapılan analiz çok yüzeysel olacaktır.

Yaygın metinsel veri formları, e-postalar ve sosyal medya verileri olmakla birlikte ses transkripsiyonu ve optik karakter tanımlama yoluyla elde edilen metin verileri de bulunmaktadır (Inmon, 2016, s. 104). Metinsel belirsizliği giderme, uygulama verilerine uygulanan koşullandırma süreçlerinden oldukça farklı bir yöntemdir. Bu yöntem ile metin verisinin standart, tek bir biçimde yapılandırılmış bir veri tabanında ve bağlamı ile birlikte saklanması sağlanır. Böylece metin standart analitik işlemciler tarafından okunup analiz edilebilir. Metnin standart bir veri tabanı formatında saklanması için bir kaydın olduğu formda saklanması gerekir. Her bir kayıt bağlamı ile birlikte işlenen metni, metnin baytını ve belgenin adını içerir.

Bu örnek metin bir şahıs ile bir özel şirket arasında imzalanan ve şartları içeren bir kiralama sözleşmesine aittir. Metinsel belirsizliği giderme işleminden geçirilen metin işlenmiş ve belgenin adı, yakalanan metnin bayt adresi, metnin kendisi ve

bağlamını içeren bir veri tabanı formatına indirgenmiştir (Inmon, 2016, s. 107-110).

Dil doğası gereği karmaşıktır ve bu nedenle metinsel belirsizlik giderme de oldukça karmaşık bir süreçtir. Metinsel belirsizlik gidermenin 90'dan fazla farklı alt işlemi bulunmaktadır. Bu amaçla uygulanan işlemlerin bir kısmı aşağıda açıklanmıştır:

Satır içi bağlamsallaştırma: Metni inceleyerek metni ve bağlamını belirleme işlemidir. Örneğin "... kiracı Ela Ankaralı tarafından imzalanan…" metni için kiracı "Ela Ankaralı" olarak tanımlanır. Bu işlem için metnin sözleşme gibi bir veri örneğine sahip olması bağlamının tahmin edilebilir olması gerekir.

Yakınlık: Birbirine yakın kelimeler metnin anlaşılmasında kullanılabilir. Örneğin "...sınavda Hacettepe'yi kazandı." cümlesinde Hacettepe sözcüğü Hacettepe Üniversitesi'ni ifade etmektedir.

Alternatif yazım: Bir sözcüğün nasıl yazıldığı ülkelere veya bölgelere göre farklı olabilmektedir. Örneğin İngiltere'de "centre" şeklinde yazılan sözcük ABD'de "center" olarak yazılmaktadır.

Homografik çözünürlük: Alternatif yazımın karmaşık bir biçimidir. Bir ifadenin anlamının anlaşılması için ifadeyi kimin yazdığının anlaşılması önemlidir. Örneğin "ha" kısaltması bir endokrinolog tarafından hepatit A olarak yorumlanırken, bir pratisyen hekim tarafından baş ağrısı (headache) olarak yorumlanabilir.

Kısaltma çözünürlüğü: Farklı bir alternatif yazım biçimidir. Örneğin "KGM" kısaltması Karayolları Genel Müdürlüğü'nü ifade etmekle birlikte isim değişikliği ile Kütüphaneler ve Yayımlar Genel Müdürlüğü yapılmadan önce Kütüphaneler Genel Müdürlüğü'nün kısaltması olarak da kullanılmıştır.

Özel değişken tanıma: Değişkenin yapısından ne ifade ettiği anlaşılabilir ve yapısı kullanılarak tanımlanabilen birçok değişken bulunmaktadır. Örneğin (555) 555 5555 şeklinde yazılan bir numara telefon numarası olarak yorumlanır.

Taksonomi çözümleme (Taxonomy resolution): Metin belirsizliği gidermenin en önemli özelliğidir. Örneğin bir Volkswagen veya Ford bir arabayı ifade etmek için kullanılır.

Tarih standardizasyonu: Belirli bir tarih farklı şekillerde yazılabilir. Tarih verilerinin standartlaştırılması oldukça faydalıdır ve yaygın olarak kullanılmaktadır. Örneğin 24 Mayıs 2022 veya 2022/05/24 aynı günü ifade etmektedir.

Yukarıda açıklanan yöntemler metinsel belirsizlik giderme amaçlı kullanılan yöntemlerin küçük bir kısmıdır; bu yöntemler dışında birçok farklı yöntem kullanılmaktadır. Metin analizi için metnin işlenmesinden çok daha önemli olan metnin bağlamıdır ve bu oldukça zor bir işlemdir (Inmon, 2016, s. 110-111).

Bir sorgunun sonucu ile ilgili olarak şüphe oluşması durumunda, metinsel belirsizlik giderme usulüne uygun olarak yapılmışsa, ana kaynağa ulaşmak ve gerekli kontrolleri yapmak mümkün olmaktadır. Çünkü metinsel belirsizlik giderme yapılan metinlerin ana kaynağının adı ve baytı da saklanmaktadır ve bu bilgi kullanılarak orijinal kaynağa ulaşmak ve gerekli kontrolleri yapmak mümkün olmaktadır (Inmon, 2016, s. 116).

2.4.1.1.3.2. Taksonomiler ve Ontolojiler

Taksonomiler kısaca terimlerin sınıflandırılması olarak tanımlanabilir. Çok çeşitli taksonomiler bulunmaktadır. Örneğin;

Kurum: Ankara Lisesi, Ankara Valiliği, Hacettepe Üniversitesi

İl: Ankara, İstanbul, İzmir

Ontoloji ise ilgili taksonomilerin gruplanmasıdır. Örneğin;

Ülke: Türkiye, ABD, İspanya, Bulgaristan

Türkiye: Ankara, İstanbul, İzmir, Bursa

Yukarıdaki iki taksonominin ilişkisi Türkiye'nin illerden oluşmasıdır. "Ülke" ve "Türkiye" taksonomileri birlikte bir ontoloji oluşturur.

Neredeyse sınırsız sayıda taksonomi ve ontoloji bulunmaktadır (Inmon, 2016, s. 113-114).

2.4.1.1.3.3. Metin ve Bağlamın Değeri

Bir metnin bağlamı hakkında bilgi sahibi olmak oldukça önemlidir. Basit bir örnekle açıklamak gerekirse, büyük bir metin verisi içinde bir öğretmen olan Ezgi ile ilgili arama yapmak istediğimizde arama yapacağımız metnin bağlamı hakkında bilgimiz yoksa birçok farklı meslekten Ezgi isimli kişilere ulaşırız. Ancak arama yaptığımız metnin bağlamı örneğin "Milli Eğitim" ise sadece öğretmen olan "Ezgi"lere ulaşabiliriz. Bu örnek için sorgu "bağlam=Milli Eğitim olan bütün Ezgi tekrarlarını bul" şeklinde yapılabilir. Metin verisinin bağlamı tam olarak aradığımız veriye ulaşmamızı sağlar, bu nedenle önemli bir koşuldur (Inmon, 2016, s. 115-116). Yine örneğin soyadı "Ankaralı" olan bir kişi için arama yapmak istediğimizde bir metin içinde Ankaralı olduğu belirtilen birçok kişiye ulaşabiliriz ancak metnin bağlamının bilinmesi sadece soyadı "Ankaralı" olan kişilere hızlıca ulaşabilmemizi sağlayacaktır.

2.4.1.1.3.1.4. Metinsel Belirsizlik Gidermenin Mekaniği

Metinsel Belirsizlik Giderme mekaniğine bir örnek olarak duygu belirleme amaçlı taksonomiler gösterilebilir. Duygu belgelerde, sosyal medya yazışmalarında, e-postalarda ve birçok farklı metinde bulunur. Metindeki duygunun belirlenmesi duygu taksonomisi ile yapılır. Örneğin metindeki negatif ve pozitif duygular için basitçe iki farklı taksonomi oluşturulabilir:

Negatif: Nefret, mutsuz, katılmamak, berbat, çirkin…

Pozitif: Sevmek, takdir etmek, iyi, güzel, mutlu… (Inmon, 2016, s. 117).

Gerçekte duygu taksonomisi yukarıda verilen örnekten çok daha geniş anlamlar içermektedir (Inmon, 2016, s. 118).

Metinsel belirsizlik giderme ile ham metin okunur ve taksonomi için belirlenen kelimelerle metnin içeriği eşleştirilir. Eşleşen kelimeler bulunduğunda metnin taksonominin belirlediği duyguyu yansıttığı çıkarımı yapılır ve belgenin tonu ölçülebilir (Inmon, 2016, s. 119).

Duygu belirlemeye büyük bir tekstil zincirinin müşterilerinden gelen geri bildirimlerin analiz edilmesi örnek olarak gösterilebilir. Müşterilerin sunulan ürün ve hizmet hakkındaki yorumları sunulan ürün ve hizmetlerin geliştirilmesi ve müşteri memnuniyeti bakımından kritik öneme sahiptir. Ancak zincir mağazalara her gün çok fazla sayıda geri bildirim yapılmaktadır ve bunların tek tek okunarak yanıtlanması veya özümsenerek geliştirme süreçlerinde kullanılması mümkün değildir, mümkün olsa dahi çok fazla iş gücü gerektirir. Mesajların içeriği ürünün tedarik edilmesinden kalitesine, mağaza çalışanlarından mekanına kadar çok farklı konularda olabilir. Bu nedenle mesajların oluşturulan taksonomilerle kıyaslanarak duygusunun/içeriğinin belirlenmesi ve belirli içeriklere otomatik olarak belirli yanıtlar oluşturularak müşteri memnuniyetine katkı sağlanması mümkün olur (Inmon, 2016, s. 119).

Duygu analizi ile elde edilen sonuçların yorumlanması bu konuda deneyim gerektirir. Çünkü müşteriler genellikle yaşadıkları olumsuz deneyimler için geri bildirimde bulunurlar. Bu nedenle yapılan duygu analizi sonucunda geri bildirimlerin yaklaşık %85'inin olumsuz olması beklenir. Eğer olumsuz geri bildirim oranı %85'ten daha çoksa o zaman bir şeylerin yanlış yapıldığı ve düzeltilmesi gerektiği düşünülür. Olumsuz geri bildirim oranı %85'ten az ise işlerin düzgün yürütüldüğü söylenebilir. Belirli konularda hiç geri bildirim yapılmaması veya çok az yapılması da farklı anlamlar içerebilir. Örneğin fiyat konusunda hiç geri bildirim olmaması, fiyatların düşük tutulduğu şeklinde yorumlanabilir (Inmon, 2016, s. 122).

2.4.2. Veri Madenciliği

Günümüzde hızla artan veri miktarı, verinin analiz edilerek veriden anlamlı sonuçlar çıkarılması, veriden bilgi elde edilmesi, yönündeki çalışmaları artırmıştır (Yıldız ve Şeker, 2016, s. 10). Her disiplinde üretilen yüksek miktardaki verinin taşıdığı bilgiyi otomatik olarak ortaya çıkarmak için çok yönlü araçlara ihtiyaç duyulmaktadır ve bu amaçla veri madenciliği terimi ortaya çıkmıştır (Han ve diğerleri, 2012, s. 2). Veri madenciliği teriminin büyük miktarda hammaddeden küçük bir külçe bulma sürecindeki madencilik teriminin anlamını yansıtması nedeniyle kullanıldığı, bu nedenle popüler olarak kullanılan bir terim haline geldiği düşünülmektedir. “Bilgi çıkarma”, “veri analizi”, “veri arkeolojisi”, “veri tarama” terimleri de “veri madenciliği” terimine benzer anlamlar içermektedir. “Veri madenciliği, büyük miktarda veriden ilginç örüntüleri ve bilgileri keşfetme sürecidir. Veri kaynakları, veri tabanlarını, veri ambarlarını, Web'i, diğer bilgi havuzlarını veya sisteme dinamik olarak aktarılan verileri içerebilir.”. Bilgi keşif süreci farklı aşamalardan oluşmaktadır. Öncelikle veriyi veri madenciliğine hazırlamak için, veri ön işleme aşamaları uygulanır. Bu aşamalar; tutarsız verilerin giderildiği veri temizleme aşaması, birden fazla veri kaynağının kullanıldığı durumlar için veri entegrasyonu aşaması, analizde kullanılacak verinin seçimi aşaması ve verinin madenciliğe uygun formata dönüştürüldüğü veri dönüştürme aşamasıdır. Ön işlemeden sonra veri, veri madenciliği uygulamalarında kullanılmaya hazır hale gelmiştir. Bu aşamadan sonra, veri örüntüleri çıkarmak için akıllı yöntemlerin uygulandığı veri madenciliği aşaması, ilginç örüntüleri temsil eden bilginin ortaya çıkarıldığı örüntü değerlendirmesi aşaması, bilgiyi kullanıcıya sunmak için görselleştirme ve bilgi sunum tekniklerinin kullanıldığı bilgi sunumu aşaması süreçleri uygulanır (Han, Kamber ve Pei, 2012, s. 5-8).

Veriden bilgi elde edilmesi amacıyla veri madenciliği araçları, özellikle açık kaynak kodlu veri madenciliği araçları, yaygın olarak kullanılmaktadır. Rapid Miner, Knime, WEKA, MOA, Orange ve KEEL açık kaynak kodlu veri madenciliği araçlarından bir kısmıdır (Yıldız ve Şeker, 2016, s. 10).

2.4.3. Yapay Zekâ ve Makine Öğrenmesi

Yapay zekâ, insanın akıl ve anlayışıyla ilgili olan yetenek ve davranışlarının bilgisayar ortamında taklit edilerek modellenmesidir. İnsan beyninin yorumlama ve karar verme kabiliyetleri ile birlikte bilgisayarların sayısal işlemleri çok hızlı olarak tamamlayabilme yeteneklerinin bir arada kullanıldığı modeller geliştirilmesi amaçlanmaktadır. Böylece makinelerin insanlar gibi deneyimlerinden öğrenebilmesi, olayları anlamlandırabilmesi ve problemlere çözüm üretmesi sağlanır (Yılmaz, 2022, s. 4-5). Doğal zekâ yani insanların sahip olduğu zekâ ile öğrenilen bilginin zamanla unutulması mümkündür, ancak yapay zekâ için bilgi kalıcıdır. Yine doğal zekânın sahip olduğu bilginin aktarılması zordur ve uzun zaman isteyebilir, yapay zekâda ise makineler arası bilgi aktarımı kolaydır. Raporlama ve belgeleme süreçleri de yapay zekâda daha kolaydır (Yılmaz, 2022, s. 7). Yapay zekâ, makinelerin büyük veri kümelerinden öğrenmesini ve karmaşık sorunları çözmesini sağlar. Yapay zekânın kalıcılığı ve bilgi aktarımındaki kolaylıkları, onu birçok endüstride değerli bir araç haline getirir. Bu avantajlar, yapay zekânın uygulama alanlarını genişletir ve daha etkili çözümler sunmasına olanak tanır.

Yapay zekâ günümüzde birçok farklı alanda kullanılmaktadır. Sağlık alanında hastalıkların teşhis edilmesi, sanayi alanında üretim ve kontrol süreçlerinde, askeri alanda hedef tespit vb. süreçlerde, bilgisayar oyunlarında, yapay zekâ uygulamaları kullanılmaktadır (Yılmaz, 2022, s. 17). Yapay zekâ sağlık alanında hastalıkların teşhisinde ve tedavi planlarının oluşturulmasında önemli bir rol oynayabilir. Sanayi sektöründe, üretim ve kontrol süreçlerini optimize ederek verimliliği artırabilir. Askeri alanda, hedef tespiti ve stratejik planlamalar için kullanılabilirken, bilgisayar oyunlarında ise daha gerçekçi ve dinamik oyun deneyimleri sunar.

İnsanlar daha önce karşılaşmadıkları bir durumla karşılaştıklarında tecrübe kazanmış olurlar. Makine öğrenmesi ise bilgisayarların deneyim yolu ile öğrendikleri bilgiyi kullanarak daha önce tecrübe etmedikleri benzer problemler ile

karşılaştıklarında çözüm üretebilmesi olarak tanımlanabilir. Bu amaçla öğrenebilen ve bu bilgiyi kullanarak tahminde bulunabilen algoritmalar geliştirilmesi hedeflenir (Yılmaz, 2022, s. 44). Bir makine öğrenmesi algoritması, önceki verilere dayanarak yeni verilere dayalı tahminlerde bulunabilir ve zamanla daha doğru sonuçlar verebilir.

Makine öğrenmesi veri madenciliği ile benzer özellikler taşımaktadır. Ancak veri madenciliği ile geçmişteki özelliklerin ortaya çıkarılması amaçlanırken makine öğrenmesi ile öğrenmesi tamamlanan verilerle ilgili tahmin yürütülmesi amaçlanır. Makine öğrenmesi için verinin ön işlenmesi aşamasında veri madenciliğinden yararlanılabilirken, veri madenciliği uygulamalarında da makine öğrenmesi kullanılmaktadır (Yılmaz, 2022, s. 46). Bu iki alan arasındaki etkileşim, veriden daha fazla bilgi ve anlam çıkarılmasına olanak sağlar.

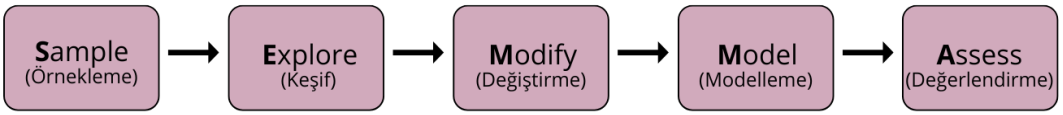
2.4.4. Veri Bilimi Yöntemleri

Veri bilimi projelerinin nasıl yönetileceğine yönelik farklı yöntemler bulunmaktadır. Bu yöntemler projeye nereden başlanacağı, projede hangi aşamaların yer alacağı ve projenin nasıl sonuçlanacağı sorularına yanıt bulmaya çalışmaktadır. Bu amaca yönelik en fazla kullanılan yöntemler SEMMA, CRISP-DM, KDD, TDSP ve OSEMN'dir.

2.4.4.1. SEMMA (Sample, Explore, Modify, Model, Assess)

SEMMA, SAS firması tarafından geliştirilen ancak günümüzde popülerliğini yitirmiş olan bir yöntemdir. SEMMA, sample (örnekleme), explore (keşif), modify (değiştirme), model (modelleme) ve assess (değerlendirme) sözcüklerinin ilk harflerinin birleşiminden oluşan bir sözcüktür. Bu yöntem verinin örneklenmesi, keşfedilmesi, dönüştürülmesi, üzerinde model oluşturulması ve değerlendirilmesi süreçlerinden oluşmaktadır. Değerlendirme süreci sonrasında ulaşılan sonuçtan

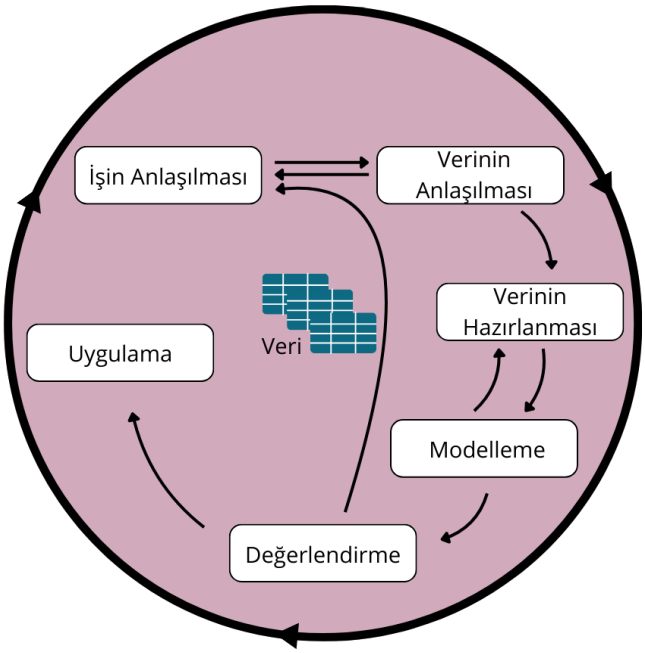
memnun kalınması durumunda süreç tamamlanmış olur ancak istenen sonuca ulaşılamaması durumunda örnekleme aşamasına geri dönülüp süreci tekrar uygulamak gerekir. Şekil 4'te SEMMA yöntemini gösteren şema verilmiştir (Şeker ve Erdoğan, t.y., s. 56).



Şekil 4. SEMMA yöntemini gösteren şema (Şeker ve Erdoğan, t.y., s. 56)

2.4.4.2. CRISP-DM (Cross-Industry Standard Process for Data Mining)

CRISP-DM (Cross-Industry Standard Process for Data Mining) veri biliminde en yaygın kullanılan modellerden biridir. Bir projeye başlanırken öncelikle problem anlaşılır, sonrasında ise veri tanınır. Bu süreç problemle ilgili veriyi elde etmek veya eldeki verinin anlaşılması olarak da ifade edilebilir. Problem tanımlama ve veri tanıma aşamalarından sonra verinin ön işlenmesini de içeren veri hazırlama aşaması başlar. Bu aşamada kirli veya gürültülü veri varsa temizlenir. Sonrasında istatistiksel bir model veya makine öğrenmesi algoritması kullanılarak model oluşturulur. Veriye uygulanan model sonucunda elde edilen sonuç değerlendirilir ve uygun olduğu düşünülürse sonuçların kullanılması aşamasına, bir ürün oluşturulacaksa üretim aşamasına, geçilir. Kullanılan modelle istenilen sonuç elde edilemezse, örneğin testler büyük hatalar veriyorsa, sürece problemi anlama aşamasından tekrar başlanır (Şeker ve Erdoğan, t.y., s. 59-60). Şekil 5'te süreci anlatan şema verilmiştir (Şeker, 2018).

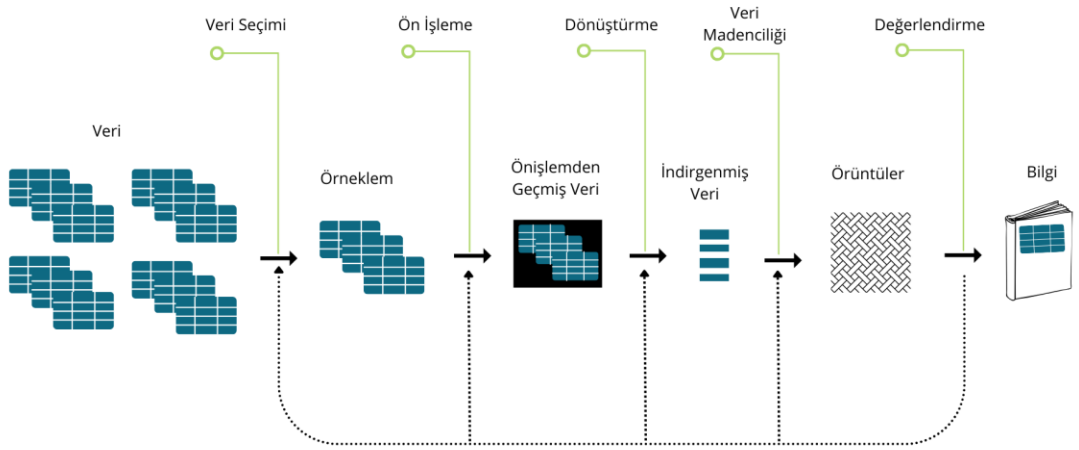


Şekil 5. CRISP-DM yöntemini gösteren şema (Şeker, 2018).

2.4.4.3. KDD (Knowledge and Data Discovery)

KDD, kısaca bilgi ve veri keşfi olarak tanımlanabilir. Probleme uygun veri seçimi, ön işleme ile eksik ve kirli verilerin temizlenerek ayrıca gerekirse boyut düşürme gibi işlemler uygulayarak verinin kullanılabilir hale getirilmesi, verinin dönüştürülmesi (metnin sayısallaştırılması, ortala alınması vb.), veri madenciliği ile örüntülerin ortaya çıkarılması ve değerlendirme süreçlerini kapsamaktadır. Veriye bu işlemler uygulanarak bilginin ortaya çıkarılması sağlanır. SEMMA ve CRISP-DM yöntemleri döngüsel yapıdayken, KDD yöntemi doğrusal yapıdadır (Şeker ve Erdoğan, t.y., s. 61). Şekil 6'da süreci anlatan şema verilmiştir.

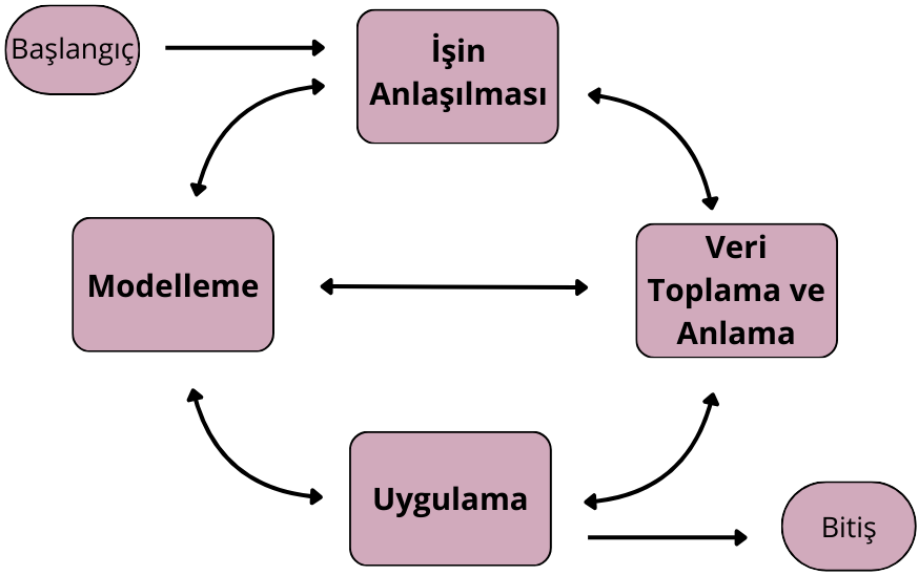
KDD, genellikle yeni karşılaşılan bir problemle ilgili olarak ilk kez yapılacak keşiflerde, araştırma amaçlı işlerde kullanılmaktadır. KDD yönteminde veri ön işleme ve dönüştürme süreçleri SEMMA ve CRISP-DM yöntemlerine göre daha fazla detaylandırılmıştır. Endüstriye, piyasaya ve ürüne yönelik işlerde CRISP-DM yöntemi daha fazla kullanılmaktadır (Şeker ve Erdoğan, t.y., s. 62).



Şekil 6. KDD yöntemini gösteren şema (Şeker ve Erdoğan, t.y., s. 61)

2.4.4.4. TDSP (The Team Data Science Process)

TDSP (Team Data Science Process), Microsoft tarafından geliştirilen ve veri bilimi projelerini yapılandırmak için kullanılan bir metodolojidir. Çevik (agile) prensipleri veri bilimi projelerine uygulamayı amaçlar. TDSP, veri bilimi projelerinin yaşam döngüsünü İşin Anlaşılması, Veri Edinme ve Anlama, Modelleme, Dağıtım ve Müşteri Kabulü olmak üzere beş ana aşamada ele alır. Bu yaşam döngüsü, bir projeyi başarıyla tamamlamak için gereken adımları ana hatlarıyla belirtir. TDSP, standart bir proje yapısı, altyapı ve kaynak önerileri ve proje yürütme için gerekli olan araçlar ve yardımcı programlar gibi bileşenleri içerir. Şekil 7'de TDSP sürecini gösteren şema verilmiştir ("Learn Azure", 2024; Hotz, 2024a).



Şekil 7. TDSP yöntemini gösteren şema (“Learn Azure”, 2024; Hotz, 2024a)

2.4.4.5. OSEMN (Obtain, Scrub, Explore, Model, iNterpret)

OSEMN yaklaşımı veri elde etme, temizleme, keşfetme, modelleme ve yorumlama adımlarından oluşur. OSEMN, bir veri projesi yaşam döngüsünü tanımlayan yapılandırılmış bir yaklaşımdır ve her aşamayı sistematik bir şekilde ele alır (Hotz, 2024b). Veri elde etme, ilgili kaynakların belirlenmesi ve verilerin toplanması; temizleme, eksik ve/veya hatalı verilerin temizlenerek veri setinin analize hazır hale getirilmesi; keşif, istatistiksel analiz ve veri görselleştirme ile veri içeriğinin anlaşılması; modelleme, uygun algoritmalarla tahmin veya sınıflandırma modellerinin oluşturulması; yorumlama ise sonuçların değerlendirilmesi ve karar aşamasında kullanılmasını ifade eder. Her adımın tanımlı olması, sürecin daha etkili bir şekilde yönetilmesini sağlar. Şekil 8’de OSEMN yöntemini gösteren şema verilmiştir (Sharma, 2024; Hotz, 2024b).

Şekil 8. OSEMN yöntemini gösteren şema (Sharma, 2024; Hotz, 2024b)

2.4.5. Veri Görselleştirme

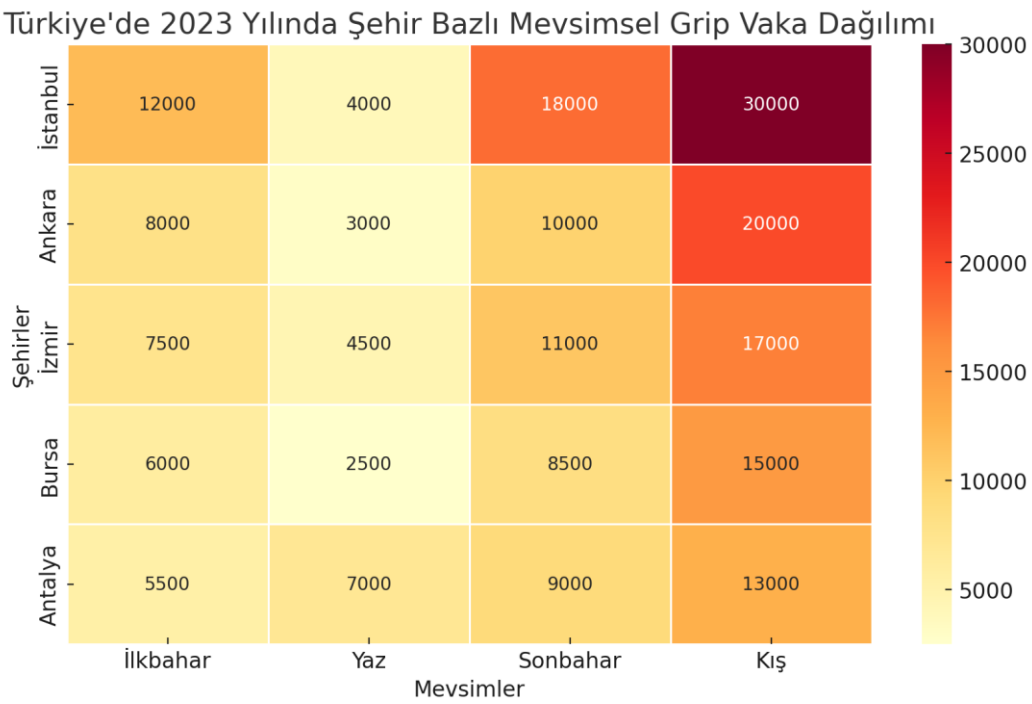
Veri görselleştirme, büyük veri setlerinden anlam çıkarma sürecinde, veriyi hızlı ve etkin bir şekilde yorumlamayı sağlayan önemli bir araç olarak öne çıkmaktadır. Veri görselleştirme teknikleri, karar alıcılara görsel içerikler sunarak doğru bilgi edinilmesini kolaylaştırmaktadır (Park, Kim ve Lee, 2021).

Veri görselleştirme araçları (ör. Tableau, Power BI), verilerin grafiksel olarak sunumunu basitleştirerek karar alıcıların bilgiye hızlı erişimini sağlar. Özellikle büyük veri setleriyle çalışırken, grafiksel analizler kullanıcıların verileri daha hızlı anlamalarını destekler (Al-Kassab, Ouertani, Schiuma ve Neely, 2014). Bu süreçte, görselleştirilmiş veriler karmaşık analizleri sadeleştirerek veri tabanları üzerindeki yükü hafifletir ve verileri yorumlamayı kolaylaştırır (Battle ve Scheidegger, 2020).

Veri görselleştirmenin kamuda kullanımının başarılı bir örneği, Birleşik Krallık'taki Açık Veri Enstitüsü'nün (Open Data Institute) çalışmalarıdır. Bu kuruluş, farklı veri setlerini ulusal düzeyde birleştirerek politika yapıcıların veriye dayalı karar almasını desteklemektedir. Böylece görselleştirilmiş verilerden sağlanan bilgi akışı, karar süreçlerinde doğruluğu ve hızı artırmaktadır (Alhadad, 2018).

Sağlık alanında, veri görselleştirme araçları halk sağlığı kararlarını desteklemekte, görselleştirilen veriler, doğru bilgiye erişimi ve yorumlamayı kolaylaştırarak karar süreçlerinde önemli bir rol oynamaktadır (Park ve diğerleri, 2021). Bu duruma kurgusal bir örnek olarak, Türkiye'de 2023 yılı boyunca çeşitli

şehirlerde gözlemlenen grip vaka dağılımı gösteren şekil oluşturulmuştur. Sağlık Bakanlığı İstanbul, Ankara, İzmir, Bursa ve Antalya gibi büyük şehirlerdeki grip vakalarını mevsimlere göre analiz etmek istemektedir. Bu analiz, hangi mevsimlerde hangi şehirlerin daha yüksek risk taşıdığını belirlemeyi kolaylaştırmaktadır. Şekil 9'da görüldüğü üzere, yılın dört mevsimi için beş şehirdeki grip vakalarının dağılımı bir ısı haritası ile görselleştirilmiştir. Bu haritada renk yoğunluğu, şehirlerdeki mevsimsel vaka yoğunluğunu yansıtmaktadır. Haritada koyu renkler yüksek vaka sayılarını, açık renkler ise düşük vaka sayılarını ifade etmektedir. Isı haritası kullanılarak oluşturulan bu görselleştirme, kış aylarında İstanbul ve Ankara'da grip vakalarının çok daha yoğun olduğunu, buna karşın yaz aylarında genel olarak tüm şehirlerde vaka sayılarının düştüğünü açıkça göstermektedir. Bu görsel analiz, karar alıcıların riskli dönemleri ve bölgeleri hızlıca fark etmesine yardımcı olur. Sağlık Bakanlığı bu bilgiyi kullanarak önleyici tedbirler planlayabilir.



Şekil 9. Kurgusal bir veri görselleştirme örneği. Türkiye'de 2023 yılında şehir bazlı mevsimsel grip vaka dağılımı ısı haritası

Görselleştirmenin etkili olması için grafiklerin karmaşıklığı ve kullanıcı deneyimi gibi insan faktörleri dikkate alınmalıdır. Örneğin, çok boyutlu veri görsellerinde ısı haritaları ve radar grafikleri gibi araçlar veriyi hızlı anlamayı sağlarken aşırı karmaşık grafikler karar süreçlerini zorlaştırabilir (Pattanaik ve Wiegand, 2021). Bu nedenle, veri görselleştirmenin kullanım amacına uygun olarak tasarlanması ve kullanıcı dostu olması önemlidir.

2.5. VERİ GÜVENLİĞİ

Veri güvenliği her zaman önemli bir kavram olmakla birlikte, internet kavrama yeni boyutlar kazandırmıştır. İnternet öncesi dönemde veri güvenliği ile ilgili olarak verinin kaybolmasının engellenmesi ön plandayken, internetle birlikte verinin yetkisiz kişilerin eline geçmesinin, değiştirilmesinin, kopyalanmasının engellenmesi vb. konuların önemi artmıştır. İnternet ortamında kimi kullanıcılar güçlerini göstermek amacıyla şifre çözme vb. işlemlerle korunmaya çalışılan veriye ulaşsa da asıl tehlike başkalarına ticari, maddi veya manevi zarar vermek amacıyla gizli verilere ulaşılmasıdır (Al, 2002, s. 38). İnternetin yaygınlaşmasıyla veri güvenliği, verinin kaybolmasını önlemenin ötesine geçerek yetkisiz erişim, veri değişikliği, kopyalama ve veri gizliliğini koruma gibi konuları içermeye başlamıştır. Bu bağlamda, siber saldırılara karşı korunma, veri sızıntılarını önleme ve etkili güvenlik protokollerinin kullanımı kritik öneme sahiptir.

Güvenlik teknolojilerinin tüm dünyada olduğu gibi ülkemizde de önemi anlaşılmıştır. Ağ uygulamalarına yönelik güvenliğin sağlanması, ulusal güvenliğin önemli bir parçasıdır. Bu nedenle bu teknolojilerde kendine yeterlik seviyesine ulaşmak önemlidir. (*T.C. Ulaştırma Bakanlığı TUENA,* 1999, s. 35). Veri güvenliği konusunda sürekli olarak yeni teknolojiler geliştirilmektedir (Al, 2002, s. 38). Güvenlik teknolojilerinde kendine yeterlik sağlamak ve sürekli olarak yeni teknolojiler geliştirmek, yeni güvenlik tehditlerine karşı etkili savunma mekanizmaları oluşturulmasına yardımcı olur.

Ülkemizde Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi tarafından Temmuz 2020'de "Bilgi ve İletişim Güvenliği Rehberi" yayımlanmıştır (Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2020). Rehberin temel amacı "bilgi güvenliği risklerinin azaltılması, ortadan kaldırılması ve özellikle gizliliği, bütünlüğü veya erişilebilirliği bozulduğunda milli güvenliği tehdit edebilecek veya kamu düzeninin bozulmasına yol açabilecek kritik bilgi/verinin güvenliğinin sağlanması için asgari güvenlik tedbirlerinin belirlenmesi ve belirlenen tedbirlerin uygulanması için yürütülecek faaliyetlerin tanımlanması" olarak belirtilmiştir. Bu rehberin uygulanmasıyla, yerli ve milli ürün kullanımının yaygınlaşması, konu hakkında farklı kurum ve kuruluşların mükerrer çalışma yapmasının engellenmesi, varlık gruplarının güvenlik tedbirlerinin derecelendirilerek, dereceye uygun tedbirlerin uygulanması, tedbirlerin ürün ve teknolojiye bağlı kalınmadan uygulanabilmesi, tedbirlerin uygulanmasının denetlenebilmesi, bütün kurum ve kuruluşlar tarafından tedbirlerin teknik olarak uygulanabilir olması, rehberin yeni koşullara uygun olarak sürdürülebilirliğinin sağlanması, rehberin diğer mevzuat ve rehberlerle ayrıca ulusal ve uluslararası standartlarla uyumlu olması hedeflenmektedir (*Türkiye Cumhuriyeti Cumhurbaşkanlığı,* 2020, s. 11). Rehberde ayrıca "Uygulama ve Veri Güvenliği" başlığı altında veri güvenliği konusu daha detaylı olarak incelenmiştir. Bu başlık altında "Kimlik Doğrulama, Oturum Yönetimi, Yetkilendirme, Dosyaların ve Kaynakların Güvenliği, Güvenli Kurulum ve Yapılandırma, Güvenli Yazılım Geliştirme, Veri Tabanı ve Kayıt Yönetimi, Hata Ele Alma ve Kayıt Yönetimi, İletişim Güvenliği, Kötücül İşlemleri Engelleme ve Dış Sistem Entegrasyonlarının Güvenliği" konularına yönelik tedbirler tanım ve seviyesi ile birlikte listelenmiş ve her tedbirle ilgili olarak denetim yöntem ve soru önerileri listelenmiştir (*Türkiye Cumhuriyeti Cumhurbaşkanlığı,* 2020, s. 84-113). Bu kapsamlı rehber, Türkiye'deki bilgi ve iletişim güvenliği uygulamalarını iyileştirmek ve standartlaştırmak amacıyla önemli bir adım olarak değerlendirilmektedir.

Bilgi ve İletişim Güvenliği rehberinin sunduğu tedbirlerin yanı sıra, açık verinin güvenli kullanımı da kritik bir öneme sahiptir. Açık verinin anonimleştirilerek kullanılması verideki kişisel ve hassas verinin açığa çıkmasını önleme

bakımından kritik öneme sahiptir. Verinin biçimi korunarak kişisel ve hassas verilere erişim engellenir. Bu konuda “Federe Öğrenme” ve “Diferansiyel Mahremiyet” gibi farklı mahremiyet modelleri geliştirilmektedir (“Türkiye Cumhuriyeti”, t.y.a).

2.5.1. Federe (Dağıtık) Öğrenme

Klasik makine öğrenmesinde bir model ve veri vardır. Model verilerle eğitilerek faydalı bir görev yapılması sağlanır. Modeli eğitmek için kullanılan veriler çoğunlukla modelin eğitildiği makinede üretilmez. Ayrıca bu veriler birden fazla cihazda bulunabilir; aynı uygulamayı çalıştıran farklı cihazlar veri üretebilir. Klasik makine öğrenmesinde tüm bu veriler merkezi bir sunucuda toplandıktan sonra modeli eğitmek için kullanılmaktadır. Örneğin web trafiğinin analiz edilmesi gibi örneklerde merkezi sunucu kullanılabilir.

Ancak, verilerin merkezi bir sunucuda bulunmadığı veya tek bir sunucudaki verilerin yeterli olmadığı durumlarda klasik makine öğrenmesi yöntemi çalışmaz. KVKK gibi veri koruma düzenlemeleri nedeniyle hassas verilerin taşınamaması, kullanıcıların verilerinin taşınmasını istememesi veya veri hacminin yüksek olması nedeniyle veri taşımanın ekonomik olmayacağı durumlar sebebiyle klasik makine öğrenmesinin kullanılması mümkün olmayabilir. Federe öğrenmede ise veriyi eğitim modeline taşımak yerine model verinin olduğu yere taşınır. Böylece örneğin tıbbi yapay zekâ modellerini farklı hastanelerin verileriyle eğitmek mümkün olmaktadır. Bütün bu süreçler merkezi bir sunucu koordinasyonunda yürütülür (“Flower framework”, t.y.b.).

2.5.2. Diferansiyel Mahremiyet

Sağlık hizmetleri, finansal işlemler vb. veri setleri değerli potansiyele sahiptir ancak bu veriler aynı zamanda hassastır ve bireysel gizliliğin korunması gerekir.

Anonimleştirme gibi geleneksel yöntemler bu gizliliği korumakta tek başına yeterli olamamaktadır. Diferansiyel mahremiyet verilerin gizliliğinin korunarak analiz edilmesine olanak tanır. Temelde veri grubunun ana kalıbını korurken bireysel verilerin gizli kalmasını sağlar. Örneğin A grubundaki kişiler ile B kişisinin verilerinin analiz edilmesiyle elde edilen sonucun sadece A grubundaki kişilerin verilerinin analiz edilmesiyle elde edilecek sonuçla neredeyse aynı olması sağlanır. Böylece B kişine ait verilerin kalabalığın içinde gizli kalması sağlanmış olur. Bunu sağlamak için kullanılan en yaygın yöntemlerden biri analizin çıktısına verideki her bireyin etkisini maskeleyecek kadar gürültü eklemektir. Böylece herhangi bir kişinin veri kümesinde olup olmadığının anlaşılmasını zorlaşır. Bu süreçte genel doğruluğun korunması önemlidir (“Flower framework”, t.y.a.).

Diferansiyel mahremiyet, özellikle hassas veri içeren alanlarda giderek daha yaygın bir kullanım alanı bulmaktadır. Bununla birlikte, küçük veri setlerinde doğruluk kaybı, diferansiyel mahremiyetin uygulanmasında önemli bir zorluk olarak öne çıkmaktadır. Bu tür kayıpların azaltılması için literatürde yeni yöntemlerin geliştirilmesi gerektiği sıklıkla vurgulanmaktadır (Nelson ve Reuben, 2019). Aynı zamanda, diferansiyel mahremiyet, merkezi derin öğrenme modelleri gibi yapay zekâ uygulamalarında da yaygınlaşmaktadır. Bu yöntem, gizlilik ihlallerine karşı etkin bir koruma sağlarken, verilerin analitik süreçlere entegrasyonunu mümkün kılmaktadır. Ancak gizlilik ile doğruluk arasında denge kurulması bu süreçte kritik bir sorun olarak varlığını sürdürmektedir. Yapılan araştırmalar, mahremiyeti ihlal etmeden daha yüksek doğruluk oranlarına ulaşılmasını sağlayacak tekniklerin geliştirilmesi gerektiğine işaret etmektedir (Demelius, Kern ve Trügler, 2023). Diferansiyel mahremiyet aynı zamanda yasal ve etik gerekliliklere uyum açısından da önemli bir araç olarak değerlendirilmektedir. Özellikle GDPR gibi düzenlemelere uyum sağlamak için bu yöntemlerin etkin bir şekilde uygulanması, veri koruma süreçlerini güçlendirebilir (Voigt ve Von dem Bussche, 2017; Prokhorenkov, 2022).

2.6. VERİ DEPOLAMA VE YÖNETİMİ

Verilerin türü, yapısı, kullanım alanı ve işlevselliği doğrultusunda değişen gereksinimlere göre organize edilen ve yönetilen çeşitli veri saklama ve yönetim yapıları bulunmaktadır. Veri gölleri, veri ambarları, veri havuzları ve veri tabanları bu yapılar arasında yaygın olarak kullanılanlardır. Bu alanlar yalnızca verilerin depolanmasını sağlamakla kalmaz, aynı zamanda veri yönetimine yönelik farklı özellikler ve işlevler de sunar. Aşağıda, bu yapıların temel tanımlarına yer verilecektir.

2.6.1. Veri tabanı (Database)

Veri tabanı, “bir ya da birkaç uygulamada kullanılmak için, gereksiz yinelemelerden arınmış olarak, düzenli biçimde bilgisayar belleklerinde saklanan birbiriyle ilişkili veriler topluluğu” olarak tanımlanmakla birlikte (Date, 2004) bu tanımın günümüzde eksik kaldığı, bu kapsama giren veri kümelerinin tamamının veri tabanı olarak kabul edilemeyeceği belirtilmektedir (Connolly ve Begg, 2005). Bu tanıma ek olarak veri tabanlarının belirli bir kurumun uygulamalarında kullandığı birbiriyle ilişkili işletimsel verilerden oluşması, birçok uygulamada ortak kullanılması, geçici verilerin veri tabanında yer almaması gerekir (Elmasri ve Navathe, 2016). Veri tabanlarındaki veriler üzerinde değişiklik yapılabilir, yine veri raporlama ve sorgulama işlemleri de yapılabilir. Veri tabanlarındaki verileri her kullanıcının işletim sistemi komutlarıyla veya genel programlama dillerini kullanarak değiştirmesi mümkün değildir. Veri tabanlarındaki verilerde değişiklik Veri Tabanı Yönetim Sistemi (VTYS) aracılığıyla yapılabilir. VTYS ile kullanıcılara yalın mantıksal yapılar sunulur ve veri tabanlarının fiziksel yapısı kullanıcılardan gizlenir (Connolly ve Begg, 2005). VTYS ile kullanıcıların veri tabanı algoritmaları ile uğraşmadan isteklerini ortaya koyması mümkün olur (Yarımağan, 2016, s. 1-2).

Veri tabanı kısaca bir kuruluşun bilgi ihtiyaçlarını karşılamak amacıyla tasarlanmış, mantıksal olarak ilişkili verilerin ve bu verilere ait açıklamaların paylaşıldığı koleksiyon olarak tanımlanabilir. Verilere ait açıklamalar sistem kataloğu, veri sözlüğü veya üst veri olarak bilinmektedir. Veri tabanlarının kendi kendilerini tanımlayan yapısı onları bağımsız yapmaktadır. Genellikle veri tabanları kurumların farklı departmanları tarafından aynı anda kullanılan büyük veri depolarıdır (Connolly ve Begg, 2015, s. 63).

Veri tabanları genellikle, veri işlemeyi verimli hale getirmek için, verileri tablolarda satır ve sütunlarda tutarlar. Böylelikle veri erişimi, güncelleme, değiştirme, organize etme gibi işlemler daha kolay yürütülür. Veri tabanlarında genellikle SQL (Structured Query Language - Yapılandırılmış Sorgu Dili) kullanılır. 1960'lı yıllarda ortaya çıkan veri tabanları daha esnek bir yapıdayken 1980'li yıllardan itibaren ilişkisel veri tabanları ve 1990'lı yıllardan itibaren nesne odaklı veri tabanları kullanılmaya başlanmıştır. İnternette yaşanan gelişmeler ise verinin daha hızlı işlenmesi ihtiyacını doğurmuş ve NoSQL (not only SQL) veri tabanları yaygınlaşmıştır ("Oracle Türkiye", t.y.). NoSQL veri tabanları verilerin yapılandırılmış, yapılandırılmamış veya yarı yapılandırılmış olarak saklanmasına olanak sağlamaktadır. Çünkü verileri saklarken ilişkisel olarak değil, anahtar-değer çifti olarak saklamaktadır (Kayabay ve diğerleri, 2016, s. 6). Günümüz ihtiyaçlarını karşılamak üzere bulut destekli veri tabanları kullanımında önemli bir artış gözlemlenmiştir ("Oracle Türkiye", t.y.).

2.6.2. Veri Ambarı (Data Warehouse)

Veri ambarı, kısaca konu odaklı, entegre, zamana bağlı olarak düzenlenmiş ve değişmez bir veri koleksiyonu olarak tanımlanabilir (Inmon, 2005, s.29). Veri ambarları, verilerin düzenlenmiş olarak saklandığı dosya veya klasörlerden oluşan, veri tabanı türünde bir veri yönetim yapısıdır ("Oracle Türkiye", t.y.). Bu yapılar, büyük ölçekli kurumsal verilerin yönetilmesi ve analiz edilmesi için kritik bir rol oynar. Örneğin, veri ambarları sayesinde veriler düzenlenmiş bir yapıda

depolanır ve bu durum veri kullanımını kolaylaştırır. Veri ambarları, çoğunlukla analiz ve raporlama işlemleri için geçmiş verilere dayanarak çalışır ve iş zekâsı uygulamalarında yaygın olarak tercih edilir.

Veri ambarlarında ETL (Extract Transform Load) işlemi kullanılır. Çeşitli kaynaklardan gelen veri temizlenip dönüştürülerek depolanır. Veri, veri ambarı için tanımlanan şemalara uygun olarak düzenlenir. Bu nedenle veri ambarının geliştirilmesi aşamasında farklı kaynaklardan gelen verinin incelenerek nasıl dönüştürüleceğinin belirlenmesi gerekir. Verilerin belirli bir şemaya uygun biçimde yapılandırılması, veriyi anlamayı ve operasyonel süreçlerde kullanmayı kolaylaştırır (“Gtech”, t.y.). Ancak bu aşamalarla verilerin uyumlu hâle getirilmesi zaman alıcı ve maliyetli bir süreçtir. Bir veri ambarında yapılandırılmamış ya da yarı yapılandırılmış verileri depolamak zordur. Şema yapılarının kısıtlı ve durağan bir yapıda olması veri ambarlarının esnekliğini sınırlayarak, verilerin daha esnek yapılarda depolanmasını gerektiren ihtiyaçlara yönelik veri gölü (data lake) olarak bilinen yeni nesil veri yönetim çözümlerine olan ilgiyi artırmıştır (Inmon, 2005).

Veri ambarları ve operasyonel veri tabanları genellikle benzer olarak görülse de bu iki sistemin kullanım amaçları ve yapıları arasında belirgin farklar vardır. Operasyonel veri tabanları (OLTP sistemleri olarak da bilinmektedir), şirketlerin günlük işleyişine destek olmak amacıyla geliştirilmiştir. Bu veri tabanları sürekli güncellenen, anlık işlemleri destekleyen ve genellikle işlem bazında veri giriş-çıkışı için optimize edilen yapılardır. Örneğin, bir bankacılık sistemindeki müşteri işlemleri ya da e-ticaret platformundaki sipariş süreçleri operasyonel veri tabanları üzerinde yürütülür.

Öte yandan veri ambarları veriler üzerinde sistemik analiz yapmayı da amaçlayan yapılardır. Analitik ve raporlama odaklı veri analizleri için optimize edilmiş olup, genellikle çok sayıda kaynaktan gelen verileri tek bir merkezde entegre ederek tutarlı bir veri seti sunar (Kimball ve Ross, 2013). Bu özellikler, veri ambarlarını stratejik karar alma süreçlerinde önemli bir veri kaynağı hâline getirir. Ayrıca veri ambarları operasyonel veri tabanlarının aksine sürekli güncellenmek yerine belirli

zaman aralıklarında toplu veri yüklemeleri yapılarak güncellenir (Watson ve Wixom, 2007).

Bu noktada, veri ambarı mimarisinin ölçeklendirilmiş ve konu-odaklı bir sürümü olarak “veri mart” (data mart) kavramından da bahsetmek gerekir. Veri mart (bazı kaynaklarda “veri pazarı” olarak Türkçeleştirilir), belirli bir departman veya iş biriminin odaklandığı veri analizi gereksinimlerini karşılayacak şekilde daraltılmış ve özelleştirilmiş bir veri ambarı yapısıdır (Kimball ve Ross, 2013). Örneğin, pazarlama departmanı için hazırlanan bir veri mart, yalnızca satış, müşteri segmentasyonu ve kampanya verilerini içererek ek bir filtreleme katmanı sunar. Veri ambarından beslenen veri martlar, departman çalışanlarına daha hızlı erişim, daha az karmaşa ve konuya özel analitik imkânlar sağlar. Böylece, kurum genelindeki büyük ve kapsamlı veri ambarı yapısı, farklı iş birimlerinin ihtiyaçlarına göre ölçeklendirilebilir ve özelleştirilebilir hale gelir.

Bununla birlikte, veri ambarlarının temel ilkelerini koruyan fakat daha daraltılmış ve konu odaklı bir mimari sunan “veri marketi” (data mart), belirli bir departman veya iş biriminin ihtiyaç duyduğu veri setlerini barındıracak şekilde ölçeklendirilmiş bir veri deposu olarak tanımlanır (Kimball ve Ross, 2013). Örneğin, pazarlama, finans ya da insan kaynakları gibi iş birimlerinin özel gereksinimlerini karşılamak için büyük kurumsal veri ambarından beslenen bu yaklaşım, kullanıcıların kendi alanlarına yönelik özelleşmiş veriler üzerinden daha hızlı ve verimli sorgulamalar yapmasına imkân tanır. Aynı zamanda, sınırlı boyutta veriye sahip küçük veya orta ölçekli kuruluşlar da ihtiyaç duydukları odaklı analizlere hızlıca ulaşmak için veri marketi yapısından yararlanabilir. Bu durum hem raporlama süreçlerinde hem de analitik uygulamalarda performans artışı ve kullanım kolaylığı sağlar.

2.6.3. Veri Gölü (Data Lake)

Veri gölleri, yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerin tamamını depolayabilen esnek veri yönetim sistemleridir. Veri göllerinde, veriler

genellikle ham halde, yani işlenmeden ve yapılandırılmadan saklanır. Bu özellik, veri göllerinin büyük miktarda veri gereksinimi duyan yeni nesil analiz, yapay zekâ ve makine öğrenimi gibi uygulamalar için ideal bir ortam olmasını sağlar. Farklı kaynaklardan gelen veriler, kaynağına bağlı olmaksızın hızlı ve düşük maliyetle veri gölüne aktarılabilir. Bu sistemde, veriler yalnızca kullanılacağı zaman, belirli bir analiz veya işlem gerektirdiğinde uygun biçime dönüştürülür (Sawadogo ve diğerleri, 2019). Bu durum, veri göllerinin hızlı ve düşük maliyetli veri erişimi gerektiren projelerde kullanılabilirliğini artırır (“Gtech”, t.y.).

Veri gölü kavramının temelinde, veri ambarlarının sınırlı yapısının getirdiği kısıtlamaların aşılması amacı bulunmaktadır. Özellikle geniş veri hacmi ve veri çeşitliliği ile çalışan veri bilimciler için veri gölleri uygun bir ortam sağlar. Bu durum, veri göllerini büyük veri analitiği uygulamalarında sıklıkla tercih edilen bir araç haline getirmiştir (Madera ve Laurent, 2016). Ancak veri gölleri, ham verilerden oluştuğu için diğer kullanıcılar açısından karmaşıktır ve veri işleme süreci zaman alabilir; bu da genel kullanıcılar için zorluk oluşturur.

Veri göllerinde genellikle ELT (Extract, Load, Transform) yöntemi kullanılır. Bu yöntemde, veriler doğrudan veri gölüne yüklenir, sonra dönüştürülür. Böylece depolama aşamasında hız ve esneklik sağlanır. Fakat bu yapı, veriyi analiz etmeye yönelik kullanım aşamasında dönüşüm gerektirdiğinden dolayı performans açısından bazı zorluklar yaratabilir (Hai ve diğerleri, 2016).

Buna ek olarak, veri gölü mimarisi “schema-on-read” (şemayı okurken tanımlama") özelliği sunar, yani veriler depolama sırasında belirli bir şemaya uymak zorunda değildir. Bu özellik, veri göllerinin geniş veri çeşitliliğine sahip ve sık veri değişikliği gerektiren projelerde daha esnek kullanım imkânı sağlanmasına katkıda bulunur (Inmon, 2016).

2.6.4. Veri Havuzu (Data Pool)

Veri havuzu bir proje veya grup için gerekli tüm verilerin depolandığı büyük "işletim alanı"dır. Farklı projeler için farklı veri havuzları oluşturulabilir ("Oracle Help Center", t.y.). Veri havuzu ticaret ortaklarının standart bir formatta veri saklayabildiği ve veriye ulaşabildiği alandır. Kullanıcılar veri havuzlarında verilerini, kolayca senkronize etmelerini sağlayan, standart biçimlerde depolar ("IBM Documentation", 2021).

Veri havuzları, veri gölleri ve veri ambarlarına göre daha küçük ölçekli ve belirli projelere veya gruplara odaklanan bir yapı sunar. Veri havuzları belirli projeler veya gruplar için tasarlanmış, verilerin hızlı ve uyumlu bir şekilde paylaşılmasına odaklanan bir veri yönetim sistemidir. Bu yapı, veri analizinden ziyade verilerin operasyonel süreçlerde paylaşılması için uygundur ve özellikle iş ortaklarıyla veri senkronizasyonu gerektiren durumlarda etkili bir çözüm sunar.

2.6.5. Bulut Tabanlı Depolama ve Dağıtık Sistemler

Dağıtık sistemler ve bulut tabanlı depolama, modern bilgi teknolojilerinin temel unsurları haline gelmiştir. Dağıtık sistemler, birden fazla bağımsız bilgisayarın bir ağ üzerinden iletişim kurarak ortak bir hedef için çalıştığı yapılardır ve yüksek performans, güvenilirlik ve kullanılabilirlik sağlar (Cointelegraph, 2024). Bulut depolama ise, verilerin internet üzerinden erişilebilen uzak sunucularda saklanmasını sağlayan bir hizmettir. Bu teknoloji, kullanıcılara esneklik, ölçeklenebilirlik ve maliyet etkinliği gibi avantajlar sunar (Mikroarea Digital, 2024). Bulut depolama, geleneksel depolama yöntemlerine kıyasla birçok fayda sağlar. Örneğin, kullanıcılar ihtiyaç duydukları kadar depolama alanını kolayca kiralayabilir ve internet üzerinden her yerden verilerine erişebilirler. Ayrıca, verilerin yedeklenmesi ve geri yüklenmesi geleneksel veri merkezlerine göre çok daha kolaydır (Bulutistan, t.y.). Bu teknolojiler, veri yönetimi ve işleme konusunda önemli avantajlar sunarken bazı zorlukları da beraberinde getirmektedir. Bulut

depolamanın internet bağlantısına olan bağımlılığı ve potansiyel güvenlik riskleri dezavantajları arasında sayılabilir (UltaHost, 2024). Bununla birlikte, bulut depolama hizmetleri, güçlü şifreleme teknikleri ve güvenlik protokolleri sayesinde güvenlik risklerini en aza indirmektedir. (Mikroarea Digital, 2024). Gelecekte, uç bilişim (edge computing), 6G/7G gibi teknolojilerin entegrasyonu ile bulut ve dağıtık sistem teknolojilerinin daha güçlü ve güvenli hale geleceği ve mevcut dezavantajların büyük ölçüde ortadan kalkacağı öngörülmektedir (Bulutistan, 2024; Codit Teknoloji, 2024). Bu gelişmeler, veri depolama ve işleme kapasitelerini daha da artırarak işletmelere ve bireylere daha verimli ve güvenli veri yönetimi çözümleri sunacaktır.

Bu noktada, bulut ve dağıtık sistemlerin sağladığı altyapının; veri ambarı, veri gölü ve veri havuzu gibi veri mimarilerine dayalı depolama çözümlerinin dönüşümünde kilit bir rol oynadığını vurgulamak önemlidir. Bulut tabanlı veri ambarları, geleneksel veri ambarlarına kıyasla daha esnek ve ölçeklenebilir bir yapı sunar, böylece kurumların değişen iş ihtiyaçlarına hızla uyum sağlayabilir (Abadi ve diğerleri, 2020). Bulut tabanlı veri gölleri, büyük miktarda yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış veriyi ham haliyle depolayarak, verinin gerektiğinde işlenip analiz edilmesine olanak tanır (Sawadogo ve diğerleri, 2019). Bulut tabanlı veri havuzları, ticaret ortakları veya kurum içindeki farklı ekipler arasında hızlı ve uyumlu veri paylaşımı sağlar. Standart formatlarda veri saklama özelliği, verinin uyumluluğunu korurken, merkezi bir bulut platformu aracılığıyla farklı kullanıcıların veri senkronizasyonu yapabilmesine olanak tanır (Gurajada, Sattler, Tinnefeld ve Höfner, 2017). Bu sistem geleneksel veri paylaşım yöntemlerine göre daha verimli ve güvenli bir yaklaşım sunar.

Bulut tabanlı depolamanın en büyük avantajlarından biri ölçeklenebilirliktir. Veri gölü veya veri havuzu gibi yapılar, bulut platformlarında ihtiyaçlara göre hızla büyütülebilir veya küçültülebilir. Bu esneklik, organizasyonların artan veri hacimlerini maliyet etkin bir şekilde yönetmesini sağlar (Buyya ve diğerleri, 2018). Ayrıca dağıtık sistemlerle verinin farklı coğrafi konumlarda depolanabilmesi veri kaybı riskini en aza indirir.

Büyük bulut sağlayıcıları, bu teknolojileri kullanarak kapsamlı çözümler sunmaktadır. Örneğin, Amazon Web Services (AWS) S3 ve Redshift hizmetleriyle veri gölü ve veri ambarı çözümleri sunarken, Google Cloud Platform (GCP) BigQuery ve Cloud Storage hizmetleriyle benzer işlevleri yerine getirir. Microsoft Azure ise Data Lake Storage ve Synapse Analytics gibi hizmetlerle bu alanda rekabet etmektedir (Hai ve diğerleri, 2016). Bu platformlar, dağıtık altyapılarıyla kullanıcılarına güvenli, esnek ve ölçeklenebilir veri yönetimi çözümleri sağlar.

Özetle, bulut tabanlı depolama ve dağıtık sistemler, veri ambarı, veri gölü ve veri havuzu gibi yapıların esnekliğini artırarak, büyük veri analitiği ve iş zekâsı uygulamalarında verinin kolay ve verimli bir şekilde yönetilmesini sağlar. Bu teknolojiler, anlık veri erişimi, veri yedekleme ve güvenlik gibi özelliklerle kurumların dijital dönüşüm süreçlerine önemli katkılar sunmaktadır. Gelecekte, yapay zekâ destekli veri yönetimi yaklaşımlarının bu alanı daha da geliştirmesi beklenmektedir (Shi, Cao, Zhang, Li ve Xu, 2016).

Kamu kurumlarının veri mimarisi için hangi yapının kullanılacağı, kurumda oluşan verilerin çeşitliliği ve boyutu, kurum hedefleri, veri kullanım senaryoları, veri tabanlı ihtiyaçlar ve stratejik hedefler dikkate alınarak belirlenmelidir. Örneğin, bir kurumun veri yapısında önemli oranda yapılandırılmamış (video, ses dosyaları vb.) veya yarı yapılandırılmış (JSON, XML vb.) veriler bulunuyorsa, bu verileri ham hâliyle depolayabilen ve ileri düzey analiz ihtiyaçlarına esnek çözümler sunabilen veri gölü (data lake) mimarisi daha uygun olacaktır (Giebler ve diğerleri, 2020). Bununla birlikte, kurumsal uygulamalardan elde edilen ilişkisel veri tabanı kayıtları, mali tablolar ve performans göstergeleri gibi büyük oranda yapılandırılmış veriler için, veri ambarı yaklaşımı önemli avantajlar sunmaktadır (Kimball ve Ross, 2013). Küçük ölçekli bir kurum için ise hem işlenmiş ve analize hazır verileri hem de ham veya yarı yapılandırılmış verileri bir arada depolayabilecek daha özelleşmiş bir sisteme ihtiyaç duyuluyorsa, veri havuzu yapısı bu gereksinimleri karşılayacak bir çözüm olarak öne çıkabilir.

Bu çalışmada kamu kurumlarının bütün verilerini yönetebilecek büyük ölçekli bir veri mimarisi modeli sunmak amaçlandığı için veri ambarlarına ve veri göllerine odaklanılmıştır.

2.6.6. Veri Göllerinin Geleneksel Veri Ambarlarından Farkı

Veri yönetimi ihtiyaçları geliştikçe hem veri gölleri hem de veri ambarları, büyük miktarda veriyi işlemek ve analiz etmek için kullanılan başlıca araçlar haline gelmiştir. Ancak bu iki çözüm, verilerin nasıl depolandığı, işlendiği ve analiz edildiği konusunda önemli farklılıklara sahiptir. Aşağıda veri gölleri ile veri ambarlarının temel farklılıkları ayrıntılı olarak ele alınmıştır.

2.6.6.1. Şema Yapısı ve Veri Formatı

Veri gölleri ve veri ambarları arasındaki en temel farklardan biri, bu sistemlerin şema ve veri formatlarını ele alma biçimidir. Veri gölleri, *schema-on-read* prensibini kullanmaktadır. Bu modelde, veriler herhangi bir dönüşüm veya işleme tabi tutulmadan ham haliyle depolanmaktadır. Verilerin şemalandırılması, analiz veya işleme aşamasında gerçekleştirilmektedir. Bu yaklaşım, büyük miktarlarda yapılandırılmamış veya yarı yapılandırılmış verinin saklanmasına ve daha esnek analiz süreçlerine imkân tanımaktadır. Örneğin sosyal medya verileri, log dosyaları veya IoT cihazlarından gelen sensör verileri ham halde veri göllerine aktarılmakta ve ihtiyaç duyulduğunda işlenmektedir (Inmon, 2016). Terrizzano, Schwarz, Roth ve Colino (2015), *schema-on-read* yaklaşımının veri keşfi ve analizi için sağladığı esnekliği vurgulamıştır. Bu esneklik, özellikle bilinmeyen veri yapıları üzerinde çalışırken veya yeni veri kaynaklarını hızlı bir şekilde alırken büyük avantaj sağlamaktadır.

Veri ambarları *schema-on-write* prensibini kullanmaktadır. Bu modelde ise veriler depolanmadan önce şemalandırılmakta ve işlenmektedir. Veri ambarlarına

aktarılan veriler, belirli bir şema veya yapıya dönüştürülerek saklanmaktadır. Veri ambarları genellikle yapılandırılmış veri kümeleri ile çalışmaktadır ve yüksek performanslı analizler için optimize edilmiştir (Kimball ve Ross, 2013). Madera ve Laurent (2016), *schema-on-write* yaklaşımının özellikle iş zekâsı uygulamaları için optimize edildiğini ve veri tutarlılığı ile performans açısından avantajlar sağladığını belirtmiştir.

Veri formatları açısından, veri mimarilerinin esneklik düzeyi ve sınırlamaları, kullandıkları veri türlerine bağlıdır. Yapılandırılmış veriler, belirli şemalara sahip olup genellikle düzenli bir şekilde saklanmaktadır. Müşteri bilgileri ve finansal raporlama sistemlerinden gelen veriler, yapılandırılmış veri kategorisine girmektedir (Inmon, 2016). Yapılandırılmamış veriler ise belirli bir şema veya formata sahip olmayan verilerden oluşmaktadır. Bu tür verilere örnek olarak metin, görüntü, video veya sosyal medya paylaşımları verilebilir. Yarı yapılandırılmış veriler ise her iki veri türünün özelliklerini bir arada bulundurmaktadır (Miloslavskaya ve Tolstoy, 2016).

Veri gölleri, geniş bir yelpazede veri formatlarını destekleyerek yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerin aynı depolama alanında tutulmasına olanak tanımaktadır. Bu esneklik, veri bilimcilerin çeşitli veri türlerini aynı platformda işleyip analiz yapabilmesini sağlamaktadır (Sawadogo ve Darmont, 2021). Örneğin yapılandırılmış veriler, JSON ve XML dosyaları gibi yarı yapılandırılmış veriler ve metin, görüntü, video gibi yapılandırılmamış veriler aynı veri gölünde bir arada saklanabilmektedir.

Veri gölleri, yapılandırılmış veriler üzerinde de etkili bir çözüm sunabilmektedir. Schema-on-read yaklaşımı sayesinde, yapılandırılmış veriler herhangi bir dönüştürme yapılmadan saklanabilir ve ihtiyaç duyulduğunda şemalandırılarak analiz edilebilir. Bu, verilerin farklı analitik senaryolarda kullanılmasına olanak sağlar ve değişen gereksinimlere hızlı uyum imkânı sunar. Veri gölleri, düşük maliyetli depolama avantajıyla büyük ölçekli yapılandırılmış verilerin verimli bir şekilde saklanmasını ve analiz edilmesini mümkün kılar.

Farklı veri kaynaklarını bütünleştirmek veya uzun vadede analiz yapmak isteyen organizasyonlar için veri gölleri stratejik bir çözüm sunmaktadır. Ancak veri gölleri, veri ambarları ve ilişkisel veri tabanlarıyla karşılaştırıldığında, yapılandırılmış verilerle çalışırken bazı dezavantajlara sahiptir. Özellikle şemalandırma ve işleme aşamalarında performans düşüşleri veya kalite sorunları yaşanabilir. Bu durum, yapılandırılmış verilerin veri göllerinde saklanırken önceden düzenlenmemesi ve sorgulama için belirli bir yapıya sahip olmamasından kaynaklanmaktadır. Bununla birlikte, gelişmiş üst veri yönetimi ve veri kalite kontrol araçlarının kullanılması, bu zorlukların üstesinden gelmek için etkili bir yöntem sunabilir. Öte yandan, veri ambarları daha çok yapılandırılmış verilerle sınırlıdır. Veriler, belirli şema kurallarına uygun şekilde (schema-on-write) depolanmakta ve işlenmektedir. Bu yapı, iş zekâsı uygulamaları için optimize edilmiş olup yapılandırılmış veri analizine ve raporlamasına odaklanmaktadır. Ancak veri ambarları yapılandırılmamış verilerle çalışmak için yeterince esnek değildir.

Her iki yaklaşımın da zorlukları vardır. *Schema-on-read* yaklaşımında veri kalitesi ve yönetişimi sorunları ortaya çıkabilirken *schema-on-write* yaklaşımında değişen veri gereksinimlerine adaptasyon zorluğu yaşanabilmektedir (Sawadogo ve diğerleri, 2019). Bu nedenle, günümüzde bazı modern veri platformları her iki yaklaşımı da destekleyen hibrit çözümler sunmaktadır (Hai ve diğerleri, 2016). Organizasyonlar, veri yönetimi stratejilerini belirlerken bu farklılıkları dikkate almalıdır. Veri gölleri ve veri ambarları, birbirini tamamlayıcı roller üstlenebilir.

2.6.6.2. Depolama Maliyetleri

Veri gölleri, büyük miktarda ham veriyi düşük maliyetlerle depolamak için tasarlanmıştır. Hadoop gibi dağıtık dosya sistemleri veya bulut tabanlı çözümler (örneğin AWS S3, Google Cloud Storage vb.) veri göllerinin büyük miktarda veriyi uygun maliyetlerle saklamasını sağlamaktadır. Veri gölleri, veri alımı esnasında veriyi işleme ihtiyacı duymadığı için bu aşamada işlem maliyeti oluşmamaktadır. Walker ve Alrehamy (2015), veri göllerinin “store-first, analyze-later” (önce

depola, sonra analiz et) yaklaşımının maliyet etkinliğini vurgulamıştır. Bu yaklaşım, organizasyonların veri toplama ve depolama süreçlerini hızlandırırken analiz maliyetlerini ihtiyaç duyulduğunda ortaya çıkarmaktadır. Bununla birlikte, veri göllerinin yönetimi için gerekli olan üst veri sistemleri, işletme maliyetlerini artırabilmektedir. Sawadogo ve Darmont (2021), veri göllerinin zaman içinde büyüdükçe yönetim maliyetlerinin artabileceğini vurgulamıştır.

Diğer yandan veri ambarları, verilerin şemalandırılması ve işlenmesi için daha fazla işlem gücü ve altyapı gerektirmektedir. Bu nedenle depolama maliyetleri veri göllerine kıyasla daha yüksektir. Veri ambarları, yapılandırılmış verilerin performanslı bir şekilde sorgulanmasını sağlamak amacıyla optimize edildiğinden, daha karmaşık ve pahalı altyapı gereksinimlerine sahip olabilmektedir (Kimball ve Ross, 2013).

2.6.6.3. Veri Yönetimi ve Performans

Veri göllerinde ham veri doğrudan depolandığı için veri yönetimi daha esnektir. Ancak bu esneklik aynı zamanda veri göllerinin karmaşık olmasına yol açabilir. Verilerin anlamlandırılabilmesi ve bulunabilmesi için güçlü bir üst veri yönetim sistemi gereklidir. Aksi takdirde, veri gölleri veri bataklığı olarak bilinen kullanışsız veri yığınlarına dönüşebilir. Veri gölleri, büyük veri analitiği ve makine öğrenimi gibi veri yoğun uygulamalarda üstün performans sağlamaktadır (Miloslavskaya ve Tolstoy, 2016). Ancak belirli yapılandırılmış veri analizlerinde veri ambarlarına göre daha yavaş olabilmektedir (Fang, 2015). Terrizzano ve diğerleri (2015) veri göllerinin büyük veri analitiği için sağladığı performans avantajlarını detaylı bir şekilde incelemiştir. Veri göllerinin çeşitli veri türlerini bir arada işleyebilme yeteneği, karmaşık analitik işlemlerde önemli performans artışları sağlayabilmektedir. Ayrıca veri göllerinin dağıtık mimarisi, büyük ölçekli paralel işleme imkânı sunarak geleneksel sistemlere göre daha hızlı sonuçlar elde edilmesine olanak tanımaktadır.

Veri ambarları, yapılandırılmış verilerin hızlı bir şekilde sorgulanması ve raporlanması için optimize edilmiştir. Verilerin şemalandırılmış olması, sorgu performansını artırmakta ve kullanıcıların hızlı bir şekilde veri analizi yapmasına olanak tanımaktadır. Veri ambarları, büyük veri kümeleri üzerinde karmaşık sorgular çalıştırmak için daha iyi bir performans sağlamaktadır. Ancak veri ambarları, yapılandırılmamış veya yarı yapılandırılmış veri türleri üzerinde aynı avantajları sunamamaktadır (Sawadogo ve Darmont, 2021). Abadi ve diğerleri (2020), çalışmalarında modern veri ambarı çözümlerinin performans iyileştirme tekniklerini incelemiştir ve sütun tabanlı depolama, verilerin doğrudan bellekte işlenmesi ve adaptif sorgu optimizasyonu gibi tekniklerin, veri ambarlarının performansını önemli ölçüde artırdığı vurgulamıştır. Bu gelişmeler, veri ambarlarının özellikle yapılandırılmış veriler üzerinde çok hızlı analiz ve raporlama yapabilmesini sağlamaktadır. Bununla birlikte veri ambarları büyük ölçekli yapılandırılmamış veri işleme konusunda veri göllerine göre dezavantajlı durumdadır.

Kapsamlı literatür taraması sonucunda oluşturulmuş bu bölüm, “Veri gölü mimarilerinin uygulanması, kamu kurumlarının veri analiz kapasitesini artırabilir” hipotezini destekler niteliktedir.

3. BÖLÜM
KAMUSAL VERİ KAYNAĞI OLARAK KURUMSAL BELGE VE DİJİTAL ARŞİV UYGULAMALARI

Kurumlar, faaliyetlerini sürdürürken çeşitli bilgiler üretir ve kullanır. Bu bilgiler, hem günlük operasyonların etkin bir şekilde yürütülmesi hem de kurumun yönetimi için kritik öneme sahiptir. Bu bağlamda kurumsal bilgi, “Kurumun etkinliği, denetimi, yönetimi, geliştirilmesi vb. ile ilgili olarak üretilen, alınan, kullanılan ve bu nedenlerle özel bir nitelik kazanan her türlü malumat, fikir ve olgulardır.” olarak tanımlanmaktadır. Bilgi yönetimi ise, “Bir kurumun entelektüel kaynaklarının üretimini, tanımlanmasını, değerlendirilmesini, tutulmasını, geliştirilmesini ve dağıtımının belirli bir plan dahilinde yürütülmesini sağlayan teori ve uygulamalardır.”. Kurumsal bilgi, doğrudan kurumun faaliyetleri esnasında ihtiyaç duyulan işletim bilgisi veya kurum faaliyetlerinin geliştirilmesi ve denetlenmesinde ihtiyaç duyulan yönetim bilgisi vb. bilgileri içerir. Bu faaliyetler esnasında üretilen belgeler ise kurumsal bilgi kaynaklarını oluşturur. Bu belgeler kurumsal hafızanın devamlılığı bakımından önemli olmakla birlikte, milli arşivin kaynakları olarak milletin hafızası olarak da önemli olabilmektedirler (Özdemirci, 2001, s. 2).

Bu kapsamda, 11 numaralı Cumhurbaşkanlığı Kararnamesi, kamu kurum ve kuruluşlarında belge yönetiminin etkili bir şekilde yürütülmesini sağlamak amacıyla rehberlik, eğitim ve denetim faaliyetlerini düzenlemiş ve elektronik belge yönetimi ile arşivleme sistemlerinin önemine dikkat çekmiştir (Resmî Gazete, 2018b). Elektronik belge yönetimi, belgelerin daha etkili bir şekilde üretilmesi, saklanması ve paylaşılması için güçlü bir altyapı sunarak kurumların bilgi yönetim süreçlerini desteklemekte ve arşivcilik faaliyetlerine uyumlu bir yapının oluşturulmasına katkı sağlamaktadır. Bu düzenlemeler, hem kurumsal bilgiyi modern teknolojilerle entegre etmeyi kolaylaştırmakta hem de belgelerin güvenli bir şekilde korunmasını sağlamaktadır.

3.1. VERİ ARŞİVLERİ

Geçmişte, çok az belge üretilen dönemlerde belgelere erişim büyük bir problem oluşturmazken günümüzde katlanarak artan belge sayısı önemli ve gelecekte ihtiyaç duyulabilecek belgelerin diğer belgelerden ayrılarak saklanması ihtiyacını doğurmuştur. Bunun sağlanması için de öncelikle kurumların programlı bir şekilde arşivlerini yapılandırması gerekir (Arısoy, 2018, s. 1). Bu yapılandırma, belgelerin düzenli bir şekilde saklanması, erişim süreçlerinin optimize edilmesi ve uzun vadeli bilgi yönetiminin sağlanması açısından kritik bir adımdır.

Arşivlerin önemi ancak arşivlerde yer alan bilgiye ihtiyaç duyulduğunda ortaya çıkmaktadır (Arısoy, 2018, s. 64). Bu nedenle kurumlarda üretilen bilgi ve belgeler, ihtiyaç duyulduğunda başvurulmak üzere düzenlenerek ve üst bilgileri tanımlanarak kurumsal arşivlerde saklanmalıdır. İhtiyaç duyulan bilgi/belgelere etkin bir şekilde erişim sağlanması öncelikle bilgi/belgelerin belirli kurallar dahilinde, oluşturulan arşivlerde depolanmasını ve korunmasını gerektirir. Geleneksel arşivlerde bu süreç basılı dokümanlar için yürütülürken, dijital arşivlerde elektronik ortamda üretilen veya basılı olarak üretildikten sonra dijitalleştirilen dokümanlar esastır.

Elektronik arşiv sistemlerinde farklı türdeki arşiv malzemesinin saklanmasına yönelik olarak birçok ülkede ulusal ve uluslararası düzeyde çalışmalar yapılmaktadır. Arısoy çalışmasında İngiltere, Fransa, Kanada, Kuzey İrlanda, Amerika Birleşik Devletleri, Estonya, İsviçre, Avustralya, Hollanda ve Yeni Zelanda ülkelerinin arşiv politikalarını değerlendirmiştir. Bu konuda Türkiye'nin durumu da ele alarak diğer ülke uygulamalarının ülkemizde kullanılabilirliğini incelemiştir (2018).

Veri arşivleri, verilerin düzenli bir şekilde toplanması, depolanması, yönetilmesi ve erişilebilir olması ile uzun süreli saklanmasını amaçlayan yapılardır (Akmon, Zimmerman, Daniels ve Hedstrom, 2011; Çalım ve Tuna, 2022; "Enfobilişim", t.y.). Araştırma merkezleri, devlet kurumları, sanayi kuruluşları gibi pek çok kurum bilimsel araştırma, analiz, proje vb. verilerini uzun süreli korumak ve tekrar

kullanmak için veri arşivlerine ihtiyaç duymaktadır. Bu arşivler sayesinde kurumlar bilgi birikimlerini koruyarak gelecekte yapılacak çalışmalara katkıda bulunabilmektedir. Ülkemizde ve dünyada farklı kurumlar tarafından oluşturulan ve desteklenen birçok veri arşivi bulunmaktadır. Bu başlık altında, önce ülkemizdeki, sonra ise dünyadaki başlıca veri arşivleri kısaca açıklanacaktır.

3.1.1. Elektronik Belge Yönetimi ve Arşiv Uygulamaları

Kurum ve kuruluşlar bilgilendirmek ve iletişim kurmak, neyi yaptıklarını, yaptıkları iş ve işlemleri neden ve nasıl yaptıklarını kayıt altına almak, yasal ve hukuki işlemler, denetimler, vergi süreçleri ve diğer devlet kurumları süreçlerinde kullanmak, yükümlülük ve sorumluluklarını göstermek, gelecek nesillere yaptıklarını aktarmak vb. amaçlarla belge üretirler (Özdemirci 2001, s. 4; Sprehe 2000, s. 15).

Belge,

"1. Kurumların veya bireylerin iş ilişkilerini yürütebilmek veya yasal yükümlülüklerini yerine getirebilmek için ürettikleri, aldıkları ve delil olarak korudukları üzerine bilgi kayıtlı her türlü fiziksel ortam. 2. Herhangi bir bireysel veya kurumsal fonksiyonun yerine getirilmesi için alınmış, ya da fonksiyonun sonucunda üretilmiş, içerik, ilişki ve formatı ile ait olduğu fonksiyon için delil teşkil eden üzerine bilgi kayıtlı her türlü fiziksel ortam. 3. Bir şahıs, kurum ya da kuruluş tarafından üretilen ya da kuruluşa gelen fiziksel yapısı ya da özelliği ne olursa olsun, gelecekte kültürel, idari ve hukuki nedenlerden dolayı kullanılabilecek ve dolayısıyla kanıt niteliği de taşıyabilecek her türden kaydedilmiş bilgi. 4. İşlemlerin veya yasal zorunlulukların yerine getirilmesinde bir kişi veya organizasyon tarafından delil olarak üretilen, kabul edilen (alınan) ve korunan enformasyondur."

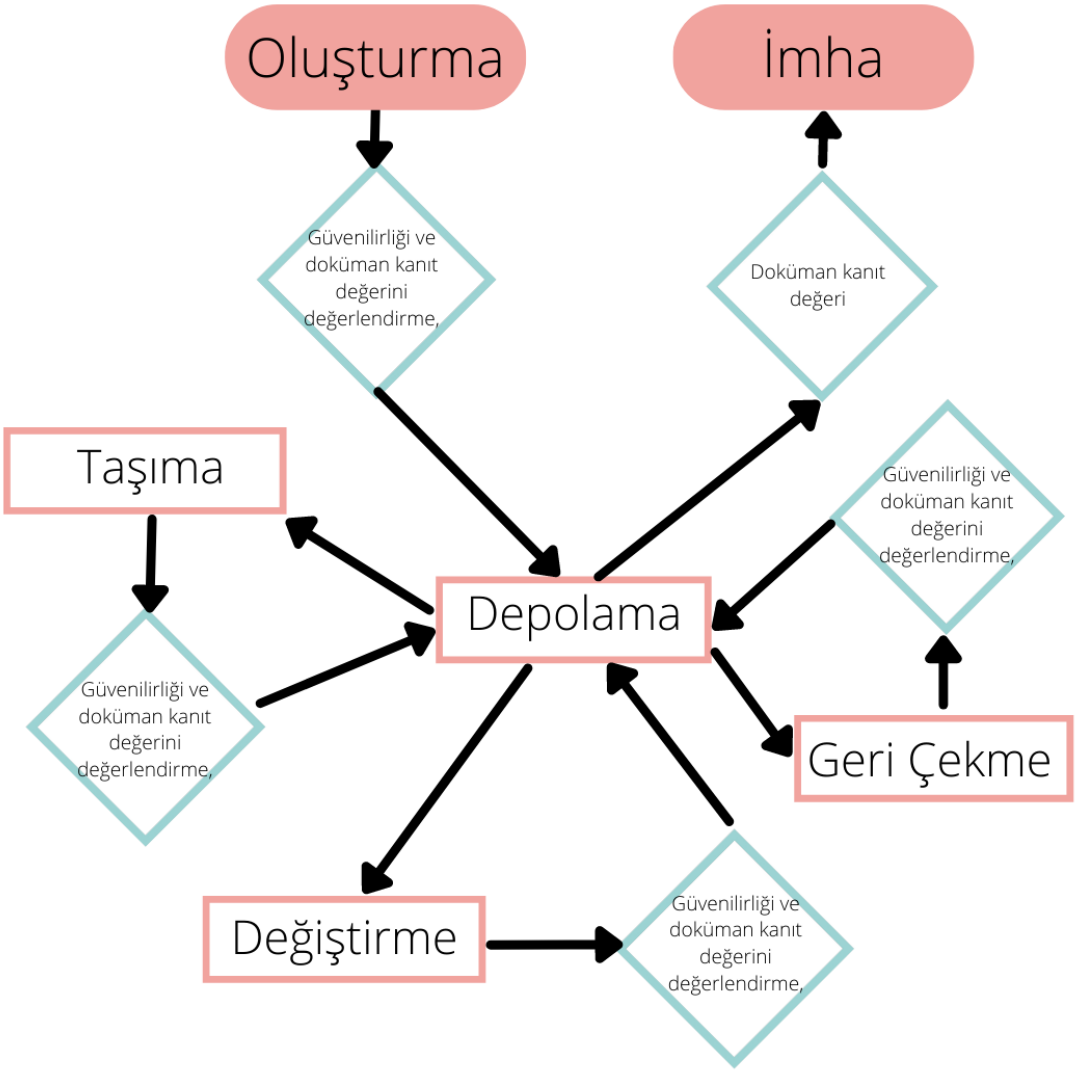
olarak farklı şekillerde tanımlanmaktadır (Özdemirci, Torunlar ve Saraç, 2009, s. 2).

Belge yönetimi ise, kurumsal bilgi kaynaklarının yönetimi ile birlikte bu belgelerden arşivsel değer taşıyanların belirlenmesi süreçlerini kapsamaktadır. Belge yönetimi, kısaca, ihtiyaç duyulan bilgi-belgelere etkin erişim sağlanması amacıyla bilgi kaynaklarının düzenlenmesi ve depolanması, tarihi değer belgelerin seçiminin yapılması olarak tanımlanabilir (Özdemirci, 2004, s. 198-199). Belge yönetimi kurumsal bilginin üretilmesinden tekrar erişilmesine kadarki onay, kullanım, depolama, koruma gibi bütün süreçleri kapsar. Belgeler kurumların yönetiminde oldukça önemlidir bu nedenle belge yönetiminin de önemi açıktır. Belge yönetiminin etkin bir şekilde yürütülebilmesi için kurumların belge yönetim sistemlerine ihtiyacı bulunmaktadır. Belge yönetiminin sistematik bir şekilde gerçekleştirilmesi, bilgi isteklerine etkin bir şekilde cevap verilebilmesi, kırtasiyeyi azaltma, tarihi belgeleri koruma vb. süreçlere de katkı sağlar. Ayrıca kurumsal bilginin bireylere bağlı olarak yürütülmesine engel olur. Kurumlar süreçlerini personel değişikliği, personel kaybı vb. olaylardan en az etkilenecek şekilde yürütebilirler. Yapılan işi, işi yapan kişi veya diğer koşullardan en az etkilenecek şekilde sürdürmeye olanak sağlanır böylece kalite, güvenlik ve sürekliliğin temeli sağlanmış olur (Özdemirci, 2001, s. 4, 5). Özdemirci bilgi ve belge yönetiminin amacını “kurumun faaliyetleri sonucu üretilen tüm belgeleri, üretim aşamasından itibaren ele alarak belirlenmiş ölçütler çerçevesinde değerlendirmek, düzenlemek istenildiğinde hizmete sunmak için sürekli olarak kullanıma hazır bulundurmaktır.” olarak tanımlamıştır (Özdemirci, 2001, s. 5). Belge yönetimi, kurumsal bilginin sistematik bir şekilde düzenlenmesi ve korunmasını sağlar. Bu süreç, kurumsal bilgi kaynaklarının etkin yönetimi ile erişilebilirliğini artırarak, kurumların operasyonel verimliliğini ve sürekliliğini destekler.

Bilgi ve belge yönetimi en az bilgiyi üretmek kadar önemlidir. Özdemirci’nin belirttiği gibi “Belgesini yönetemeyenler kurumlarını, bilgisini yönetemeyenler insanları ve toplumları yönetemezler.” (2019c). Bilgi ve belge yönetim sistemleri farklı sistemlerle entegre olarak çalışmayı gerektirir. Bilgi yönetimi öncelikle bilginin üretilmesi ve kayıt altına alınmasını gerektirir. Bu süreçten sonra bilgi güvenli ortamlarda saklanarak ihtiyaç duyulduğunda erişilebilir ve analiz edilebilir.

Günümüzde bu süreçler elektronik ortamlarda yenilikçi bilgi teknolojileri kullanılarak gerçekleştirilmektedir (Özdemirci, 2019b, s. 4).

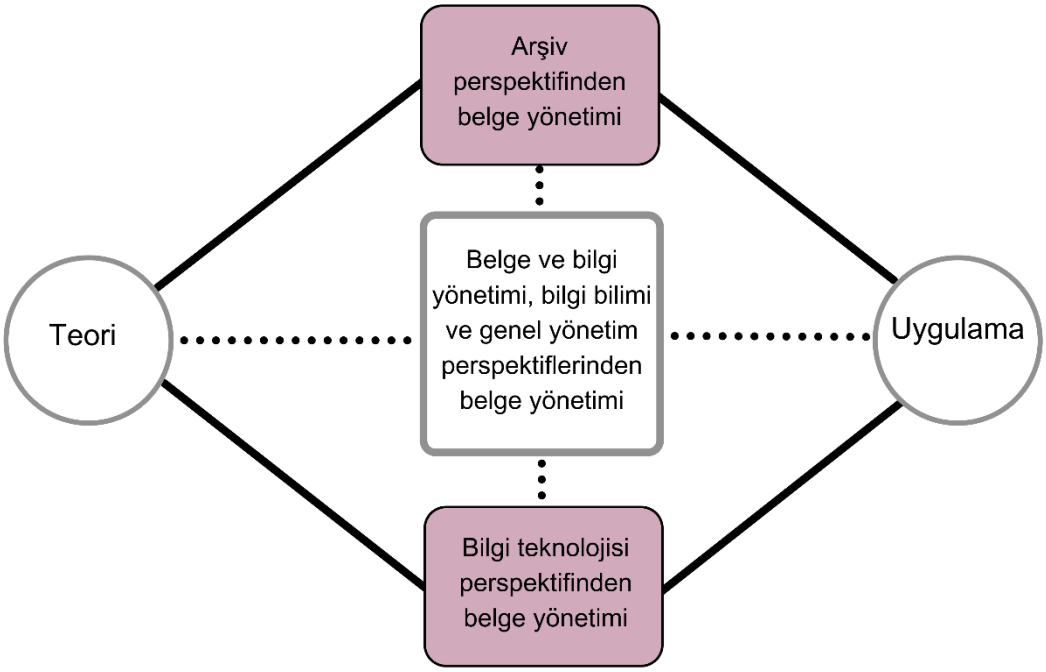
Bilgi yönetimi sistemleri, donanım-yazılım ve iş-işlem süreçleri olmak üzere iki ana yapıdan oluşmaktadır. Bilgi yönetim sistemleri "teknoloji odaklı bilgi yönetimi, bilgi odaklı yönetim sarmalında oluşan bir platform" olarak değerlendirilmelidir. Bilginin üretilmesi kadar bilginin güvenliğinin ve güvenirliğinin sağlanması da önemlidir ve günümüzde bu süreç bilgi üretiminden daha maliyetli olmaktadır (Özdemirci, 2019b, s. 5). Şekil 10'da kamu bilgisinin güvenilirlik ve kanıt değerinin belgenin oluşturulmasından imhasına kadarki süreçte hangi aşamalarla ilgili olduğunu gösteren model verilmiştir (Ma, Abie, Skramstad ve Nygård 2011, s. 13). Bilişim alanında yaşanan gelişmelerle birlikte bilgi ve belge yönetim süreçleri ve işlemleri daha hızlı ve ekonomik şekilde sürdürülmekte, kurumsal bilgi sistemleri, e-arşiv sistemleri, elektronik belge yönetim sistemleri kurumlara ve kullanıcılara kolaylık sağlamaktadır (Özdemirci, 2019b, s.4). Bilgi yönetim sistemleri etki alanı ve boyutları sürekli olarak gelişen dinamik platformlardır (Özdemirci, 2019b, s. 9).



Şekil 10. Kamu bilgisi güvenilirlik ve kanıt değeri (Ma ve diğerleri, 2011, s. 13)

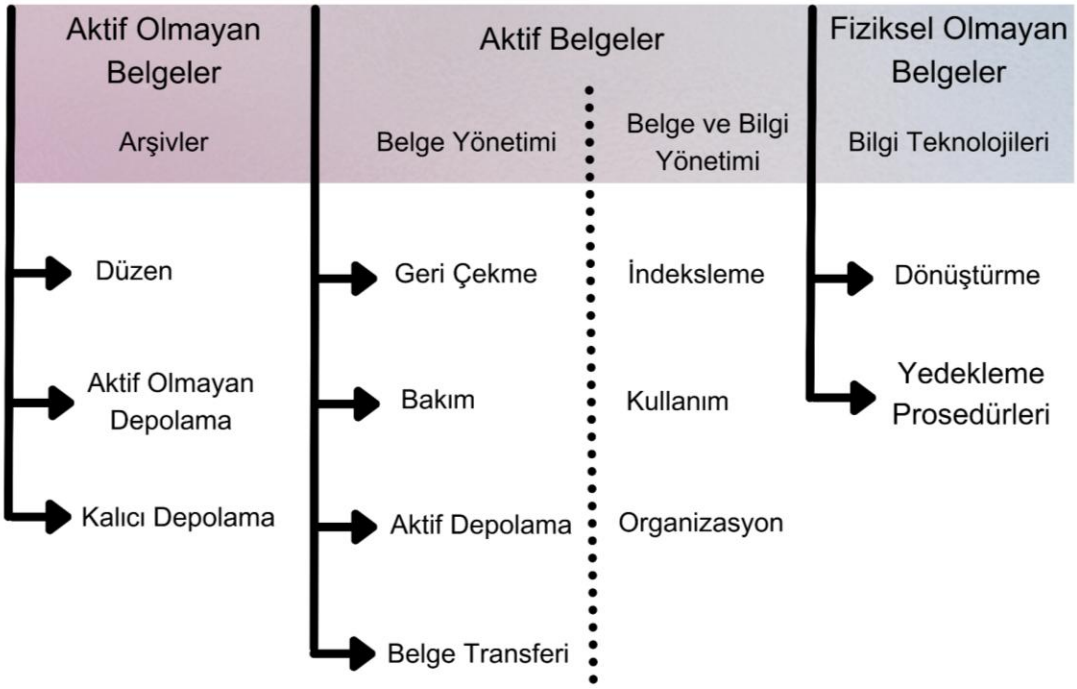
Bilgi yönetim sistemleri yönetilirken ve kullanılırken mahremiyet dikkate alınmalıdır. Bu bağlamda bu konuda yetkinliği bulunan Belge Yönetimi ve Arşiv Birimleri'nin kurumsal bilgi ve belge süreçlerinde aktif rol alması önemlidir. (Özdemirci, 2019b, s. 4, 5).

Şekil 11'de, belge yönetimi sürecinin farklı bilim dallarının iş birliği ile nasıl daha etkili bir şekilde yürütülebileceği gösterilmektedir (Yusof ve Chell, 2002, s. 59).



Şekil 11. Bilgi ve Belge Yönetimi (Yusof ve Chell, 2002, s. 59)

Şekil 12'de genel hatlarıyla belge yönetimi çerçevesi farklı süreçlerde gerektirdiği farklı alan yetkinlikleriyle birlikte gösterilmiştir (Yusof ve Chell, 2002, s. 59).



Şekil 12. Belge yönetimi çerçevesi (Yusof ve Chell, 2002, s. 59)

Bu doğrultuda, bilgi ve belge yönetimi süreçlerinin etkili bir şekilde yürütülmesi, kurum içi ve dışı bilgi sistemlerinin uyumlu bir şekilde çalışmasıyla mümkün olmakta, bu entegrasyonun kurumsal süreçlere katkısı ise giderek daha fazla önem kazanmaktadır. Bu bağlamda, kurumların bilgi ve belge yönetiminde hedeflerini gerçekleştirebilmeleri için elektronik belge yönetim sistemleri kritik bir adım olarak değerlendirilmektedir. Kurumlarda üretilen bilginin idari, mali ve hukuki değer taşıyan belgelere dönüşmesi sürecinde Elektronik belge yönetim sistemi (EBYS) entegrasyonu gereklidir. EBYS'ler, belgelerin düzenli, erişilebilir ve güvenli bir şekilde yönetilmesini sağlamanın yanı sıra kurumsal belleğin oluşmasında ve kurumların sürdürülebilirliklerinin sağlanmasında önemli bir rol oynayan belge tabanlı çalışan bilgi yönetim sistemleridir.

Bir kurumdaki EBYS verilerinin etkin bir şekilde kullanılabilmesi için, kurum içindeki veri ambarı veya veri gölü gibi bir veri mimarisiyle entegre edilmesi gereklidir. Bu entegrasyon, EBYS'nin ürettiği uygulama ve metin verilerinin kurumsal düzeyde depolanmasını, yönetilmesini ve analiz edilmesini sağlayacaktır. Bütünleşmiş bir veri mimarisi tasarlanırken EBYS'den gelen verilerin dikkate alınması önemlidir.

EBYS'lerin diğer bilgi sistemleriyle de entegre çalışması önemlidir. İnsan Kaynakları Bilgi Sistemi, Proje Bilgi Sistemleri, Mali Bilgi Sistemleri kurum içi bilgi sistemlerinden birkaçıdır. Bunlara ek olarak EBYS'lerin Devlet Teşkilatı Sistemi (DETSİS), Elektronik İmza (e-İmza), Zaman Damgası, İmza Yetkilileri Modülü (İYEM), Kayıtlı Elektronik Posta (KEP), e-Devlet gibi kurum dışı bilgi sistemleriyle de entegre olması gereklidir (Özdemirci, 2019b, s. 5, 6). EBYS'lerin yasal düzenlemelere uyumlu hale getirilmesi, özellikle elektronik imza ve zaman damgası gibi teknolojilerin entegrasyonu ile mümkün olmaktadır (Külcü, 2009). EBYS mimarileri oluşturulurken ve sistemler yönetilirken kurumların yapısı ve ihtiyaçları göz önüne alınmalıdır (Özdemirci, 2019b, s. 8).

Literatürde, EBYS uygulamalarının yalnızca belge yönetim süreçlerini iyileştirmekle kalmayıp, aynı zamanda kurumsal faaliyetlerin verimliliğini ve karar

alma mekanizmalarının etkinliğini artırdığı vurgulanmaktadır (Dzikria, Narulita, Hermanto ve Kusnanto, 2022). Ayrıca bu sistemlerin kullanıcı dostu bir tasarımla geliştirilmesi ve veri güvenliğinin sağlanması, etkili bir belge yönetimi için temel unsurlardır. EBYS'lerin etkili kullanımı, kurumsal bilgi yönetim süreçlerini desteklerken, kullanıcı memnuniyetini artırmak ve bilgiye hızlı erişim sağlamak gibi avantajlar sunar. Bu bağlamda, teknolojik altyapının sürekli geliştirilmesi ve sistemlerin kullanıcı ihtiyaçlarına göre optimize edilmesi, EBYS'lerin uzun vadeli başarısı için büyük öneme sahiptir (Azmir ve Wijayanti, 2022).

3.1.2. Veri Arşivleri ve Türkiye'deki Durum

3.1.2.1. Aperta - Türkiye Açık Arşivi

TÜBİTAK ULAKBİM tarafından yürütülen TÜBİTAK Açık Arşivi'nin (TÜBİTAK desteğiyle üretilen veya TÜBİTAK araştırmacılarının ürettiği yayınlara ve verilere açık erişim sağlayan arşiv (*TÜBİTAK Açık Bilim Politikası*, t.y., s. 2) kapsamı üniversiteler ve kamu kurumlarında çalışan araştırmacılara açılarak genişletilmiş, Aperta Türkiye Açık Arşivi oluşturulmuştur. Aperta'nın amacı araştırma verilerinin saklanması, korunması ve yönetilmesini ve taranarak verilere ücretsiz olarak erişilebilmesini sağlamak, ulusal bir veri deposu oluşturmaktır. Araştırmacıların açık erişimli yayınları ile birlikte, ilgili verileri ve yazılımları da Aperta platformu üzerinden paylaşması amaçlanmaktadır. Aperta araştırmacılara ücretsiz olarak DOI desteği vermekte, veri setlerinin; kitap, makale, patent, rapor, tez vb. yayınların; şekil, çizim, fotoğraf vb. resimlerin; yazılım, video, ders vb. materyallerin güvenli olarak paylaşılmasına ve kullanılmasına olanak sağlamaktadır ("Aperta - Türkiye Açık Arşivi", t.y.).

Aperta, Creative Commons Zero lisansı sayesinde üst verilerin yasal kısıtlama olmaksızın kullanılmasını ve yeniden düzenlenmesini sağlamaktadır. Bu sistem, OAI-PMH protokolünü desteklediği için Dublin Core, MARCXML ve DataCite gibi farklı üst veri formatlarını kullanmakta ve bu sayede farklı kullanıcı gruplarının

verilere kolayca erişebilmesine olanak tanımaktadır. Ayrıca Aperta, Re3data, Harman ve OpenDOAR gibi uluslararası rehberlerde listelenmiş olup bu rehberler üzerinden erişilebilirliği artırmaktadır. DOI'ye sahip olmayan veri setlerine DOI atayarak bu veri setlerinin görünürlüğünü ve erişilebilirliğini desteklemektedir. Üst veriler, DataCite standartlarına uygun olarak hazırlanmakta ve Schema.org formatı kullanılarak arama motorları için optimize edilmektedir. Bu özellik, yayın, veri seti ve yazılım gibi farklı türdeki kaynaklar arasında semantik bağlantılar kurulmasına olanak tanımaktadır. Aperta'nın teknik altyapısı, CERN tarafından geliştirilmiş Zenodo yazılımının özelleştirilmiş bir sürümüne dayanmaktadır. Ayrıca, API entegrasyonu sayesinde diğer sistemlerle kolayca entegre olmakta ve araştırmacılara metin ve veri madenciliği faaliyetleri için destek sağlamaktadır.

3.1.2.2. T.C. Sanayi ve Teknoloji Bakanlığı Girişimci Bilgi Sistemi

Ulusal Sanayi Veri Master Planı'nda yer verilen Girişimci Bilgi Sistemi (GBS), Sanayi ve Teknoloji Bakanlığı'nın, kamu kurumlarının ekonomik faaliyetleriyle ilgili verilerinin belirli standartlara uygun olarak bir veri tabanında toplanmasını amaçlayan projesidir. GBS ile, kamu kurumları, üniversiteler ve diğer araştırma birimlerindeki karar alıcılar, politika yapıcılar ve araştırmacılar için doğru ve güvenilir veri sunulan bir ortam oluşturulmuştur. GBS'nin yetenek ve kapsamının genişletilerek sanayi ile ilgili verilerin bulunduğu bir Veri Merkezi oluşturulacağı belirtilmektedir ("T.C. Sanayi ve Teknoloji Bakanlığı", t.y.).

3.1.2.3. Kamu Uygulamaları Merkezi

Kamu Uygulamaları Merkezi, T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı bünyesinde faaliyet gösteren, kamu kurumlarının hizmet sunumu ve veri paylaşımı yapabileceği ayrıca hizmetlerle ilgili istatistiksel verilere ulaşabileceği platformdur ("E-Devlet Kapısı", t.y.).

3.1.2.4. AçıkVeri Projesi

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi Başkanlığı tarafından yürütülen AçıkVeri Projesi, kamu kurumlarının iş ve hizmet sunum süreçlerinde ürettikleri verilerin paylaşımını ve bu süreçlerdeki şeffaflık ve hesap verilebilirlik ilkelerinin benimsenmesini teşvik etmeyi amaçlayan bir girişimdir. Proje ile, kişisel veriler ile ulusal ve ticari sırların mahremiyet ilkelerinin ve bu süreçlerle ilgili gizlilik ve güvenlik önlemlerinin dikkate alınarak, ayrıca veri paylaşımında uyulması gereken diğer ilkeler uygulanarak katma değerli hizmetlerin üretilmesini desteklemek hedeflenmektedir (“Türkiye Cumhuriyeti Cumhurbaşkanlığı”, t.y.a).

3.1.2.5. Türkiye Ulusal Kamu Entegre Veri Merkezi Projesi

Kamu Entegre Veri Merkezi Projesi, 5809 sayılı Elektronik Haberleşme Kanunu kapsamında Ulaştırma ve Altyapı Bakanlığı tarafından yürütülmektedir. 2017 yılında yüklenici firma ile sözleşme imzalanarak çalışmalara başlanmıştır. Faz 1'de mevcut durum, ihtiyaç analizi ve stratejik rapor tamamlanmış olup, Faz 2'de kavramsal ve ayrıntılı veri merkezi tasarımları, arazi etüdü ve mimari projeler üzerinde çalışılmaktadır (*T.C. Ulaştırma ve Altyapı Bakanlığı,* 2020, s.99)

Ulusal Kamu Entegre Veri Merkezi (UKEVM), kamuya ait verilerin tek bir merkezde güvenli ve verimli şekilde yönetilmesini sağlayan bir ekosistemdir. UKEVM, kamu kurumlarının ayrı ayrı veri merkezi kurma ihtiyacını ortadan kaldırarak maliyet, enerji ve operasyonel verimlilik sağlar. Siber güvenlik tehditlerinin tek noktadan yönetilmesi ve standardizasyon sayesinde daha hızlı tespit ve önleme imkânı sunar. Kurumlar arası güvenli iletişim protokolleriyle bilgi güvenliğinin artması hedeflenmiştir. Güncel teknolojilerle bilgi hizmetleri daha ekonomik ve verimli hale gelirken, Kamu bilgi teknolojisi stratejilerinin gelişimine de katkıda bulunur (“T.C. Ulaştırma ve Altyapı Bakanlığı”, 2024).

3.1.2.6. TÜİK Veri Arşivi

Türkiye İstatistik Kurumu (TÜİK) tarafından yönetilen İstatistikVeri Portalı, Türkiye'nin resmi istatistiklerini ve verilerini toplayan, düzenleyen ve yayınlayan kapsamlı bir veri arşividir. Bu portal, ülkenin sosyal, ekonomik ve demografik verilerini geniş bir yelpazede sunmaktadır. TÜİK'in veri arşivi, adalet ve seçimden bilim ve teknolojiye, çevre ve enerjiden dış ticarete kadar 18 ana kategoride istatistiksel bilgi sağlamaktadır. Araştırmacılar, politika yapıcılar ve genel kullanıcılar için tasarlanan bu platform, 169 konu başlığı altında 3.508 bülten, 624.938 gösterge ve 60 milyondan fazla veri noktası içermektedir. Kullanıcılar, ihtiyaç duydukları verilere kolayca erişebilir, çeşitli formatlarda indirebilir ve analiz edebilirler. TÜİK'in veri arşivi, Türkiye'nin sosyo-ekonomik durumunu anlamak, politika geliştirmek ve akademik araştırmalar yapmak için önemli bir kaynak oluşturmaktadır (“Türkiye İstatistik Kurumu”, t.y.b).

3.1.3. Diğer Ülkelerdeki Veri Arşivleri ve İlgili Kuruluşlar

3.1.3.1. Data.gov Veri Arşivi

Data.gov, Amerika Birleşik Devletleri hükümetinin federal kurumlar tarafından yayınlanan veri setlerine erişim sağladığı, böylelikle hükümetin açık verilerine halkın erişimini kolaylaştırarak, kurumların misyonlarını gerçekleştirmesine destek olmayı, ekonomik faaliyetleri canlandırmayı ve şeffaf bir hükümet anlayışının benimsenmesini hedefleyen açık veri web sitesidir (“The Home of the U.S. Government's Open Data”, t.y.).

3.1.3.2. Avustralya Veri Arşivi (Australian Data Archive: ADA)

Avustralya Veri Arşivi (ADA), 1981 yılında kurulan ve Avustralya Ulusal Üniversitesi'ndeki Sosyal Araştırma ve Yöntemler Merkezi bünyesinde faaliyet gösteren “Core Trust Seal” sertifikalı bir veri deposudur. Core Trust Seal, bir veri

deposunun güvenilirlik, uyumluluk ve sürdürülebilirlik açısından uluslararası standartlara uygun olduğunu, yani verilerin uzun vadeli korunduğunu ve erişilebilir olduğunu gösterir. ADA siyasi, sosyal ve ekonomik konulardaki dijital verilerin toplanarak korunması ve farklı disiplinlerdeki araştırmacılar için erişilebilir hale getirilerek ileri analizler için kullanılması amacıyla hizmet vermektedir. 5000'in üzerinde veri setiyle Avustralya'da sosyal bilimler alanındaki en kapsamlı kataloga sahiptir. Ülkedeki anket, nüfus sayımı vb. ile elde edilen verileri de barındırmaktadır. Ayrıca Asya Pasifik bölgesindeki diğer ülkelerden de veri içermektedir. DDI ve OAIS gibi uluslararası en iyi uygulamalara uygun olarak standartlar geliştirip uygulamaktadır ve uluslararası iş birliği projelerine önem vermektedir. Veri arşivcilerinden oluşan bir ekip, Avustralya toplumuna hizmet sunmaktadır. Çevrimiçi veri hizmetlerini ve gerekli altyapıyı "The National Computational Infrastructure" sağlamakta ve geliştirmektedir ("Australian Data Archive", t.y.).

3.1.3.3. Birleşik Krallık Veri Arşivi (UK Data Archive)

Birleşik Krallık Veri Arşivi (UK Data Archive), 1967'de Sosyal Bilimler Araştırma Konseyi Veri Bankası (Social Science Research Council Data Bank) olarak kurulmuştur. 2005 yılında ise, Ulusal Arşiv için kamu kayıtlarının toplandığı bir depo haline gelmiştir. Araştırmacılara destek ve eğitim vermekle birlikte ülkenin sosyal, ekonomik ve nüfus verilerine erişim sunmaktadır. Ulusal İstatistik Ofisi (Office for National Statistics) tarafından yürütülen anket verilerini ve ulusal sosyal araştırma merkezi verilerini de sunmaktadır ("UK Data Archive", t.y.).

3.1.3.4. Japonya Ulusal Bilişim Enstitüsü Araştırma Veri Bulutu (Japan National Institute of Informatics Research Data Cloud)

Japonya Ulusal Bilişim Enstitüsü (NII) araştırma veri bulutu Japonya'daki araştırma verilerinin yönetimini, depolanmasını ve paylaşımını kolaylaştırmak

amacıyla kurulmuş bir platformdur. Araştırmacıların veri setlerini güvenli bir şekilde depolamaları ve bu verilere kolayca erişebilmeleri için gelişmiş bir altyapı sunar. Ayrıca üniversiteler, araştırma enstitüleri ve özel sektör ile iş birliğine büyük önem vermektedir ve veri paylaşımı ve iş birliği olanaklarını artırmak için çeşitli araçlar ve hizmetler sağlamaktadır. Araştırma veri yönetimi platformu (GakuNin RDM), yayın platformu (WEKO3) ve keşif platformu (CiNii) bulunmaktadır (“NII Inter-University”, t.y.; “Research Center”, t.y.).

3.1.3.5. Eurostat (Eurostat is the statistical office of the European Union

Eurostat, Avrupa Birliği'nin istatistik ofisidir ve misyonu Avrupa’da yüksek kaliteli veri ve istatistik sağlamaktır (“Eurostat”, t.y.b). Avrupa'nın çeşitli ülkelerinden veri toplar, analiz eder ve yayımlar. Eurostat, ulusal istatistik kurumları tarafından iletilen ham veriler üzerinde kontroller ve doğrulama denetimleri gerçekleştirir (“Eurostat”, t.y.a).

3.1.3.6. İsviçre Federal Veri Arşivi

Federal İstatistik Ofisi (FSO) bünyesindeki Açık Devlet Verileri Ofisi tarafından yönetilen İsviçre Federal Veri Arşivi, İsviçre açık devlet verileri için merkezi portalıdır. Kamu yararına olması durumunda belediyeler vb. yarı devlet kuruluşları veya özel paydaşların verilerinin kaydını da tutmaktadır. Amacı kamu verilerinin kullanımını teşvik eden bir ortam yaratmak ve İsviçre veri altyapısının önemli bir parçasının oluşturulmasını desteklemektir.

Portal veri saklamamakta, sadece veri setlerine referans vermektedir. Veri seti sahipleri tarafından saklanır. FSO kurumlara açık verilerini yayınlaması konusunda destek olur, ayrıca kaliteyi sürekli olarak takip eder (“Opendata.swiss”, t.y.).

3.1.3.7. DataShop

Carnegie Mellon Üniversitesi, The Pittsburgh Science of Learning Center bünyesinde yürütülen DataShop, öğrencilerin eğitim yazılımlarıyla etkileşimi sonucunda oluşan verileri içeren, dünya çapında önemli bir veri deposu olmakla birlikte web tabanlı bir arayüz üzerinden analiz ve görselleştirme araçları sunmaktadır. Çevrimiçi dersler ve öğrenme-bilimi çalışmalarına ilişkin verilerle birlikte zekâ destekli eğitim sistemleri ve sanal laboratuvarlar gibi etkileşimli öğrenme ortamlarından gelen tıklama verilerini, ayrıca yayın, dosya, sunum vb. dokümanları da depolayabilmektedir. Veri DataShop veri tabanına kaydedildikten sonra, web uygulaması veri analizi için destek olmaktadır. Ayrıca veriyi dışa aktararak farklı işleme ve analiz araçları kullanılabilir. DataShop'a ücretsiz üyelik yaparak erişim sağlanabilmektedir ("DataShop@CMU", t.y.).

3.1.3.8. Dryad Digital Repository

Dryad araştırma verilerinin açık erişilebilirliğine ve yeniden kullanımına odaklanmış bir veri platformudur. Nihai amacı bilimsel keşiflerin hızlanması için tüm araştırma verilerinin açık olarak paylaşılmasıdır. Araştırma verilerinin korunmasında şeffaflık, kapsayıcılık, açıklık ve güvenilirliği ön planda tutar. Araştırma kurumları, fon sağlayıcıları, yayıncılar vb. paydaşlardan oluşan geniş bir yapıdır ("Dryad", t.y.).

3.1.3.9. Ulusal Akıl Sağlığı Enstitüsü Veri Arşivi (The National Institute of Mental Health Data Archive - NDA)

Amerika Birleşik Devletleri Ulusal Akıl Sağlığı Enstitüsü Veri Arşivi (The National Institute of Mental Health (NIMH) Data Archive - NDA), farklı alanlardaki araştırma projelerinden toplanan ve ortak bir standarda uyumlu hale getirilmiş insan denek verilerini yetkili araştırmacıların kullanımına sunmaktadır, özet verileri ise isteyen

herkes kullanabilmektedir. Araştırma verileri ile birlikte, farklı araç, yöntem ve analizleri paylaşmak ve işbirlikçi bilim ve keşfi mümkün kılmak için altyapı sağlamaktadır. NDA verileri ortak bir standartta sunarak ve araştırma sonuç raporları ile bilimsel araştırma ve keşfi hızlandırmayı hedeflemektedir (“NDA”, t.y.).

3.1.3.10. Ulusal Atmosferik Araştırma Merkezi (National Center for Atmospheric Research) Veri Arşivi

Amerika Birleşik Devletleri Ulusal Atmosferik Araştırma Merkezi (National Center for Atmospheric Research - NCAR) Hesaplamalı Sistemler ve Bilgi Sistemleri Laboratuvarı (Computational and Information Systems Laboratory) tarafından yönetilmektedir.

Meteorolojik ve oşinografik gözlemler, operasyonel ve yeniden analiz modeli çıktılarından oluşan geniş bir koleksiyonu bulunmaktadır. “Core Trust Seal” sertifikalı bir veri deposudur (“NCAR Research Data Archive”, t.y.).

3.1.3.11. Avrupa Veri Portalı (European Data Portal)

Avrupa Veri Portalı (European Data Portal), Avrupa Birliği, uluslararası, ulusal, bölgesel ve yerel veri portallarındaki verilere erişim sunan merkezi portaldır. Vatandaşlar, iş dünyası ve kurumların açık verilere erişimini sağlamak ve verilerin yeniden kullanılmasını teşvik etmeyi, Avrupa ülkeleri ve Avrupa Birliği kurumlarının daha kaliteli veri ve üst veri paylaşmasına destek olmayı ayrıca vatandaşları ve kurumları açık verinin yaratacağı fırsatlar konusunda eğitmeyi hedeflemektedir (“Documentation of data.europa.eu”, t.y.).

3.1.3.12. Üniversiteler Arası Politik ve Sosyal Araştırmalar Konsorsiyumu (Inter-university Consortium for Political and Social Research - ICPSR)

ABD'deki Michigan Üniversitesi'nde bulunan ICPSR 1962 yılında kurulmuştur ve sosyal bilimler alanında geniş kapsamlı bir veri arşivi sunmaktadır. Birçok devlet kurumu ve özel vakıf ICPSR'ye maddi destek sağlamaktadır ("ICPSR", t.y.a). ICPSR verilerin yeniden kullanılabilmesi ve araştırma bulgularının doğrulanabilmesini sağlamak amacıyla verileri depolar ve erişime sunar. Ancak verileri erişime sunmadan önce iyileştirerek verinin değerini ve veriye erişimi artırmayı, aynı zamanda verinin daha uzun vadeli korunmasını sağlamayı amaçlar. Uygulanan veri iyileştirmesi verinin düzenlenmesi, tanımlanması, temizlenmesi, geliştirilmesi ve korunmasını içerir ("ICPSR", t.y.b).

3.1.3.13. Çocuk ve Aile Veri Arşivi (Child & Family Data Archive - CFData)

Çocuk ve Aile Veri Arşivi (Child & Family Data Archive - CFData), ICPSR'ın bir projesidir. ICPSR, CFData'ya teknik destek ve finansman sağlamaktadır. Küçük çocuklar, aileleri, yaşadıkları çevre ve onlara hizmet eden programlara ait verileri saklayan ve kullanıma sunan CFData 2019'da kurulmuştur. Yüzlerce veri seti ve ilgili dokümanı bünyesinde barındırmaktadır. Araştırmacı, eğitimci ve karar alıcı yetkililerin verileri bulmasına yardımcı olan web tabanlı araçlar sunmaktadır, böylece verilerin keşfedilmesi ve kullanılmasını desteklemektedir. Veri kümelerinin kullanımı ile ilgili olarak düzenli eğitimler düzenlemektedir ("Child & Family Data Archive", t.y.).

3.1.3.14. Deniz Veri Arşivi (Marine Data Archive - MDA)

Deniz Veri Arşivi (Marine Data Archive - MDA) bireylere veya enstitülere veri dosyalarını yöneterek, kişisel veya kurumsal bir arşiv/yedekleme sistemi olarak

ayrıca veri yayımlanması için bir depo olarak hizmet vermektedir (“Marine Data Archive”, t.y.).

3.1.3.15. Dataverse Projesi

Dataverse Projesi, araştırma verilerinin paylaşılması, korunması, keşfedilmesi, analiz edilmesine olanak sağlayan, böylece veriye ulaşmayı kolaylaştıran açık kaynaklı bir web uygulamasıdır. Dataverse deposu farklı Dataverse koleksiyonlarından oluşan yazılım kurulumudur. Her Dataverse koleksiyonunun içinde verilerle birlikte veriye ait üst veri ve veri dosyaları (dokümantasyon ilgili kod) da bulunmaktadır. Harvard Üniversitesi tarafından dünya çapında birçok ortakla geliştirilen Dataverse Projesi’nin temel hedefi arşivcilerin yaptığı çoğu işi otomatikleştirmek ve veri üreticileri için servis sağlamaktır. Dataverse ile araştırmacıların verilerini uzun vadeli korumayı garanti edemeden kendilerinin saklaması veya uzun süreli korumayı garanti altına almak için profesyonel arşivlere göndermesine bir alternatif oluşmuştur. Araştırmacının kendi web sitesine Dataverse koleksiyondan araştırma verisine erişilebilmesi için link eklenebilmektedir. Yine Dataverse üzerinden veriye ulaşan araştırmacılara da verilerin ait olduğu çalışmaya ait bilgi (alıntı) verilmektedir. Dataverse Projesi’nin güncel hedefleri arasında sunduğu hizmetleri genişletmek, birlikte çalışabilirliği artırmak, kullanıcı deneyimini iyileştirmek ve hassas veri ile büyük veri işleme becerisi geliştirmek vb. yer almaktadır (“Dataverse Project”, t.y.).

3.1.3.16. Zenodo

Açık Bilim'e herkesin katılabilmesini sağlamak amacıyla OpenAIRE projesi tarafından Avrupa Komisyonu (European Commission) destekli araştırma verilerinin depolanması için oluşturulmuştur. OpenAIRE ortağı ve açık kaynak, açık erişim ve açık veride öncü olan CERN tarafından 2013 yılında, büyük veri yönetimi için araçlar ve açık veri için dijital kitaplık geliştirilmiştir. Zenodo ile bu

araçların etkili bir biçimde paylaşılması sağlanmıştır. Zenodo'nun sunduğu tümünü yakalama (catch-all) deposu bütün alanlardaki araştırmacılar için bir ihtiyaç olduğundan dünyanın her yerinden ve her disiplinden araştırmacılar tarafından kullanılmaya başlanmıştır. Dijital devrim verilerin ve yazılımların ele alınmasında akademik süreçlerin geliştirilmesini gerektirmiştir. Bir araştırmanın tam olarak anlaşılması için bütün detaylarını bilmek gerekir ve Zenodo bütün bu detayların paylaşılmasına olanak sağlamaktadır. Ayrıca bir araştırma henüz yayınlanmadan araştırma ile ilgili materyalleri Zenodo'ya yüklemek ve korumalı olarak bu bağlantıları hakemlerle paylaşmak mümkündür. Zenodo araştırmaların atıf almasına yardım etmektedir ve yayınların atıf bilgisi de bilimsel toplayıcılara iletilmektedir ("Zenodo", t.y.).

Zenodo, OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting, Açık Arşivler Girişimi Üst Verisi Hasatlama Protokolü) protokolünü desteklemekte olup Dublin Core, DataCite ve MARCXML gibi uluslararası üst veri şemalarıyla uyumludur. Bu özellik, platformdaki materyallerin harmanlanmasını ve diğer sistemlerle entegrasyonunu mümkün kılmaktadır. Ayrıca, RESTful API desteği sayesinde toplu yükleme ve veri madenciliği gibi süreçleri kolaylaştırarak araştırmacılara geniş bir kullanım alanı sunmaktadır. Bu yazılımın GitHub'da barındırılan versiyonu, araştırmacıların platformun özelliklerini özelleştirme veya genişletme ihtiyaçlarına yönelik açık bir altyapı sunmaktadır (CERN, t.y.).

3.1.3.17. Veri Arşivleme ve Ağ Hizmetleri (Data Archiving and Networked Services - DANS)

DANS Hollanda ulusal araştırma veri deposudur. Araştırma verilerinin yeni araştırmalarda kullanılmasına ayrıca yayınlanmış araştırmaların doğrulanmasına olanak tanır. DANS bünyesinde 200.000'den fazla veri seti bulunmaktadır. DANS gelecekte Hollanda veri ortamının daha başarılı olması için gerekli ihtiyaçları belirleyerek, etkili bir şekilde katkı sunmayı hedeflemektedir ("Data Archiving and Networked Services", t.y.).

3.1.3.18. Figshare

Figshare, kullanıcıların makaleler, veri setleri ve araştırma sonuçları gibi tüm araştırma çıktılarını alıntılanabilir, paylaşılabilir ve keşfedilebilir bir şekilde sunabilecekleri bir depodur. 2012 yılında piyasaya sürülmüştür (“Figshare”, t.y.). Platform, kullanıcılara dosya biçimi açısından geniş bir yelpaze sunarken her veri kaydı için DOI atayarak içeriğin kalıcı olarak tanımlanmasını ve sürüm takibini mümkün hâle getirir. Bu özellikleri ile Figshare, özellikle araştırma bütünlüğü ve tekrarlanabilirliği açısından önem taşıyan sürüm yönetimi ve arşivlemeyi desteklemektedir.

Platformun üst veri yönetimi, Dublin Core ve DataCite gibi yaygın şemalara uyumlu alanlar sunarak araştırma çıktılarını daha bulunabilir ve yeniden kullanılabilir hâle getirmektedir. Figshare ayrıca API düzeyinde entegrasyon ve otomasyon imkânları sağlayarak toplu veri yükleme ve düzenlemeyi kolaylaştırmaktadır. Kullanıcıların görüntüleme ve indirme istatistiklerine erişebilmesi, akademik çıktıların etki düzeyini göstermekte ve alternatif atıf ölçümlerine katkı sunmaktadır. OAI-PMH gibi protokollerle de uyumlu biçimde çalışabilmesi, verilerin farklı dijital arşivler arasında standartlara dayalı olarak paylaşılmasına zemin hazırlamaktadır (Open Archives Initiative, t.y.).

3.1.3.19. FinData

Finlandiya'da kurulan FinData adlı ulusal veri arşivi, sağlık verilerini güvenli bir ortamda toplamakta, araştırma ve yenilikçilik için yeniden kullanılabilir hale getirmektedir. Bu veriler, ilaç geliştirme, sağlık teknolojileri ve kişisel sağlık çözümleri gibi çeşitli alanlarda inovasyonu desteklemektedir. Bu veri arşivi sağlık sisteminin maliyetlerini düşürmek ve hizmet kalitesini artırmak için kullanılmaktadır. FinData, 1 Mayıs 2019'da yürürlüğe giren Sağlık ve Sosyal Verilerin İkincil Kullanımı Yasası (Act on the Secondary Use of Health and Social Data) kapsamında kurulmuştur. Bu yasa, ulusal sağlık ve sosyal hizmet

kayıtlarının yanı sıra birincil bakım, uzman sağlık hizmetleri ve sosyal hizmetlerdeki operasyonel müşteri ve hasta sistemlerinden gelen verilere erişim sağlamaktadır. FinData araştırma, geliştirme ve yenilikler için izin, danışmanlık ve veri erişimi sunan tek nokta olarak hizmet vermektedir (“Finnish Social and Health Data Permit Authority Findata”, t.y.).

Yukarıda açıklanan veri arşivleri, günümüzün karmaşık ve geniş çaplı veri yönetimi gereksinimlerini karşılamak için tasarlanmış çeşitli yaklaşımları ve uygulamaları temsil etmektedir. Her biri, veri toplama, saklama, yönetim ve paylaşım süreçlerinde belirli alanlara odaklanarak önemli katkılar sunmaktadır.

Öncelikle, bu arşivlerin temel özelliklerinden biri, veri çeşitliliği ve çok kaynaklılığa odaklanmalarıdır. Sosyal bilimlerden sağlık verilerine, atmosferik gözlemlerden öğrenme süreçlerine kadar farklı disiplinlerdeki veriler bir araya getirilerek kullanıcılara sunulmaktadır. Bu çeşitlilik, farklı veri türlerinin (yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış) yönetimini ve analizini mümkün kılmaktadır.

Veri arşivleri, aynı zamanda standartlaşma ve uluslararası iş birliği süreçlerini destekleyen önemli platformlardır. Örneğin ADA, uluslararası standartlar olan DDI ve OAIS çerçevesinde çalışarak, verilerin uzun vadeli erişilebilirliğini ve sürdürülebilirliğini sağlamakta, aynı zamanda küresel iş birliği projelerine katkıda bulunmaktadır. Benzer şekilde, Zenodo ve Figshare, açık bilim ve açık veri ilkelerini benimseyerek disiplinler arası veri paylaşımını kolaylaştırmaktadır. Bu standartlaşma süreçleri, verilerin yeniden kullanılabilirliğini artırmakta ve araştırmaların doğrulanmasını desteklemektedir.

Bu arşivler, ayrıca üst veri yönetimi ve kullanıcı odaklı veri erişimi açısından dikkat çekici çözümler sunmaktadır. Çoğu veri arşivi, veri setlerine eklenen üst veri ile verilerin keşfedilebilirliğini ve analiz edilebilirliğini artırmaktadır. Örneğin Dataverse Projesi, kullanıcıların verilerini organize edip paylaşmalarına olanak tanırken, veri üreticileri ve kullanıcıları arasında bir köprü oluşturmaktadır. DataShop gibi platformlar ise kullanıcılarına görselleştirme ve analiz araçları

sunarak, verilerin yalnızca erişilebilir değil, aynı zamanda kullanılabilir olmasını sağlamaktadır.

Veri güvenliği ve sürdürülebilirlik, bu arşivlerin ortak özelliklerinden biridir. Core Trust Seal sertifikasına sahip olan arşivler, güvenilir veri yönetimi uygulamalarıyla dikkat çekmektedir. Örneğin Dryad ve NCAR Veri Arşivi, verilerin uzun vadeli korunmasını ve erişilebilirliğini sağlarken, veri güvenliğini arttırmaktadır. Bu, özellikle kişisel ve hassas verilerin işlendiği sağlık ve sosyal veri arşivleri için kritik öneme sahiptir.

Son olarak, bu veri arşivleri, disiplinler arası veri paylaşımı ve iş birliğini teşvik eden kapsayıcı modeller sunmaktadır. European Data Portal ve ICPSR gibi platformlar, ulusal ve uluslararası düzeyde veri paylaşımını artırmakta, farklı disiplinlerden araştırmacılara geniş bir veri yelpazesi sunmaktadır. Bu tür platformlar, yalnızca verileri toplamakla kalmayıp, aynı zamanda verilerin toplum yararına kullanılmasını ve karar alma süreçlerini desteklemesini sağlamaktadır.

Tüm bu özellikler, bu veri arşivlerinin modern veri yönetimi sistemlerinin temel bileşenlerini nasıl etkili bir şekilde birleştirdiğini göstermektedir. Standartlaşma, üst veri kullanımı, güvenlik ve açık erişim politikalarının benimsenmesi, bu arşivlerin hem bireysel hem de kurumsal düzeyde sürdürülebilir, erişilebilir ve iş birliğine dayalı bir veri ekosistemi oluşturmasını sağlamaktadır. Türkiye'de ulusal bir veri arşivi tasarlanırken, bu uluslararası örneklerden ilham alınarak, disiplinler arası uyumlu ve kullanıcı dostu bir model geliştirilebilir. Bu model, farklı veri türlerini destekleyen, uzun vadeli erişim sağlayan ve ulusal düzeyde verilerin etkin kullanımını teşvik eden bir yapı sunabilir.

3.2. İLGİLİ MEVZUAT VE UYGULAMALAR

3.2.1. Türkiye'deki Mevzuat ve Standartlar

Bu bölümde, Türkiye'deki kurumsal belge ve dijital arşiv yönetimi uygulamalarını şekillendiren yasal çerçeve ve standartlar incelenecektir.

3.2.1.1. Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (9/6/2020 Tarih ve 2646 Sayılı)

İlgili yönetmelik madde 1'de yönetmeliğin amacı "Güvenli elektronik imza kullanılarak elektronik ortamda veya el yazısıyla atılan imza ile fiziksel ortamda yapılan resmi yazışmalara ilişkin kuralları belirlemek, bilgi veya belge alışverişini, hızlı ve güvenli bir biçimde yürütmek ve uygulama birliğini sağlamaktır." olarak ifade edilmiştir. Bu yönetmeliğin tüm kamu kurum ve kuruluşlarını kapsadığı belirtilmiştir. Ayrıca zorunlu haller veya olağanüstü durumlar dışında, kamu kurum ve kuruluşlarınca resmi yazışmaların, e-Yazışma Teknik Rehberi'ne uygun olacak şekilde, elektronik ortamda hazırlanıp imzalanan belgelerle yapılacağı, elektronik imza ile imzalanan belgelerin ayrıca çıktısının alınarak imzalanmayacağı ve fiziksel ortamda saklanmayacağı belirtilmiştir (Resmi Yazışmalarda Uygulanacak Usul, 2020).

Yönetmelikte belge özellikleri; oluşturulacak nüsha sayısı, belgenin şekli özellikleri ve belgenin yazı tipi ve harf büyüklüğünü içerecek şekilde detaylandırılmıştır. Belgenin bölümleri ise yazı alanı, kurumsal logo, başlık, sayı, tarih, konu, muhatap, ilgi, metin, imza, ek, dağıtım, olur, paraf, koordinasyon, belge doğrulama bilgileri, iletişim bilgileri, gizlilik dereceli belgeler, süreli ve kişiye özel yazışmalar, sayfa numarası ve üst veri elemanları başlıkları altında detaylandırılmıştır. Belgenin çoğaltılması, elektronik ve fiziksel ortamda gönderilmesi, alınması ve iade edilmesi durum ve şartları açıklanmıştır. Görüş, bilgi ve belge talep yazılarının yerine getirilmesi gereken süre açıklanmıştır. Ayrıca bu yönetmeliğe uygun olarak hazırlanmayan belgeler için uygulanacak

işlem hakkında bilgi verilmiştir. 9/6/2020 tarihli ve 2646 sayılı Cumhurbaşkanı kararının yürürlüğe girmesiyle, “15/12/2014 tarihli ve 2014/7074 sayılı Bakanlar Kurulu Kararı ile yürürlüğe konulan Resmî Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik yürürlükten kaldırılmıştır.” (Resmi Yazışmalarda Uygulanacak Usul, 2020).

3.2.1.2. TSE K 523: Bilgi Varlıklarının Gizlilik Derecelerine Göre Sınıflandırılması Kriteri

TSE K 523, 2016 yılında yayımlanan ve kişisel ile kurumsal veri, bilgi ve belgelerin gizlilik gereksinimlerine uygun olarak sınıflandırılmasını amaçlayan bir standarttır. Bu standart, bilgi varlıklarının gizliliğini koruma ihtiyacından doğmuştur ve bilgi güvenliği yönetimi alanında önemli bir kaynak olarak kabul edilmektedir. Hem elektronik ortamlarda hem de taşınabilir teknolojiler aracılığıyla erişilebilen veri ve belgelerin, paylaşım ve güvenlik risklerine karşı korunması hedeflenmektedir.

Standardın temel amacı, organizasyonların sahip olduğu bilgi varlıklarının tespit edilmesi ve bu varlıkların gizlilik derecelerine göre sınıflandırılmasıdır. TSE K 523, bilgi varlıklarını fiziksel/elektronik bilgi varlıkları, yazılım, insan, soyut değerler (itibar, imaj vb.), hizmetler ve projeler olmak üzere altı başlık altında ele almaktadır. Bu varlıklar, gizlilik seviyelerine göre dokuz kategoriye ayrılmıştır: “Çok Gizli”, “Gizli”, “Özel”, “Hizmete Özel”, “Ticari Gizli”, “Ticari Özel”, “Kişiye Gizli”, “Kişiye Özel” ve “Tasnif Dışı”.

Bu gizlilik seviyeleri, verilerin nasıl korunacağına dair rehberlik sağlamakta ve veri güvenliğini sağlamak amacıyla kurum içindeki erişim haklarını düzenlemektedir. Standart, kişisel ve kurumsal bilgilerin, gizlilik ihlallerini önleyici bir yaklaşımla ele alınmasını ve güvenlik risklerinin minimize edilmesini hedeflemektedir (TSE K 523, 2016).

3.2.1.3. TS 13298 Elektronik Belge Yönetimi Standardı

Türkiye'de elektronik belge yönetimi alanında önemli bir yere sahip olan TS 13298 Elektronik Belge Yönetimi Standardı, kurumların EBYS uygulamalarını şekillendiren temel bir rehber niteliğindedir. Bu standart, EBYS'lerin tasarımı, uygulanması ve yönetilmesi için gerekli kriterleri belirlemektedir. TS 13298 standardı sistem ve belge kriterlerine ek olarak üst veri elemanlarını içeren üç ana bölümden oluşmaktadır. TS 13298 standardı:

- EBYS'lerin Devlet Arşivleri Genel Müdürlüğü'nün tanımladığı standart dosya planları ile uyumlu olası ve dosya tasnif planlarını içermesini,
- Kamu kurum ve kuruluşlarında yeni birimlerin kurulması vb. Durumlara entegre edilebilir olmasını,
- EBYS'lerin interneti çok az kullanan kişiler tarafından kullanılabilecek ve kullanıcı hatalarına engel olacak şekilde tasarlanmasını,
- Elektronik belgelere yapılacak işlemlerin kaydının alınmasını ve bu kayıtların muhafaza edilmesini, böylece usulsüzlüklerin önüne geçilmesinin sağlanmasını,
- Elektronik belgelerin imha şartlarının belirlenmesini, böylece belge kaybının engellenmesini ayrıca belgelerin çapraz referanslarla bağlanarak belgelerin kopyalarının oluşturulmasına engel olunmasını,
- EBYS'lerin belirli periyotlarla kullanım raporlarını oluşturabilecek yeteneğe sahip olmasını,
- Bir klasörde yer alan belge sayısının dosya muhteviyat raporu vb. ile takip edilebilmesini,
- Sistemlerde belgelerin saklama planlarının oluşturulmasını, tasfiye vb. işlemlerin otomatik yapılabilmesini ve bu sürecin raporlanabilmesini,

- Her türlü elektronik belgenin üst verisi ve ekleri ile birlikte sisteme dahil edilebilmesini, standart formatlarda hazırlanan belgelerin, belgelerin üretildiği program olmadan görüntülenebilmesini, makro vb. içermesi nedeniyle güncellenmeye açık belgelerin içeriğinin ilk haliyle korunmasını,
- Uygun belgelerde tam metin arama yapılabilmesini, en son yapılan aramaya ait ölçütlerin kaydedilebilmesini, kullanıcı işlemlerinin raporlanabilmesini,
- Kullanıcıların tanımlı rol ve yetkileri dahilinde sistemi kullanabilmesini, sistemde işlem yapan kullanıcıların kayıtlarının korunmasını, ayrıca sistem içerisinde kullanıcılar için çevrimiçi yardım özelliğinin bulunmasını,
- Sistemde e-imza kullanılarak e-imzanın sunduğu kimlik doğrulama vb. özellikleri kullanarak güvenlikle ilgili problemlerin engellenmesini,
- Elektronik belgelere ait belgenin üretildiği program, oluşturulduğu işletim sistemi vb. bilgilerin kayıt altına alınarak uzun yıllar sonra belgelere ulaşabilmenin mümkün olmasını,
- Belgelere ek olarak elektronik ortamda üretilmiş diğer dokümanların da sistemde içeriklerinin taranmasını ve iş akışlarında kullanılabilmesini,
- Elektronik olmayan dosya ve belgelerin tanımlama ve yer bilgilerinin sistemde tutulmasını ve böylece fiziksel belgeye erişimin kolaylaşmasını,
- Sisteme sayısal görüntüleme ile aktarılan fiziksel belgelerin metinleştirilerek doküman olarak veya görüntü olarak kaydedilmesini,
- Sistemin üst veri için bir sınırlama getirmemesini ve üst verilerin standarda uygun olmasını istemektedir (Yılmaz, 2012, s. 2-5).

3.2.1.4. Devlet Arşiv Hizmetleri Hakkında Yönetmelik (DAHHY)

İlgili yönetmelik madde 1'de yönetmeliğin amacı "kamu kurum ve kuruluşlarının iş ve işlemleri sonucunda oluşan belgelerin; düzenlenmesine, gerekli şartlar altında korunmalarının teminine, herhangi bir sebepten dolayı kaybının engellenmesine, Devletin, gerçek ve tüzel kişilerin ve bilimin hizmetinde değerlendirilmelerine, kurum ve kuruluşlar ile şahıslar elinde bulunan arşiv belgeleri ve ileride arşiv belgesi haline gelecek arşivlik belgelerin tespit edilmesine, saklanmasına gerek görülmeyen belgelerin ayıklanmasına, imhasına ve arşiv belgelerinin Devlet Arşivleri Başkanlığına devrine ilişkin usul ve esasları düzenlemektir." olarak belirtilmiştir. Madde 3'te ise bu yönetmeliğin "11 sayılı Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi" hükümlerine dayanılarak hazırlandığı belirtilmiştir (Devlet Arşiv Hizmetleri Hakkında Yönetmelik, 2019, s. 1). Yönetmeliğin ikinci bölümünde belgelerin korunması, belge yöneticileri ve arşiv personeli, belgelerin gizliliği ve gizliliğin kaldırılması, arşivlerden yararlanma, arşivlerin oluşturulması; üçüncü bölümünde belgelere dosya kodu verilmesi, belgelerin dosyalanması ve dosya etiketi gibi dosyalama işlemleri; dördüncü bölümünde arşivlere devredilecek belgelerin hazırlanması, belgelerin arşivlere devri, arşivlerde düzenleme, Başkanlığa devredilecek arşiv belgelerinin hazırlanması ve teslimi başlıkları altında belge devir işlemleri, beşinci bölümünde kurum arşivi ve merkezi arşivlerde ayıklama ve imha, ayıklama işlemine tâbi tutulmayacak belgeler, ayıklama ve imha komisyonlarının oluşturulması, ayıklama ve imha komisyonlarının çalışma esasları, imha listelerinin düzenlenmesi ve kesinlik kazanması, imha şekilleri başlıkları altında ayıklama ve imha işlemleri; altıncı bölümünde ise belgelerin dijitalleştirilmesi, dosya planları ve saklama planları, denetim, belge yönetimi ve arşiv hizmetleri faaliyet raporları, özel arşivlerin satın alınması, özel arşivlerin yurt dışına izinsiz çıkarılamayacağı, tereddütlerin giderilmesi ve iş birliği, ilga edilen, yetkisi devredilen ve özelleşen kurum ve kuruluşların belgeleri, müsadere olunan belge vb. çeşitli, geçici ve son hükümlere yer verilmiştir. Altıncı Bölüm Madde 24'te dosyalama işlemlerinde "Standart Dosya Planı"nın uygulanmasının zorunlu olduğu ifade edilmiştir (Devlet Arşiv Hizmetleri Hakkında Yönetmelik, 2019, s. 3-11).

3.2.1.5. Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi (Kararname Numarası: 11) (DABHCK)

Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi, Türkiye'nin arşivcilik sistemini geliştirme ve modernleştirme yolunda önemli bir adımı temsil etmektedir. 13 Temmuz 2018 tarihli ve 30473 sayılı Resmî Gazete'de yayımlanan 11 numaralı Cumhurbaşkanlığı Kararnamesi ile Devlet Arşivleri Başkanlığı'nın görev ve yetki alanları tanımlanarak devlet arşivlerinin daha etkin bir şekilde yönetilmesi sağlanmıştır (Resmî Gazete, 2018a). İlgili yönetmelik madde 1'de kararnamenin amacı, "Devlet Arşivleri Başkanlığının kuruluş, görev ve yetkilerine ilişkin usul ve esasları düzenlemek" olarak belirtilmiştir (Devlet Arşivleri Başkanlığı Hakkında, 2018, s. 1).

Kararnamede, Devlet Arşivleri Başkanlığının görevleri, Başkan'ın sorumlulukları, "Cumhurbaşkanlığı Arşivi Dairesi Başkanlığı, Belge Tespit ve Değerlendirme Dairesi Başkanlığı, Tasnif ve Araştırma Hizmetleri Dairesi Başkanlığı, Bilgi İşlem ve Elektronik Arşiv Dairesi Başkanlığı, Muhafaza ve Bakım Dairesi Başkanlığı, Personel ve Eğitim Dairesi Başkanlığı, Destek Hizmetleri Dairesi Başkanlığı, Dış İlişkiler ve Tanıtım Dairesi Başkanlığı, Strateji Geliştirme Dairesi Başkanlığı, Özel Kalem Müdürlüğü" başlıkları altında başkanlığın hizmet birimlerinin görev ve yetkileri, bu Kararname kapsamına giren kurum ve kuruluşların elindeki arşiv belgesi ve arşivlik belge ile ilgili yükümlülükleri, yöneticilerin sorumlulukları, personel istihdamı, çalıştırılması ve görevlendirme, kadrolar vb. konularda maddeler bulunmaktadır (Devlet Arşivleri Başkanlığı Hakkında, 2018).

Bu kararname ile Devlet Arşivleri Başkanlığı'nın yapılanmasında dijital arşivleme, elektronik belge yönetimi ve veri güvenliği gibi günümüz arşivcilik dünyasında önem kazanan uygulamalara yer verilmiştir. Bu yeni yapı ve uygulama esasları ile Türkiye'de arşivcilik, daha güvenilir ve ulaşılabilir bir sistem anlayışını benimsemektedir.

3.2.1.6. Standart Dosya Planı ile İlgili 2005/7 Sayılı Başbakanlık Genelgesi (SDP 2005/7)

Başbakanlık Devlet Arşivleri Genel Müdürlüğü koordinasyonunda başlamış olan "Standart Dosya Planı" çalışmaları ile kamu kurum ve kuruluşlarının oluşturduğu belgelere ihtiyaç duyulması halinde hızlı erişim sağlanabilmesi amacıyla belgelerin dosyalanması aşamasının belirlenen genel yönteme uygun olarak yapılması hedeflenmiş, bu amaçla kamu kurum ve kuruluşlarının ana hizmet birimlerine ilişkin dosya planı hazırlamaları, ayrıca tüm kamu kurum ve kuruluşlarının ortaklaşa kullanacağı yardımcı hizmet, danışma ve denetim birimlerine ait belgeler için ise Başbakanlık Devlet Arşivleri Genel Müdürlüğü tarafından hazırlanan ortak alanlara ait dosya planını kullanmaları talimatlandırılmıştır (Standart Dosya Planı, 2005).

3.2.1.7. Veri Paylaşımı Kurulu Yönetmeliği (VPKY)

İlgili yönetmelik madde 1'de yönetmeliğin amacı, "Genel Müdürlük (Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü) merkezi veri tabanında tutulan kayıtlardan yararlanacak Genel Müdürlük birimlerini, kurumları, kamu hizmeti sunan tüzel kişilikleri ve adrese dayalı kamu hizmeti sunan kuruluşları tespit etmek, çevrimiçi ve çevrimdışı paylaşımın kapsamı, yöntemi ve güvenliğine ilişkin genel usul ve esasları saptamak ve Genel Müdürlük bünyesinde oluşturulan Veri Paylaşımı Kurulunun çalışma usul ve esaslarını belirlemektir." olarak verilmiştir (Veri Paylaşımı Kurulu Yönetmeliği, 2020). Yönetmelik, Genel Müdürlük merkezi veri tabanında tutulan kayıtların güvenli, düzenli olarak yetkili kurumlar arasında paylaşımını sağlamak amacıyla çevrimiçi ve çevrimdışı veri paylaşım prosedürlerini detaylandırmaktadır. Bu kapsamda veri paylaşımı süreçlerinde güvenlik önlemleri ve izlenecek yöntemler belirlenmiş böylece kamusal veri yönetiminde bütünlük ve gizlilik esasları gözetilmiştir.

3.2.1.8. Kişisel Verilerin Korunması Kanunu (KVKK) (No: 6698)

6698 sayılı Kişisel Verilerin Korunması Kanunu (KVKK), Türkiye'de kişisel verilerin korunması ve işlenmesi konusunda önemli bir yasal çerçeve oluşturmaktadır. 7 Nisan 2016 tarihinde Resmî Gazete'de yayımlanarak yürürlüğe giren bu kanun, kişisel verilerin işlenmesinde bireylerin temel hak ve özgürlüklerini korumayı ve veri işleyen gerçek ve tüzel kişilerin yükümlülüklerini düzenlemeyi amaçlamaktadır. KVKK, kişisel verilerin işlenmesinde hukuka ve dürüstlük kurallarına uygunluk, doğruluk ve güncellik, belirli ve meşru amaçlar için işlenme, amaçla bağlantılı ve sınırlı olma gibi temel ilkeleri belirlemektedir. Kanun ayrıca veri sorumlularının yükümlülüklerini ve ilgili kişilerin haklarını da detaylı olarak düzenlemektedir (Kişisel Verilerin Korunması Kanunu, 2016).

Dijital arşivler açısından değerlendirildiğinde KVKK, kişisel verilerin nasıl yönetilmesi gerektiğine dair önemli yükümlülükler belirlemektedir. Bu bağlamda dijital arşivler, kişisel verileri işlerken kanunun belirlediği ilkelere uymak ve veri güvenliğini sağlamak zorundadır.

3.2.1.9. Bilgi ve İletişim Güvenliği Rehberi (BİGR)

Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi tarafından koordine edilen "Bilgi ve İletişim Güvenliği Rehberi", 6 Temmuz 2019 tarihli ve 2019/12 sayılı Cumhurbaşkanlığı Genelgesi doğrultusunda hazırlanmış ve 24 Temmuz 2020'de onaylanmıştır. Bu rehber, kamu kurumları ve kritik altyapı hizmeti sunan işletmelerin bilgi ve iletişim güvenliği tedbirlerini belirlemeyi amaçlamaktadır. Rehberin, gelişen teknoloji ve ulusal stratejiler doğrultusunda güncellenmeye devam edeceği belirtilmiştir. Tüm kamu kurumları ve kritik altyapı hizmeti veren işletmelerin, mevcut ve yeni bilgi sistemlerinde rehberde belirtilen tedbirlere uyması zorunlu tutulmaktadır. Bu rehber, yerli ve millî siber güvenlik ürünlerinin kullanımını teşvik ederek, Türkiye'nin siber güvenlik alanındaki uluslararası

rekabet gücünü artırmayı hedeflemektedir (*Cumhurbaşkanlığı Dijital Dönüşüm Ofisi,* 2020).

3.2.1.10. Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik

Türkiye Cumhuriyeti Cumhurbaşkanlığı tarafından hazırlanan Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik, gizlilik dereceli belgelerin tüm süreçlerini düzenlemek üzere 2023 yılında yürürlüğe girmiştir. Yönetmelik, ulusal güvenliğin korunmasını amaçlamakta olup gizlilik dereceli bilgi ve belgelerin yetkisiz erişimden korunması, doğru sınıflandırılması, saklanması, taşınması ve imha edilmesi gibi hususları detaylı bir şekilde ele almaktadır (T.C. Cumhurbaşkanlığı, 2023).

Bu kapsamda, belgelerin "çok gizli," "gizli," "özel" ve "hizmete özel" gibi derecelendirilmesine yönelik kurallar, erişim yetkilendirme süreçleri, yetkisiz erişim durumlarında uygulanacak yaptırımlar ve uluslararası standartlarla uyumlu güvenlik tedbirleri gibi konular yönetmelikte açıklanmıştır. Ayrıca, bu belgelerle ilgili yazışma ve işlem süreçleri de detaylı bir şekilde tanımlanmıştır.

Kamu kurum ve kuruluşlarının yanı sıra, gizlilik dereceli belgelerle çalışan özel sektör temsilcileri ve iş ortakları için de bağlayıcı olan yönetmelik, bu tür belgelerin ele alınmasında yüksek güvenlik standartlarının sağlanmasını hedeflemektedir (T.C. Cumhurbaşkanlığı, 2023).

3.2.1.11. E-Yazışma Teknik Rehberi

E-Yazışma Teknik Rehberi, Türkiye’de kamu kurum ve kuruluşları arasında gerçekleştirilen elektronik yazışmaların teknik ve idari esaslarını düzenlemek amacıyla hazırlanmıştır. Rehber, standart bir altyapı sunarak yazışma süreçlerinin güvenilir, hızlı ve maliyet etkin bir şekilde yürütülmesini sağlamayı hedeflemektedir. Bu bağlamda,

elektronik yazışmalarda kullanılan veri formatları, güvenlik protokolleri ve uyum süreçlerine dair detaylı teknik spesifikasyonlar içermektedir (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2019b).

Rehberin temel bileşenlerinden biri, e-Yazışmaların ulusal mevzuatlara ve uluslararası standartlara uygun olarak tasarlanmasını sağlamaktır. Elektronik belgelerin gönderiminde "Elektronik Belge Yönetim Sistemi (EBYS)"nin kullanımı zorunlu hale getirilmiş ve belgelere ilişkin güvenlik önlemleri, zaman damgası uygulamaları ve şifreleme yöntemleri gibi kritik detaylar ele alınmıştır. Ayrıca, rehberde kamu kurumları arasındaki uyumluluğu artırmak için belge formatlarına ve elektronik imza kullanımlarına dair standartlar belirlenmiştir (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2019b). E-Yazışma Teknik Rehberi, Türkiye'deki dijital dönüşüm süreçlerinin temel taşlarından biri olarak, kamu yönetiminde verimliliği artırmayı ve güvenliği sağlamayı amaçlamaktadır. Bu rehber, e-Devlet hizmetlerinin modernleşmesine önemli katkılar sunmaktadır.

3.2.2. Uluslararası Standartlar ve Düzenlemeler

Diğer ülkelerdeki kurumsal belge ve dijital arşiv yönetimi uygulamalarını yönlendiren yasal çerçeve ve standartlara yönelik örneklere bu bölümde yer verilmiştir.

3.2.2.1. ISO 14721: OAIS - Open Archival Information System – (Açık Arşivsel Bilgi Sistemi)

ISO 14721 (2012) dijital bilginin uzun vadeli korunmasını sağlamak amacıyla tasarlanmış bir referans modelidir (Lee, 2010, s. 4020). OAIS dijital koruma ve dijital küratörlük alanında birçok meslek grubu için temel bir referans noktası haline gelmiştir. Bu model, bilgi profesyonellerinin sadece içeriği anlamalarını değil, aynı zamanda içeriğin nasıl geliştirildiğini ve yayımlandığını anlamalarını da içerir (Lee, 2010, s. 4029).

Dijital korumada ilk uluslararası standart olarak ortaya çıkan referans modeli (Beagrie, 2003, s. 5), üst düzey standarda ihtiyaç duyulan bir dönemde geliştirilmiştir ve birçok standardizasyon, araştırma ve geliştirme çalışmasında önemli bir rol oynamaktadır. Model yeni terimleri tanıtmakla birlikte önceden var olan kaynakların da önemli ölçüde uyarlanarak yeniden kullanılmasına olanak tanır (Lee, 2010, s. 4029). OAIS'in 6 temel bileşeni bulunmaktadır. Bu temel bileşenler veri girişi (ingest), arşiv depolama (archival storage), koruma planlaması (preservation planning), erişim (access), veri yönetimi (data management) ve yönetimdir (administration) (Rahmanto ve Riasetiawan, 2018, s. 1). Bu bileşenler, bilginin zaman içinde güvenilir kalmasını ve gelecekteki kullanıcılar tarafından erişilebilir olmasını sağlar.

3.2.2.2. ISO/DIS 16363: Audit and Certification of Trustworthy Digital Repositories (Güvenilir Dijital Depoların Denetimi ve Sertifikasyonu)

ISO/DIS 16363 (2012), güvenilir dijital arşivlerin sertifikasyonu için kapsamlı bir denetim çerçevesidir. Standart, dijital verileri güvenli bir şekilde saklamak ve uzun süreli erişim sağlamak amacıyla dijital depoların değerlendirilmesi için kullanılır. Bu standartta "depo" terimi, yalnızca dijital verilerin depolandığı teknik altyapıyı değil, dijital korumadan sorumlu olan tüm organizasyonu ifade eder. ISO/DIS 16363, 100'den fazla ölçütü üç ana başlık altında toplar:

Organizasyonel Altyapı: Yönetişim, organizasyon yapısı, personel, mali sürdürülebilirlik, lisanslama, yasal sorumluluklar gibi unsurları içerir.

Dijital Nesne Yönetimi: Dijital verilerin alınması, doğruluğunun ve bütünlüğünün korunması ve erişim yönetimini kapsar. Uzun vadeli dijital koruma planlaması bu başlıkta önemli bir yer tutar.

Altyapı ve Güvenlik Risk Yönetimi: Teknik altyapı ve güvenlik risklerinin yönetilmesine odaklanır.

ISO/DIS 16363, OAIS (Açık Arşiv Bilgi Sistemi) modeli ile uyumlu olup, bu modelin kavram ve terimlerini kullanır. ISO 16919 ve ISO 17021 standartları ISO/DIS 16363 standardını destekler (“DPC Digital”. t.y.).

Veri depolarının, bu standartta belirtilen kriterlere uygun olarak oluşturulması dijital varlıkların uzun süreli erişim ve güvenliğinin sağlanması konusundaki yetkinliğini gösterir.

3.2.2.3. ISO/IEC 27001: Information Security, Cybersecurity and Privacy Protection - Information Security Management Systems – Requirements ve TS ISO/IEC 27001: Bilgi Güvenliği, Siber Güvenlik ve Kişisel Gizliliğin Korunması - Bilgi Güvenliği Yönetim Sistemleri - Gereklilikler

ISO/IEC 27001, bilgi güvenliği yönetimi için uluslararası alanda en çok kabul gören standartlardan biridir. Uluslararası Standardizasyon Örgütü (ISO) ve Uluslararası Elektroteknik Komisyonu (IEC) tarafından yayımlanan bu standart, kurumların Bilgi Güvenliği Yönetim Sistemi (BGYS) oluşturması, yönetmesi, sürdürmesi ve sürekli iyileştirmesi için gereklilikleri belirlemektedir. Standart, organizasyonların bilgi varlıklarının gizliliğini, bütünlüğünü ve erişilebilirliğini koruma amacıyla bilgi güvenliğine ilişkin riskleri yönetmesine yardımcı olmaktadır. Ayrıca, siber güvenlik ve gizlilik koruma konularını da kapsamaktadır (Advisera, t.y.).

ISO/IEC 27001, şirketlerin bilgi güvenliğini sistematik ve maliyet etkin bir şekilde sağlamaktadır. Bu standart, kurumsal risk analizleri yapılarak bilgi varlıklarını tehdit eden potansiyel güvenlik risklerine karşı önleyici ve düzenleyici tedbirlerin alınmasını zorunlu kılmaktadır. Bilgi güvenliği politikalarının oluşturulması, yalnızca yetkili kişilerin verilere erişebilmesi ve verileri değiştirebilmesi ilkelerine dayanmaktadır. Standart, organizasyonların yasal düzenlemelere uyum sağlamasına ve daha iyi bir organizasyon yapısı kurmasına yardımcı olmaktadır (ISO 27001, 2022).

Türkiye’de, ISO/IEC 27001’in karşılığı olan TS ISO/IEC 27001, Türk Standartları Enstitüsü (TSE) tarafından yayımlanmıştır. TS ISO/IEC 27001, Türkçe olması sayesinde Türkiye’deki kurumların uluslararası standarttan kolaylıkla faydalanmasını sağlamaktadır. Standart, kurumların bilgi güvenliği yönetim sistemlerini ulusal bir referans noktası olarak yapılandırmalarına olanak tanımakta ve bilgi güvenliği politikalarını oluştururken uyulması gereken kılavuz ilkeleri sunmaktadır (TS ISO/IEC 27001, 2023). Türkçe çeviri, bilgi güvenliği ihtiyaçlarının karşılanmasında ulusal erişilebilirlik ve uygulama kolaylığı sağlamaktadır.

3.2.2.4. ISO 8000: Data Quality (Veri Kalitesi)

ISO 8000 standardı, veri kalitesi için geliştirilmiş önemli bir uluslararası standarttır. Bu standart, veri kalitesini tanımlamak, ölçmek ve iyileştirmek için kapsamlı bir çerçeve sunmaktadır (ISO 8000-1, 2022). ISO 8000'in temel amacı, kuruluşların yüksek kaliteli veri üretmesini, yönetmesini ve paylaşmasını sağlamaktır. ISO 8000 standardı, veri kalitesinin çeşitli yönlerini kapsayan farklı bölümlerden oluşmaktadır. Örneğin Bölüm 1 Genel Bakış ve Temel İlkeler, Bölüm 2 Terminoloji, Bölüm 61 Veri Kalitesi Yönetimi, Bölüm 100-140 ise Ana Veri Kalitesi yönlerini kapsamaktadır.

Ana veri kalitesi bölümleri, semantik kodlama, veri doğruluğu, tamlığı ve kökeni gibi konuları ele almaktadır. Örneğin, ISO 8000-110 bölümü, kuruluşlar ve sistemler arasında ana veri değişimi için gereksinimleri belirlemektedir (Mohammed, Eram ve Talburt, 2017). ISO 8000-61, veri kalitesi yönetimi için bir süreç referans modeli sunmaktadır. Bu model, veri kalitesinin üç temel ilkesini vurgulamaktadır:

- Veri kalitesi süreçler aracılığıyla elde edilir.
- Sürekli iyileştirme gereklidir.

- Herkes veri kalitesi yönetiminde bir role sahiptir.

ISO 8000 standardı, kuruluşlara birçok fayda sağlamaktadır. ISO 8000 standardı, kuruluşların veri kalitesini sistematik bir şekilde yönetmelerine ve geliştirmelerine yardımcı olan kapsamlı bir çerçeve sunmaktadır. Bu standart, veri kalitesinin önemini vurgulayarak, kuruluşların daha iyi kararlar almasına ve operasyonel verimliliği artırılmasına katkıda bulunmaktadır (Benson, 2008).

3.2.2.5. ISO/IEC TR 10032:2003: Information Technology - Reference Model of Data Management (Bilgi Teknolojisi - Veri Yönetimi Referans Modeli)

ISO/IEC TR 10032:2003, ISO tarafından yayınlamış bir teknik rapordur ve veri yönetimi için kapsamlı bir referans model tanımlamaktadır. Bu standart, veri yönetimi için ortak bir temel ve terminoloji sağlarken, mevcut ve gelecekteki veri yönetimi standartlarının geliştirilmesini koordine etmek için bir çerçeve sunmaktadır. Bilgi sistemlerindeki kalıcı verilerin yönetimiyle ilgili kavramları ve ilişkileri açıklayan bu rapor, veri tabanı yönetim sistemleri, veri sözlüğü sistemleri gibi spesifik veri yönetimi bileşenlerinin sağladığı hizmetleri tanımlamak için kullanılmaktadır. ISO/IEC TR 10032, veri yönetiminin teknik yönlerine odaklanarak, veri yönetimi alanında standartlaşmayı kolaylaştırmak, ortak bir anlayış oluşturmak ve veri yönetiminin teknik boyutlarını ele almak için geliştirilmiş bir modeldir (ISO/IEC TR 10032, 2003).

3.2.2.6. ISO 15489-1: Records Management – Belge Yönetimi

ISO 15489, Uluslararası Standardizasyon Organizasyonu (ISO) tarafından kayıt yönetimi için oluşturulmuş uluslararası bir standarttır. Kayıtların oluşturulması, yakalanması ve yönetilmesi için yönergeleri belirleyen kavramları ve ilkeleri sunar.

Kuruluşlar bu standart ile kayıt ve belgelerini etkili bir şekilde yönetmeyi hedefler. ISO 15489 işletmelerin güvenli ve verimli bir kayıt tutma ve kayıt yönetim

süreçlerine sahip olmalarına yardımcı olur. Böylece işletmelere güvenlik yönetimi, veri gizliliğinin korunması, sürekli iyileştirme vb. konularda fayda sağlar (“SafetyCulture”, t.y.).

3.2.2.7. ISO/IEC 11179: Information Technology - Metadata Registries ve TS ISO/IEC 11179: Bilgi Teknolojisi - Metaveri Kayıtları

ISO/IEC 11179 standardı, Uluslararası Standardizasyon Örgütü (ISO) ve Uluslararası Elektroteknik Komisyonu (IEC) tarafından ortaklaşa geliştirilen ve üst veri yönetimi için kapsamlı bir çerçeve sunan uluslararası bir standarttır (ISO/IEC 11179-1, 2023). Bu standart, verilerin daha anlaşılır ve paylaşılabilir olmasını sağlamak amacıyla üst verilerin standartlaştırılması ve kaydedilmesi süreçlerini düzenlemektedir. Özellikle sağlık, finans, kamu yönetimi ve bilgi teknolojileri gibi veri yoğun alanlarda, veri standardizasyonu ve yönetimi açısından önemli bir rehber olarak kabul edilmektedir.

Bu standart verilerin kurum içi ve kurumlar arası düzeyde tutarlı bir şekilde tanımlanmasını teşvik etmekte, zaman, mekân ve uygulama farklılıklarına rağmen verilerin yeniden kullanılabilirliğini artırmayı hedeflemektedir. Ayrıca, veri tanımlarının etkin yönetimine olanak sağlamakta ve bu tanımların yeniden kullanımını teşvik ederek veri uyumluluğu ve standardizasyonunu desteklemektedir. Bu yönleriyle ISO/IEC 11179, organizasyonların veri yönetimi süreçlerini iyileştirmelerine ve veri kalitesini artırmalarına katkı sunmaktadır.

Türkiye'de ise bu standart, TS ISO/IEC 11179 adı altında Türk Standardları Enstitüsü (TSE) tarafından Türkçe'ye çevrilerek yayımlanmıştır. TS ISO/IEC 11179-3 (2023), uluslararası standardın tüm bölümlerini ve içeriğini Türkçe olarak sunarak Türkiye’deki kurumların bu standartları daha kolay anlamasını ve uygulamasını mümkün kılmaktadır. Böylelikle kurumlar üst veri yönetimi süreçlerinde uluslararası standartlarla uyumlu olarak hareket edebilmektedir.

ISO/IEC 11179 ve TS ISO/IEC 11179 veri yönetimi süreçlerinde yüksek standartların benimsenmesini teşvik ederek, veri paylaşımını kolaylaştırmakta, veri kalitesini artırmakta ve veri yönetişimini güçlendirmektedir.

3.2.2.8. ISO/IEC 19583: Information Technology - Concepts and Usage of Metadata (Bilgi Teknolojileri ve Üst Verinin Konseptleri)

ISO/IEC 19583, bilgi teknolojilerindeki üst verilerin kullanımı, yönetimi ve kaydedilmesine yönelik kavramları, yöntemleri ve örnek uygulamaları açıklayan bir standart serisidir (ISO/IEC 19583-1, 2019). Bu standart, özellikle ISO/IEC 11179 ve ISO/IEC 19763 standartlarını destekleyen teknik raporlardan oluşmaktadır. ISO/IEC 19583 serisi, dört ana teknik rapordan oluşmaktadır. Her bir rapor, üst verinin farklı yönlerini ele almaktadır:

ISO/IEC 19583-1:2019: Metadata Concepts (Üst Veri Kavramları): Bu bölüm, üst veri kavramlarını ve üst verilerin veri yönetimi süreçlerindeki rolünü açıklamaktadır. Ayrıca üst veri konseptinin temel unsurlarını tanımlamaktadır ve üst verilerin, veri modelleri ve meta modellerle olan ilişkisini detaylandırmaktadır. Bu bölüm, üst veri yönetimi için teorik bir temel sağlamaktadır.

ISO/IEC 19583-21:2022: Data Model in SQL (SQL'de Veri Modeli): Bu teknik rapor, ISO/IEC 11179-3'te tanımlanan meta modelin bir SQL veri tabanı dili kullanılarak nasıl uygulanabileceğini açıklamaktadır. Kayıtların somut örneklere dönüştürülmesi için SQL örnekleri sunmaktadır. Bölüm, geliştiricilere ve veri tabanı yöneticilerine, üst veri modellerini SQL ile uygulamak için rehberlik etmektedir.

ISO/IEC 19583-22:2018: Registering and Mapping Development Processes (Geliştirme Süreçlerini Kaydetme ve Eşleme): Bu rapor ISO/IEC 19763 serisinin yeteneklerini kullanarak süreç modellerinin kaydedilmesi ve eşleştirilmesine yönelik bir kullanım senaryosu sunmaktadır. Ayrıca süreç modelleme ve eşlemenin, yeniden kullanım için nasıl optimize edilebileceğini göstermektedir. Bu

bölüm, süreç modellerinin daha etkili şekilde paylaşılması ve yönetilmesini hedeflemektedir.

ISO/IEC 19583-23:2020: Data Element Exchange - DEX (Veri Öğesi Değişimi): Bu teknik rapor veri öğesi tanımlarının bir ISO/IEC 11179-3 üst veri kaydı üzerinden nasıl değiştirilebileceğini açıklamaktadır. Ayrıca standart bir veri iletişim formatı sağlamaktadır. Bu bölüm, veri öğelerinin farklı sistemler arasında güvenli ve etkili bir şekilde paylaşılmasına yardımcı olmaktadır.

ISO/IEC 19583, bilgi teknolojilerinde üst veri yönetimi ve süreç modelleme için hem teorik hem de pratik bir temel sağlamaktadır. Standart, üst verilerin kayıt altına alınması, paylaşılması ve yeniden kullanımı konusunda kapsamlı rehberlik sunarak bu süreçleri daha etkili ve verimli hale getirmeyi hedeflemektedir. Ayrıca geliştirme süreçlerinin iyileştirilmesi ve süreç modellerinin yeniden kullanımı için pratik örneklerle kullanıcıların bu alanlarda daha iyi çözümler üretmesine yardımcı olmaktadır. SQL veri modelleri ve mesaj değişim çerçeveleri gibi somut araçlarla uygulayıcılara teknik destek sağlayarak, karmaşık meta veri yapılarının kolayca uygulanmasını mümkün kılmaktadır. Böylece geliştiricilerin ve yöneticilerin üst veri yönetimi süreçlerini optimize etmelerine olanak tanımaktadır.

3.2.2.9. ISO 30301: Information and Documentation - Management Systems for Records — Requirements (Bilgi ve Dokümantasyon - Belge Yönetim Sistemleri — Gereksinimler)

ISO 30301, organizasyonların kayıt yönetim süreçlerini stratejik hedefleriyle uyumlu hale getirmeyi amaçlayan bir standarttır. Bu standart, kayıt yönetimini yalnızca bir arşivleme süreci olarak ele almanın ötesine geçerek, kurumsal performansı destekleyen ve hesap verebilirliği artıran bir araç olarak konumlandırmaktadır. Bu bağlamda, ISO 30301, organizasyonların kayıtların oluşturulması, kullanımı, bakımı ve imhasını kapsayan süreçlerini sistematik bir

şekilde düzenlemelerine olanak tanımaktadır. Standart, kayıtların kurumsal iş süreçlerine ve stratejik hedeflere olan katkısını vurgulamaktadır.

ISO 30301, tüm sektörlerde uygulanabilir niteliktedir. Kayıtların etkin yönetimi sayesinde organizasyonlar, bilgiye hızlı ve güvenli erişim sağlayabilir, yasal uyumluluğu artırabilir ve süreç verimliliğini destekleyebilir. Kayıtların yaşam döngüsünü yöneten bu standart, aynı zamanda organizasyonların risk yönetimi kapasitesini geliştirip şeffaflık ve izlenebilirliği artırmaktadır.

Standart, diğer yönetim sistemleri standartları (örneğin ISO 9001 ve ISO 14001) ile entegre edilebilecek bir yapıya sahiptir. ISO 30301'in temel amacı, kayıtların bir kurumsal varlık olarak değer üretebilmesini sağlamaktır. Böylece, organizasyonlar hem operasyonel süreçlerini iyileştirebilir hem de uzun vadeli hedeflerine ulaşmada kayıt yönetimini stratejik bir araç olarak kullanabilirler (ISO 30301, 2011).

3.2.2.10. ISO 30302: Information and documentation - Management Systems for Records - Guidelines for Implementation (Bilgi ve Belge Yönetimi - Belgeler için Yönetim Sistemleri - Uygulama Rehberi)

ISO 30302, kayıt yönetim sistemlerinin etkin bir şekilde planlanması, uygulanması ve değerlendirilmesi için rehberlik sağlayan bir standarttır. Bu standart, organizasyonların bilgi ve kayıt yönetimi süreçlerini stratejik hedefleriyle uyumlu hale getirmelerine yardımcı olur. ISO 30302, ISO 30300 serisinin bir parçası olarak geliştirilmiş olup, diğer yönetim sistemi standartlarıyla uyumlu bir çerçeve sunmaktadır ve organizasyonların kayıt yönetimi sistemlerini sürekli iyileştirme sürecine entegre etmelerini kolaylaştırmaktadır.

ISO 30302, kayıt yönetiminin planlanması ve uygulanması sırasında organizasyonel bağlam, riskler ve fırsatlar gibi faktörlerin dikkate alınmasını önermektedir. Standart, kayıt yönetim sistemlerinin performansını değerlendirmek ve bu sistemleri geliştirmek için gerekli ilkeleri sunmaktadır. Bu

süreçte organizasyonların kayıt yönetimi politikalarını tanımlaması, kayıtların yaşam döngüsünü düzenlemesi ve uygun kaynak tahsisiyle sistemlerini desteklemesi teşvik edilmektedir.

ISO 30302'nin amacı, kayıtların kurumsal süreçlere ve hedeflere olan katkısını en üst düzeye çıkarmaktır. Kayıt yönetim sistemlerinin verimliliği, şeffaflığı ve hesap verebilirliği artırmasına olanak tanırken organizasyonların yasal ve düzenleyici gerekliliklere uyum sağlamalarını da kolaylaştırmaktadır. Bu standart, organizasyonların kayıt yönetimi uygulamalarını diğer yönetim sistemleriyle entegre ederek bilgiye hızlı erişim, güvenlik ve sürdürülebilirlik sağlamaktadır.

ISO 30302, kurumsal süreçlerini geliştirmek isteyen kurumlar için pratik bir rehber sunmaktadır. Kayıt yönetimini iş süreçlerini ve performansı iyileştiren bir yönetim aracı olarak ele almaktadır (ISO 30302, 2020).

3.2.2.11. ISO/IEC 19773:2011: Information Technology - Metadata Registries (MDR) Modules - Bilgi Teknolojileri - Üst Veri Kayıtları Modülleri

ISO/IEC 19773 (2011) standardı, bilgi teknolojisi sistemlerinde yapısal ve tutarlı bir şekilde üst verilerin temsil edilmesi için standart bir çerçeve sunmaktadır. Standardın amacı, üst veri işleme süreçlerini düzenleyerek bilgilerin yüksek doğruluk ve verimlilikle depolanmasını ve yönetilmesini sağlamak böylece farklı sistemlerin sorunsuz bir şekilde entegrasyonunu kolaylaştırmaktır. Bu standart, çeşitli üst verilere yönelik veri yapısı genişletmelerini tanımlayarak farklı platformlar arasında birlikte çalışabilirliği ve uyumluluğu artırmayı hedeflemektedir. Özellikle veri tabanı yönetim sistemlerinde üst verilerin uygulanmasını ele almakta, yapı ve arayüzler sunmaktadır.

3.2.2.12. ISO 16175-1: Information and Documentation - Processes and Functional Requirements for Software for Managing Records (Bilgi ve Dokümantasyon - Belge Yönetimi Yazılımları için Süreçler ve İşlevsel Gereksinimler)

ISO 16175, elektronik ortamda kayıt yönetimi için rehberlik ilkelerini belirleyen bir standarttı. Bu standart, organizasyonların dijital kayıtlarını etkin bir şekilde yönetmelerini ve elektronik bilgi sistemleriyle uyumlu bir kayıt yönetim sistemi geliştirmelerini amaçlamaktadır. ISO 16175, özellikle elektronik bilgi sistemleri yoluyla oluşturulan ve yönetilen kayıtların güvenilirliğini, doğruluğunu ve erişilebilirliğini sağlamaya odaklanmaktadır.

Standardın temel odak noktası, organizasyonların iş süreçleriyle entegre edilmiş bir elektronik kayıt yönetim sistemi oluşturmasını teşvik etmektir. ISO 16175, bu bağlamda kayıtların oluşturulmasından, saklanmasına, erişimine ve sonunda imhasına kadarki yaşam döngüsünü yönetmek için gerekli ilke ve süreçleri açıklamaktadır. Elektronik kayıtların yönetimi, organizasyonel verimlilik ve yasal uyumluluk açısından büyük önem taşımaktadır.

ISO 16175 ayrıca, elektronik bilgi sistemleri üzerinden kayıt yönetimi gerçekleştiren organizasyonlar için rehberlik sağlarmaktadır. Bu sistemlerin tasarımı, uygulanması ve değerlendirilmesi sırasında dikkate alınması gereken teknik gereklilikler ve yönetim ilkeleri standart içinde detaylı şekilde ele alınmıştır. Bu süreçte, kayıtların güvenliğini sağlamak, bilgiye hızlı erişimi mümkün kılmak ve kayıtların yasal geçerliliğini korumak ana hedefler arasındadır.

Bu standart, bilgiye dayalı karar alma süreçlerini desteklemek, kayıtların doğruluğunu ve bütünlüğünü korumak ve uzun vadeli erişim gereksinimlerini karşılamak için organizasyonlara somut bir rehber sunmaktadır (ISO 16175, 2020).

3.2.2.13. ISO 30300: Information and Documentation — Records Management — Core Concepts and Vocabulary (Bilgi ve Belgeleme — Belge Yönetimi — Temel Kavramlar ve Kelime Bilgisi)

ISO 30300, organizasyonların kayıt yönetimi süreçlerini geliştirmek ve bilgi varlıklarını stratejik bir kaynak olarak yönetmek amacıyla oluşturulmuş bir yönetim sistemi standardıdır (ISO 30300, 2011). Bu standart, kayıtların yaşam döngüsünü, yani oluşturulmasından imhasına kadar geçen tüm süreçleri kapsayarak bilgi yönetimine sistematik bir yaklaşım getirmektedir.

ISO 30300, diğer yönetim sistemi standartları (örneğin ISO 9001 veya ISO 14001) ile aynı çerçeveyi takip etmektedir. Bu uyumluluk, organizasyonların farklı yönetim sistemlerini entegre etmelerini kolaylaştırmaktadır. Ayrıca, uluslararası düzeyde geçerli olan ortak bir dil ve terminoloji sunarak kayıt yönetimi alanında standartlaşmayı teşvik etmektedir. Standardın temel amacı, kayıt yönetimi süreçlerini kurumsal stratejiyle uyumlu hale getirerek hesap verebilirliği, şeffaflığı ve yasal uyumu sağlamaktır. Aynı zamanda, kayıtların güvenliği ve erişilebilirliği için ilkeler sunarak bilgiye hızlı erişimi mümkün kılmakta ve organizasyonel süreçlerin etkinliğini artırmaktadır (ISO/TC 46/SC 11, t.y.).

ISO 30300, tüm sektörlerde uygulanabilir olup her ölçekteki organizasyon için uygundur. Bu kapsamda, organizasyonların yalnızca belgeleri arşivlemekle kalmayıp, kayıtların oluşturulması, saklanması, düzenlenmesi ve imhası gibi tüm süreçlerini kapsamaktadır. Bilgi yönetimi süreçlerine stratejik bir yaklaşım getirerek bu süreçlerdeki riskleri en aza indirmekte ve kayıtların güvenliğini, bütünlüğünü ve kullanılabilirliğini artırmaktadır. Kayıt yönetimi sistemlerini kurarken organizasyonlara planlama, uygulama, değerlendirme ve sürekli iyileştirme adımları konusunda rehberlik etmektedir. Bu süreç, organizasyonların değişen ihtiyaçlarına ve teknolojik gelişmelere uyum sağlamalarını kolaylaştırmaktadır.

ISO 30300, bilgi yönetimini yalnızca teknik bir süreç olarak değil, kurumsal başarıyı artıran bir stratejik araç olarak ele almaktadır. Standart,

organizasyonların uzun vadeli sürdürülebilirlik hedeflerine ulaşmalarına ve iş süreçlerini optimize etmelerine yardımcı olmaktadır (ISO/TC 46/SC 11, t.y.).

3.2.2.14. ISO/IEC 27002: Information Security, Cybersecurity And Privacy Protection — Information Security Controls ve TS ISO/IEC 27002: Bilgi Güvenliği, Siber Güvenlik ve Gizlilik Koruması - Bilgi Güvenliği Kontrolleri

ISO/IEC 27002 standardı, bilgi güvenliği yönetim sistemleri için kapsamlı bir rehber sunmakta olup, organizasyonların siber güvenlik odaklı bir Bilgi Güvenliği Yönetim Sistemi kurması, uygulaması ve geliştirmesi için yol gösterici bir kaynak sağlamaktadır (ISO/IEC 27002, 2022). Bu standardın ana özellikleri arasında erişim kontrolü, kriptografi gibi temel siber güvenlik konularındaki uygulamaları ve kontrol hedeflerini içeren kapsamlı bir güvenlik çerçevesi sunulması yer almaktadır. Ayrıca organizasyonların bilgi güvenliği risklerini tanımlaması, değerlendirmesi ve etkili bir şekilde yönetmesi için risk yönetimi süreci içermektedir. Standart, hassas verilerin korunmasına yönelik taahhütler ile organizasyonların güvenilirliğini artırmaktadır. Çeşitli yasal ve düzenleyici veri koruma gereksinimlerine uyumu kolaylaştırmakta ve iş operasyonlarını kesintiye uğratabilecek güvenlik olaylarının olasılığını azaltarak operasyonel dayanıklılık sağlamaktadır.

ISO/IEC 27002, ISO/IEC 27001 standardının gerektirdiği BGYS'nin uygulanması için pratik bir plan sunmakta olup, organizasyonların siber tehditlere karşı kritik bilgilerini korumak amacıyla proaktif bir yaklaşım benimsemelerine yardımcı olmaktadır (ISO/IEC 27002, 2022). Türkçe versiyonu olan TS ISO/IEC 27002 standardı ise orijinal ISO/IEC 27002 standardının Türkçe çevirisi olarak Türk Standartları Enstitüsü (TSE) tarafından yayımlanmıştır. TS ISO/IEC 27002, orijinal standardın tüm içeriğini ve güncellemelerini yansıtacak şekilde periyodik olarak güncellenmektedir (TS ISO/IEC 27002, 2022).

3.2.2.15. ISO/IEC 27014: Information Security, Cybersecurity and Privacy Protection — Governance of Information Security ve TS ISO/IEC 27014: Bilgi Güvenliği, Siber Güvenlik ve Gizlilik Koruması - Bilgi Güvenliği Yönetimi

ISO/IEC 27014, bilgi güvenliğinin kurumsal yönetişimle entegrasyonunu sağlamak amacıyla geliştirilmiş bir standarttır. Bu standart bilgi güvenliğinin kurumsal yönetişime nasıl entegre edileceği ve bilgi güvenliği performansının kurumsal hedeflere ulaşma kapasitesini nasıl destekleyeceği hakkında kapsamlı bir rehberdir. Kurumların bilgi güvenliği yönetim süreçlerini etkili bir şekilde yönetebilmeleri için bir çerçeve sunmaktadır. ISO/IEC 27014, bilgi güvenliğine dair alınacak önlemlerin planlanması, izlenmesi, değerlendirilmesi ve iletilmesi süreçlerini yönlendirmektedir (ISO/IEC 27014, 2020).

Türkiye'de bu standart, Türk Standartları Enstitüsü (TSE) tarafından TS ISO/IEC 27014 adıyla yayımlanmıştır ve kurumların bilgi güvenliği yönetimini yönetişim yapısıyla uyumlu hale getirebilmeleri için yerel bir rehber niteliği taşımaktadır. TS ISO/IEC 27014, kurumların BGYS süreçlerini üst yönetim düzeyinde değerlendirmeleri, yönlendirmeleri ve izlemeleri için gerekli prensipleri içermektedir. Bilgi güvenliği yönetimini kurumsal stratejiyle uyumlu bir hale getirmek isteyen tüm organizasyonlar için bu standart yol gösterici niteliktedir (TS TS ISO/IEC 27014, 2022).

3.2.2.16. ISO/IEC 29100: Information Technology — Security Techniques — Privacy Framework

ISO/IEC 29100, bilgi ve iletişim teknolojisi sistemlerinde kişisel bilgilerin korunması için kapsamlı bir gizlilik çerçevesi sunan uluslararası bir standarttır. Kişisel bilgilerin korunmasını sağlamak amacıyla kurumsal, teknik ve ilkesel unsurları bir araya getiren bu standart, gizlilik yönetim süreçlerinde kullanılacak ortak bir terminoloji sunarak, kişisel bilgilerin korunmasında görev alan unsurların net bir şekilde tanımlanmasına olanak tanımaktadır. Standart ayrıca, kişisel

verilerin işlenmesiyle ilgili olarak gizliliğin korunması için gerekliliklerin belirlenmesi ve gizlilik ilke ve kontrollerinin tasarımı, uygulanması, yönetimi ve değerlendirilmesine yönelik bir temel oluşturmaktadır. Bu çerçeve, bilgi ve iletişim teknolojisi sistemlerinde kişisel bilgilerin korunmasını sağlayacak politikaların ve önlemlerin geliştirilmesi, uygulanması ve sürekli iyileştirilmesi için organizasyonlara yol göstermektedir (ISO/IEC 29100, 2024).

ISO/IEC 29100 standardı, bilgi ve iletişim teknolojisi sistemlerinde veya hizmetlerinde gizlilik kontrolleri gerektiren durumlarda kullanılmak üzere tasarlanmış olup, özellikle organizasyonların kişisel verileri koruma konusundaki yaklaşımlarını standartlaştırmalarına ve iyileştirmelerine yardımcı olmaktadır. Standart, dört ana bileşen üzerine odaklanmaktadır:

- Ortak bir gizlilik terminolojisi belirlemek,
- Kişisel verilerin işlenmesinde aktörleri ve rollerini tanımlamak,
- Gizlilik koruma önlemlerini açıklamak,
- Bilgi teknolojisi için bilinen gizlilik ilkelerine referanslar sağlamak.

Bu kapsamda gizlilik ilkeleri, rıza ve seçim, amaç meşruiyeti ve belirtimi, veri minimizasyonu, kullanım ve saklama sınırlaması gibi temel prensipleri içermektedir (National Law Review, 2021). Standart, organizasyonlara bilgi ve iletişim teknolojisi sistemlerinde kişisel bilgilerin korunmasını sağlayacak politikaların ve önlemlerin geliştirilmesi, uygulanması ve sürekli iyileştirilmesi için rehberlik etmektedir. Günümüz dijital çağında giderek sıkılaşan veri koruma düzenlemeleri karşısında ISO/IEC 29100, gizlilik yönetimi konusunda global bir referans noktası oluşturarak organizasyonlara yol göstermektedir (ISO/IEC 29100, 2024). ISO/IEC 29100, kişisel bilgilerle ilgili olarak organizasyonların uyması gereken 11 temel gizlilik ilkesini içermektedir. Tablo 2’de bu ilkeler sunulmuştur.

Tablo 2. ISO/IEC 29100'in 11 temel gizlilik ilkesi

Gizlilik İlkesi	Açıklama
Rıza ve Seçim	Bireylerin verilerinin işlenmesine dair rızalarının alınması ve tercihlerine saygı gösterilmesi.
Amacın Meşruiyeti ve Belirtilmesi	Kişisel bilgilerin belirli, meşru ve açık bir amaç için toplanması.
Toplama Sınırlaması	Kişisel verilerin yalnızca gerekli olduğu durumlarda toplanması.
Veri Minimizasyonu	Sadece gerekli miktarda kişisel veri toplanması.
Kullanım, Saklama ve İfşa Sınırlaması	Kişisel verilerin belirlenen amaç dışında kullanılmaması, saklanma süresinin sınırlandırılması ve yetkisiz ifşalardan korunması.
Doğruluk ve Kalite	Kişisel bilgilerin doğru ve güncel olmasının sağlanması.
Açıklık, Şeffaflık ve Bilgilendirme	Verilerin toplanma ve işlenme süreçlerinin şeffaf olması ve bireylerin bilgilendirilmesi.
Bireysel Katılım ve Erişim	Bireylerin kendileriyle ilgili bilgilere erişebilmesi ve gerektiğinde düzeltme talep edebilmesi.
Sorumluluk	Verilerin işlenmesinde ilgili tarafların sorumluluğunun belirlenmesi.
Bilgi Güvenliği	Kişisel verilerin güvenliğinin sağlanması için gerekli teknik ve organizasyonel önlemlerin alınması.
Gizlilik Uyumluluğu	Kişisel verilerin işlenmesinde yasal düzenlemeler ve gizlilik politikalarına uyumun sağlanması.

3.2.2.17. ISO/IEC 27701: Security Techniques — Extension to ISO/IEC 27701 and ISO/IEC 27002 for Privacy Information Management — Requirements and Guidelines

ISO/IEC 27701, bilgi güvenliği yönetimi çerçevesinde gizlilik ve veri koruma gereksinimlerini karşılamak amacıyla geliştirilmiş bir standarttır. ISO/IEC 27701 ve ISO/IEC 27002 üzerine inşa edilen bu standart, kuruluşların kişisel verileri koruma süreçlerini yönetmelerine yardımcı olacak bir Gizlilik Bilgi Yönetim Sistemi (PIMS - Privacy Information Management System) oluşturmaktadır. ISO/IEC 27701 veri koruma yasaları ile uyum sağlamak isteyen kurumlar için önemli bir rehberdir. Standart, veri sorumluları ve veri işleyenler için uygun gizlilik kontrollerini ve gereksinimleri tanımlamaktadır. Kişisel verilerin toplanması,

işlenmesi, saklanması ve paylaşılması gibi süreçlerde gizlilik risklerini en aza indirmek için gerekli önlemleri belirlemektedir (ISMS.online, 2024.).

3.2.2.18. ISO/IEC 38500: Information Technology - Governance of IT for the Organization (Bilgi Teknolojisi - Kuruluşlar için Bilgi Teknolojisi Yönetim Rehberi)

ISO/IEC 38500, bilgi teknolojisi (BT) yönetimi için uluslararası bir çerçeve sunan bir standarttır. Bu standart, BT'nin bir organizasyonda stratejik hedeflere ulaşmasını desteklemek için üst yönetim tarafından nasıl yönlendirilmesi ve kontrol edilmesi gerektiğine dair rehberlik sağlamaktadır. Standart BT ile ilgili kararların etik, sorumlu, adil ve şeffaf bir şekilde alınmasını destekleyen ilkelere dayanmaktadır. ISO/IEC 38500, BT'nin yalnızca teknik bir destek fonksiyonu olarak değil, stratejik bir kaynak olarak ele alınması gerektiğini vurgulamaktadır (ISO/IEC 38500, 2015).

3.2.2.19. ISO/IEC 19763: Information Technology - Metamodel Framework for Interoperability (MFI) ve TS ISO/IEC 19763: Bilgi Teknolojileri - Birlikte Çalışabilirlik için Meta Model Çerçevesi

ISO/IEC 19763 standardı, modellerin kaydına ve modeller arasındaki eşlemelere odaklanarak bilgi sistemlerinin birlikte çalışabilirliğini artırmayı amaçlayan çok bölümlü bir standarttır. Bu standart, veri yönetimi ve modelleme için ortak bir temel oluşturan ISO/IEC 11179'da belirtilen ilkeleri temel almaktadır (ISO/IEC 11179-1, 2015; ISO/IEC 19763, 2020). Standart, çeşitli meta modeller ve çerçeveler sunarak organizasyonların farklı model türlerini etkili bir şekilde yönetmelerine ve entegre etmelerine olanak tanımaktadır. ISO/IEC 19763 standardının Türkçe çevirisi, Türk Standartları Enstitüsü (TSE) tarafından TS ISO/IEC 19763 adıyla yayımlanmıştır. Bu çeviri, standardın Türkiye'deki uyumluluk ve uygulanabilirlik süreçlerini kolaylaştırmayı hedeflemektedir. Bilgi sistemlerinin birlikte çalışabilirliğini artırmak için kullanılan meta modelleme çerçevelerini ulusal

gereksinimlere uygun bir şekilde sunmaktadır (TS ISO/IEC 19763, 2020). Tablo 3'te bu standardın bölümleri ve kısa açıklamaları yer almaktadır.

Tablo 3. ISO/IEC 19763 Standardının bölümleri

Bölüm No	Bölüm Başlığı	Açıklama
1	Çerçeve	Modellerin kaydı ve eşlemesi için genel çerçeve sunar.
2	Meta Modelin Özellikleri ve İlişkileri	Meta model yapılarını ve ilişkilerini açıklar.
3	Ontoloji Kaydı İçin Meta Model	Ontoloji modellerinin kaydı ve yönetimi için yapı sağlar.
4	Varlık-İlişki Meta Modeli	Varlık ve ilişkilerin tanımlanması için meta model sunar.
5	Süreç Modeli Kaydı İçin Meta Model	İş süreçlerini temsil eden modellerin kaydını destekler.
6	Kayıt Özeti	Standart içindeki kayıtların genel bir özetini sunar.
7	Hizmet Kaydı İçin Meta Model	Hizmet odaklı mimarilerdeki hizmetlerin kaydı için yapı oluşturur.
8	Rol ve Hedef Kaydı İçin Meta Model	İş rolleri ve hedeflerinin tanımlanması için bir yapı sağlar.
9	Talep Üzerine Model Seçimi	Model seçimi ve yönetimi için rehberlik sunan kısımdır.
10	Çekirdek Model ve Temel Eşleme	Modeller arasında temel eşleme yöntemlerini ve çekirdek model yapılarını açıklar.
11	Bilgi Değişimi İçin Temel Şema	Bilgi değişimi için ortak bir şema oluşturur.
12	Bilgi Modeli Kaydı İçin Meta Model	Bilgi modellerinin kaydı ve yönetimi için yapı sunar.
13	Form Tasarımı Kaydı İçin Meta Model	Form tabanlı sistemlerin kaydı için gerekli ilkeleri açıklar.
14	İşlem Modelleri için Meta Model	İşlem modellerini kaydetmek için meta model sunar.
15	Kavram Modeli Kaydı İçin Meta Model	Kavram modellerinin kaydı için meta model sağlar.
16	Belge Modeli Kaydı İçin Meta Model	Doküman ve belgelerin meta model bazında kaydını destekler.

Bölüm 1 ve bölüm 10 (Çerçeve ve Çekirdek Model), tüm bu bölümleri bir araya getiren ve birlikte çalışabilirliği sağlayan çekirdek bir model sunmaktadır. Bu model, farklı veri ve model türleri arasında bir bağ kurarak birlikte çalışabilirliği artırmaktadır. Bölüm 3 ve bölüm 5 (Ontoloji ve Süreç Modelleri) ise iş süreçleri ve ontoloji temelli sistemler için meta modeller, operasyonel mükemmellik ve bilgi paylaşımı için kritik öneme sahiptir.

3.2.2.20. ISO/IEC 19510: Information Technology — Object Management Group Business Process Model and Notation (Bilgi Teknolojisi — Nesne Yönetimi Grubu İş Süreci Modelleme ve Notasyonu)

ISO/IEC 19510 standardı iş süreçlerinin modellenmesi ve yönetimi konusunda küresel bir referans noktası oluşturmaktadır. Standart iş süreçlerinin modellenmesi ve notasyonu için kapsamlı bir çerçeve sunmaktadır. Standardın temel amacı, iş analistlerinden teknik geliştiricilere ve süreçleri yönetecek iş insanlarına kadar tüm iş kullanıcıları tarafından kolayca anlaşılabilecek bir notasyon sağlamak, böylece iş süreci tasarımı ile süreç uygulaması arasındaki boşluğu dolduran standartlaştırılmış bir köprü oluşturmaktadır (ISO/IEC 19510, 201, Object Management Group, t.y.).

ISO/IEC 19510, iş birliği diyagramları, süreç diyagramları ve koreografi diyagramlarının notasyonunu ve semantiğini tanımlamaktadır. Farklı modelleme notasyonları ve bakış açıları karşısında, iş süreci modeli ve notasyonunu standartlaştırmayı amaçlamaktadır (Object Management Group, t.y.).

Standart 507 sayfalık kapsamlı bir dokümandır ve ISO/IEC JTC 1 Teknik Komitesi tarafından geliştirilmiştir. Standardın kullanımı, organizasyonların iş süreçlerini daha etkili bir şekilde tasarlamasına, uygulamasına ve yönetmesine olanak tanımaktadır. Ayrıca, farklı paydaşlar arasındaki iletişimi geliştirerek, iş süreçlerinin daha iyi anlaşılmasını ve optimize edilmesini sağlamaktadır (Chinosi ve Trombetta, 2012).

Bu standart, düzenli olarak gözden geçirilmekte ve güncelliğini korumaktadır. ISO/IEC 19510, sürdürülebilir kalkınma hedeflerine de katkıda bulunmakta, bu da standardın sadece iş süreçleri için değil, aynı zamanda daha geniş toplumsal hedefler için de önemini vurgulamaktadır (Recker, 2010).

3.2.2.21. ISO 15836 - Dublin Core Üst Veri Eleman Seti

Dublin Core Üst Veri Eleman Seti hem dijital hem de fiziksel kaynakların tanımlanması için kullanılan 15 temel üst veri elemanından oluşan bir standarttır. Bu standart; başlık, konu,dil, format, açıklama, tür gibi unsurları içerir ve çok çeşitli bilgi kaynaklarının tanımlanmasını sağlar. Dublin Core, internet üzerinde bilgi paylaşımını ve birlikte çalışabilirliği kolaylaştırmak amacıyla geliştirilmiştir. Kullanımının kolaylığı ve esnekliği sayesinde, birçok farklı alanda yaygın olarak benimsenmiştir.

1995 yılında geliştirilen Dublin Core, sonraki yıllarda ISO standardı haline gelerek ISO 15836 olarak kabul edilmiştir. Bu standardizasyon, Dublin Core'un uluslararası düzeyde tanınmasını sağlamış ve dijital bilgi kaynaklarının yönetimi ve paylaşımı için geniş çapta kullanılabilir hale getirmiştir. Örneğin, kütüphaneler, müzeler ve dijital arşivler, bu standart sayesinde dijital koleksiyonlarını sistematik bir şekilde tanımlayabilmekte ve bu koleksiyonları diğer kuruluşlarla paylaşabilmektedir. Dublin Core aynı zamanda diğer üst veri standartları için bir temel teşkil eder ve bu sayede üst verilerin birbiriyle uyumlu olmasını sağlar (Weibel, 2009; Dublin Core Metadata Initiative, 2020).

Dublin Core'daki "identifier" elemanı, dijital nesnelerin kalıcı tanımlayıcılarını da (Persistent Identifier – PID) içerebilecek şekilde tasarlanmıştır. Örneğin, Handle Sistemi üzerinden üretilen bir "handle constant" (örn. hdl:20.500.12345/ABC123) bu alana kaydedilerek nesnenin kalıcı erişim bağlantısı güvence altına alınabilir (Dublin Core Metadata Initiative, 2020; CNRI, 2023).

3.2.2.22. ISO 20652:2017 (Producer-Archive Interface Methodology Abstract Standard - PAIMAS)

ISO 20652:2017, PAIMAS (Üretici-Arşiv Arayüzü Metodolojisi Soyut Standardı) olarak bilinen, dijital içerik üreten kurumlar ile arşivler arasındaki bilgi alışverişini yapısallaştıran uluslararası bir standarttır. PAIMAS, dijital bilginin üreticiden

arşive nasıl aktarılacağını tanımlayan süreçleri belirler ve bu sürecin her aşamasında kullanılan terminoloji, formatlar ve iş akışlarını belirler. Bu standart, arşivleme sürecinde yer alan tarafların sorumluluklarını açıkça tanımlar ve böylece üretici ve arşiv arasındaki iletişim ve iş birliğini optimize eder. PAIMAS, dijital içeriklerin uzun vadede güvenli bir şekilde saklanmasını ve arşivlenmesini sağlamak için önemli bir çerçevedir (“DCC”, 2009).

3.2.2.23. ISO 42010: Software, Systems and Enterprise — Architecture Description (Yazılım, Sistemler ve İşletme — Mimari Tanımlama)

ISO/IEC/IEEE 42010, sistem ve yazılım mühendisliğinde mimari tanımlamalar için uluslararası bir standarttır. Bu standart, çeşitli varlıkların—yazılımlar, sistemler, işletmeler, sistemler bütünü, ürün aileleri, hizmet hatları, teknolojiler ve iş alanları—mimari tanımlarının yapısını ve ifadesini belirler (ISO/IEC/IEEE 42010, 2022). Mimari tanımlamalar, bu varlıkların yapısını, davranışını, tasarımını ve evrimini anlamak için kritik öneme sahiptir (ISO/IEC/IEEE 42010, t.y.).

Standart, mimari tanımlamalar için bir kavramsal model sunar ve mimari bakış açıları, paydaş perspektifleri, mimari yönler, model türleri, mimari tanımlama çerçeveleri ve dilleri gibi unsurları içerir (ISO/IEC/IEEE 42010, 2022). Bu unsurlar, mimari tanımlamaların tutarlılığını ve etkinliğini artırmayı amaçlar. Standart ayrıca, mimari tanımlamaların oluşturulması, analizi ve sürdürülmesi süreçlerinde rehberlik eder (ISO/IEC/IEEE 42010, t.y.). ISO/IEC/IEEE 42010, mimari tanımlamaların paydaşların endişelerini ele almasını ve mimari kararların belgelenmesini teşvik eder. Bu standart, mimari tanımlamaların kalitesini ve anlaşılabilirliğini artırarak, sistemlerin ve yazılımların başarılı bir şekilde geliştirilmesine katkıda bulunur.

3.2.2.24. ISO 14721: TRAC (Trustworthy Repositories Audit & Certification)

TRAC (Güvenilir Depolar Denetim ve Sertifikasyonu), dijital arşivlerin güvenilirliğini değerlendirmek için kullanılan uluslararası bir standarttır. Bu standart, dijital arşivlerin uzun vadede güvenli bir şekilde işlevini sürdürebilmesi için gerekli olan yönetimsel, teknik ve kurumsal kriterleri belirler. TRAC metrikleri,

ISO 14721:2012 standardına dayanmaktadır. TRAC kontrol listesinin nihai versiyonu 2007 yılında yayınlanmıştır. TRAC, dijital arşivlerin denetim ve sertifikasyon süreçlerini kolaylaştırır ve bu sayede bu arşivlerin uzun süre boyunca güvenilir kalmasını sağlar. Özellikle dijital koruma ve arşivleme alanında faaliyet gösteren kuruluşlar, TRAC standardını kullanarak dijital varlıklarının korunmasını güvence altına alır. Arşivlerin teknolojik değişimlere ayak uydurmasını ve güvenilirliklerini sürdürmelerini sağlayan en önemli araçlardan biridir (*National Archives and Records Administration,* 2007; “CRL”, t.y.).

3.2.2.25. ISO 19115: Geographic Information - Metadata (Coğrafi Bilgi - Üst Veri)

ISO 19115 (Geographic Information - Metadata) standardı, coğrafi bilgi ve hizmetlerle ilgili üst verileri tanımlayan bir uluslararası standarttır. Bu standart, coğrafi bilgilerin yönetimi, kullanımı ve paylaşımı için gerekli üst veri bileşenlerini detaylandırır. ISO 19115, özellikle Coğrafi Bilgi Sistemleri (Geographic Information System - GIS) kullanılarak üretilen verilerin yönetilmesi, depolanması ve paylaşılmasına yönelik bir rehber sunar. Coğrafi verilerin farklı sistemler arasında uyumlu olmasını sağlamak için geliştirilen bu standart, verilerin daha etkin ve verimli bir şekilde kullanılmasını kolaylaştırır. Ayrıca coğrafi bilgilerin uzun vadede erişilebilir ve güvenilir kalmasını sağlamak için gerekli olan üst veri yapısını da belirler (“NASA Earthdata”, t.y.; “Esri ArcGIS Pro”, t.y.).

3.2.2.26. ISO 23081-1: Information and Documentation - Records Management Processes - Metadata for Records (Bilgi ve Dokümantasyon - Belge Yönetimi Süreçleri - Belgeler için Üstveri)

ISO 23081-1:2017 standardı, belgelerin yönetimi süreçlerinde kullanılan üst verilerin nasıl oluşturulacağını ve yönetileceğini belirlemektedir. Bu standart, belgelerin izlenebilirliğini ve uzun vadede erişilebilir olmasını sağlamak için gerekli

üst veri gereksinimlerini tanımlamaktadır. ISO 15489-1:2016 gereklilikleri çerçevesinde üst veri setlerini belirlemektedir. ISO 23081-2:2021 ve 23081-3:2011 uygulama aşaması için daha fazla detaylı açıklama içermektedir (“International Standard ISO 23081-1”, 2017).

3.2.2.27 ISO/IEC JTC 1/SC 42: Artificial intelligence (Yapay Zekâ)

ISO/IEC JTC 1/SC 42 yapay zekâ alanında uluslararası standartların geliştirilmesinden sorumlu bir alt komitedir. Bu komite, AI teknolojilerinin endüstri, hükümet ve akademi tarafından daha etkin ve sorumlu bir şekilde kullanılmasına katkıda bulunarak, bu teknolojilerin potansiyelini en üst düzeye çıkarmayı hedeflemektedir. ISO ve IEC iş birliğiyle oluşturulan bu komite, AI teknolojilerinin güvenilir, şeffaf ve etik bir şekilde uygulanmasını sağlamak amacıyla çalışmaktadır. SC 42’nin kapsamı, AI sistemleri için terminoloji, risk yönetimi, veri kalitesi, güvenlik ve etik gibi kritik konuları içermektedir.

ISO/IEC 42001 bu komitenin yönettiği standartlardandır (ISO/IEC JTC 1/SC 42, 2017). AI yönetim sistemleri için gereklilikleri belirlemekte ve organizasyonların bu teknolojileri sorumlu bir şekilde kullanmalarına rehberlik etmektedir.

3.2.2.28. ISO/IEC 42001: Information Technology - Artificial Intelligence - Management System (Bilgi Teknolojisi — Yapay Zekâ — Yönetim Sistemi)

ISO/IEC 42001, ISO/IEC JTC 1/SC 42 komitesinin yürüttüğü, yapay zekâ yönetim sistemleri için uluslararası bir standarttır. Bu standart, organizasyonların yapay zekâ tabanlı ürün ve hizmetleri sorumlu bir şekilde geliştirmeleri, uygulamaları ve sürekli iyileştirmeleri için gereklilikleri belirlemektedir. Standart yapay zekâ sistemlerinin etik, şeffaf ve güvenilir olmasını sağlamayı hedeflemektedir (ISO/IEC 42001, 2023).

Standart, yapay zekâ yönetim sistemlerinin kurulumu, uygulanması, sürdürülmesi ve sürekli iyileştirilmesi için bir çerçeve sunmaktadır. Bu çerçeve, organizasyonların yapay zekâ ile ilgili riskleri ve fırsatları etkin bir şekilde yönetmelerine yardımcı olmaktadır. Ayrıca, yapay zekâ sistemlerinin yaşam döngüsü boyunca güvenlik, güvenilirlik ve etik ile ilgili sorumlulukların yerine getirilmesini teşvik etmektedir (ISO/IEC 42001, 2023).

3.2.2.29. ISO/IEC 20000: IT Service Management – A Practical Guide (BT Hizmet Yönetimi – Pratik Bir Rehber)

ISO/IEC 20000, bilgi teknolojisi hizmet yönetimi (ITSM) için uluslararası bir standarttır ve kuruluşların etkili bir şekilde BT hizmetleri sunmalarını, yönetmelerini ve sürekli iyileştirmelerini sağlamak için rehberlik etmektedir (ISO/IEC 20000, 2019). Standart, hizmet yönetimi süreçlerini planlama, uygulama, izleme ve iyileştirme için bir çerçeve sunmaktadır. Standart, özellikle hizmet seviyesi yönetimi, olay yönetimi, problem yönetimi, değişiklik yönetimi ve varlık yönetimi gibi süreçlere odaklanmaktadır. Bu çerçeve, ITIL gibi hizmet yönetimi uygulamalarıyla da uyumlu olup, kuruluşların BT hizmet kalitesini artırarak rekabet avantajı sağlamalarına yardımcı olmaktadır (Taylor, 2019).

3.2.2.30. COBIT (Control Objectives for Information and Related Technologies - Bilgi ve İlgili Teknolojiler İçin Kontrol Hedefleri)

COBIT, ISACA (Information Systems Audit and Control Association - Bilgi Sistemleri Denetim ve Kontrol Derneği) tarafından geliştirilen ve bilgi teknolojisi (BT) yönetimi ile denetimi için uluslararası bir çerçeve sunan bir standarttır (ISACA, t.y.a). COBIT, BT'nin kurumsal hedeflere ulaşmasını desteklemek ve BT risklerini yönetmek amacıyla süreç bazlı bir yaklaşım sunmaktadır. İlk olarak 1996 yılında yayımlanan COBIT, BT yönetişimi ve yönetiminde beş ana prensibe dayanan bir yapı sunmaktadır. Bu prensipler şunlardır:

- Paydaş İhtiyaçlarının Karşılanması: COBIT, BT hizmetlerinin organizasyonun genel ihtiyaçlarını ve paydaşların beklentilerini karşılayacak şekilde uyarlanmasını sağlamaktadır (ISACA, t.y.b).

- Kapsamlı Bir Çerçeve Sağlama: COBIT, BT yönetişimi ve yönetiminde gereken tüm bileşenleri kapsayarak kuruluşlara kapsamlı bir yapı sunmaktadır.

- Sonuçların Uygulanabilir Olması: COBIT, stratejiden günlük operasyona kadar tüm BT süreçlerinin uyum içinde olmasını teşvik etmektedir.

- Yönetim ve Yönetişimin Ayrılması: COBIT, BT yönetimi ile BT yönetişimini birbirinden ayırarak her birine özgü roller ve sorumluluklar tanımlamaktadır.

- Tek Bir Entegre Çerçeve Kullanımı: COBIT, diğer yönetim ve denetim çerçeveleriyle (örneğin ITIL, ISO/IEC 27001) uyumlu olacak şekilde entegre bir yapı sunmaktadır (ISACA, t.y.b).

COBIT'in sunduğu bu çerçeve, BT'nin iş stratejileriyle uyumunu ve BT hizmetlerinin iş hedeflerine katkıda bulunmasını sağlamayı hedeflemektedir. Aynı zamanda, veri güvenliği, uyumluluk ve verimliliği artırmayı amaçlayarak BT yönetişimi için temel kontroller sunmaktadır. COBIT'in en güncel versiyonu olan COBIT 2019, kuruluşların hızla değişen iş ve teknoloji ortamına uyum sağlamalarına yardımcı olacak ek düzenlemeler sunmaktadır. COBIT 2019, önceki sürümlere göre daha esnek bir yapıya sahip olup, BT'nin kurumsal stratejilerle olan ilişkisini daha da güçlendirmiştir (ISACA, t.y.a). Bu güncel sürüm, BT yönetişiminde daha fazla özelleştirme olanağı sunarak her kuruluşun kendi ihtiyaçlarına uygun bir yapı oluşturmasını sağlamaktadır.

COBIT 2019, ayrıca risk yönetimi ve uyumluluk gibi kritik konulara da özel bir vurgu yapmakta ve organizasyonların BT yönetişimini sürekli iyileştirme ve güncel tutma süreçlerini desteklemaktedir. COBIT 2019'da yer alan yetkinlik modeli ve

yetki seviyeleri, BT yönetimi ve yönetişimini daha ölçülebilir ve denetlenebilir hale getirmektedir.

3.2.2.31. PREMIS (Preservation Metadata: Implementation Strategies - Koruma Üst verisi: Uygulama Stratejileri)

PREMIS, dijital nesnelerin uzun süreli korunması amacıyla üst verilerin yönetimini standartlaştırmak için geliştirilmiş uluslararası bir çerçevedir (“The Library of Congress”, t.y.).

PREMIS Veri Sözlüğü, nesneler (dijital varlıklar), entelektüel varlıklar, olaylar (koruma işlemleri), ajanlar (koruma işlemini gerçekleştirenler) ve haklar olmak üzere beş ana bileşen üzerine kurulmuştur. Bu bileşenler, dijital nesnelerin korunması için gerekli tüm işlemleri ve bu işlemler sırasında üretilen üst verileri tanımlar. Özellikle dijital nesnelerin özgünlüklerinin, bütünlüklerinin ve kullanılabilirliklerinin uzun süre korunabilmesi için önemli olan PREMIS, dijital varlıkların korunması sürecinde karşılaşılan zorlukları yönetmek için kapsamlı bir çözüm sunar. (PREMIS Editorial Committee, 2015, s. 7).

PREMIS, dünya genelindeki dijital arşivleme projelerinde kullanılmakta olup, pek çok ticari ve açık kaynak dijital koruma yazılımı ve sistemlerine entegre edilmiştir. Özellikle kütüphaneler, müzeler, arşivler ve diğer dijital koruma kuruluşları için vazgeçilmez bir standart haline gelmiştir (PREMIS Editorial Committee, 2015, s. 1). PREMIS’in son sürümü olan 3.0, 2015 yılında yayımlanmıştır ve dijital ortamlardaki teknolojik gelişmelere cevap verecek şekilde güncellenmiştir (PREMIS Editorial Committee, 2015, s. 4).

3.2.2.32. Deming Döngüsü (Plan – Do – Check – Act, PDCA)

Deming döngüsü süreçlerin sürekli iyileştirilmesini sağlamak amacıyla kullanılan dört aşamalı bir modeldir. İlk olarak Dr. W. Edwards Deming tarafından geliştirilen

bu döngü, organizasyonların verimliliklerini artırmalarına, kaliteyi iyileştirmelerine ve değişime uyum sağlamalarına yardımcı olmaktadır. Süreç iyileştirme planının oluşturulması (Plan), bu planın küçük ölçekte uygulanması (Do), sonuçların ölçülmesi ve değerlendirilmesi (Check) ve başarılıysa yaygınlaştırılması veya gerekiyorsa revize edilmesi (Act) adımlarından oluşmaktadır (Deming, 1986).

Veri yönetimi ve veri mimarisi projelerinde Deming Döngüsü, sürekli iyileştirme için güçlü bir çerçeve sunmaktadır. Örneğin veri kalitesinin artırılması veya veri alt yapısındaki verimliliğin geliştirilmesi için bu model kullanılabilir. PDCA, veri yönetimi süreçlerini düzenli bir şekilde analiz ederek hataları minimize etmeyi ve süreçleri optimize etmeyi mümkün kılmaktadır (Ladley, 2020).

3.2.2.33. ITIL (Information Technology Infrastructure Library - Bilgi Teknolojisi Altyapı Kütüphanesi)

ITIL, Bilgi Teknolojisi (BT) hizmet yönetimi için dünya çapında kabul görmüş bir uygulama çerçevesidir. BT hizmetlerinin etkin, verimli ve müşteri odaklı bir şekilde yönetilmesi için yol gösterici prensipler sunmaktadır. ITIL, kurumların BT süreçlerini iş stratejileriyle uyumlu hale getirmesine yardımcı olurken, müşteri memnuniyetini artırma, operasyonel maliyetleri düşürme ve BT performansını iyileştirme gibi hedeflere ulaşmalarını sağlamaktadır (Axelos, 2019).

ITIL'in temel amacı, BT hizmetlerinin yüksek kalitede sunulması, BT operasyonlarının sürekli izlenmesi ve hizmet seviyelerinin iyileştirilmesidir. Çerçeve, olay yönetimi, problem yönetimi, değişiklik yönetimi ve konfigürasyon yönetimi gibi süreçleri kapsamaktadır. Bu süreçler, BT altyapısının daha hızlı ve güvenilir bir şekilde hizmet verebilmesi için optimize edilmiştir (Hanna, Rudd, Macfarlane, Windebank ve Rance, 2007). Özellikle veri yönetimi süreçleriyle sürekli izleme ve performans optimizasyonu konularında önemli katkılar sağlamakta böylece BT operasyonlarının daha verimli ve sürdürülebilir bir yapıya kavuşmasına olanak tanımaktadır.

ITIL, veri yönetimi ve izleme alanında performans optimizasyonunu sağlayarak, organizasyonların veriyle ilgili süreçlerini sürekli iyileştirmeye yönelik yaklaşımlar geliştirmelerine de yardımcı olmaktadır. Bu, BT operasyonlarının daha etkili bir şekilde yönetilmesine ve performansın optimize edilmesine katkıda bulunmaktadır (Axelos, 2019).

ITIL'in en güncel versiyonu olan ITIL 4, dijital dönüşüm çağında daha çevik ve esnek bir hizmet yönetimi modeli sunmakta ve "Değer Akışı" gibi kavramlarla müşteri odaklı hizmet sunumuna vurgu yapmaktadır. ITIL 4, organizasyonların sürekli iyileştirme kültürünü benimsemesine ve BT'nin organizasyonel hedeflerle uyumlu hale gelmesine olanak tanımaktadır (Taylor, 2019).

3.2.2.34. Avrupa Birliği GDPR (Genel Veri Koruma Yönetmeliği) (2018)

Genel Veri Koruma Yönetmeliği (GDPR), Avrupa Birliği'nde yaşayan bireylerin kişisel verilerinin korunmasını amaçlayan bir düzenlemedir. 2018 yılında yürürlüğe giren GDPR, AB sınırları içinde kişisel verileri toplayan ve işleyen tüm kuruluşlar için bağlayıcıdır. Bu düzenleme, veri toplama ve işleme süreçlerinde bireylere önemli haklar tanır ve işletmelere ciddi yükümlülükler getirir. GDPR kapsamında bireyler, hangi kişisel verilerinin toplandığı, nasıl kullanıldığı ve kimlerle paylaşıldığı konusunda bilgilendirilme hakkına sahiptir. Ayrıca bireylerin verilerinin silinmesini isteme ve veri taşınabilirliği hakları da GDPR kapsamında güvence altına alınmıştır. GDPR, veri ihlallerine karşı ciddi cezalar öngörmekte olup, bu nedenle dünya çapında işletmelerin veri koruma politikalarını güçlendirmesine neden olmuştur ("GDPR.EU", t.y.; Voigt ve Von dem Bussche, 2017).

3.2.2.35. ABD FISMA (Federal Information Security Management Act - Federal Bilgi Güvenliği Yönetim Yasası)

ABD Federal Bilgi Güvenliği Yönetimi Yasası (FISMA), federal kurumların bilgi güvenliği risklerini yönetmek ve bu risklere karşı gerekli önlemleri almak amacıyla 2002 yılında yürürlüğe girmiştir. FISMA, federal kuruluşlar ve onların yüklenicileri için bilgi güvenliği politikaları geliştirme ve uygulama zorunluluğu getirir. Bu yasa, ulusal güvenlik açısından kritik olan federal bilgi sistemlerinin korunmasını sağlamak için bir dizi güvenlik önlemi ve risk yönetimi süreci tanımlamaktadır. Ayrıca federal kurumlar için düzenli denetimler ve raporlama süreçleri oluşturur. FISMA, ABD hükümetinin siber güvenlik stratejisinin temel taşlarından biridir (“SDxCentral”, t.y.).

3.2.2.36. ABD HIPAA (Health Insurance Portability and Accountability Act - Sağlık Sigortası Taşınabilirliği ve Hesap Verebilirlik Yasası)

ABD Sağlık Sigortası Taşınabilirlik ve Sorumluluk Yasası (HIPAA), sağlık bilgilerinin gizliliğini ve güvenliğini korumayı amaçlayan bir düzenlemedir. 1996 yılında yürürlüğe giren HIPAA, sağlık hizmet sağlayıcılarının, sigorta şirketlerinin ve diğer ilgili kuruluşların bireylerin sağlık verilerini nasıl kullanacağını ve paylaşacağını düzenler. Bu yasa, sağlık bilgilerini izinsiz kullanıma karşı koruma altına alır ve bu bilgilerin güvenli bir şekilde işlenmesini sağlar. HIPAA, bireylerin sağlık bilgilerinin gizliliği konusunda önemli haklar tanımaktadır ve sağlık sektöründe faaliyet gösteren tüm kuruluşlar için bağlayıcıdır (“CDC Public Health Law”, 2024).

3.2.2.37. ISAD(G) (General International Standard Archival Description - (Genel Uluslararası Standart Arşivsel Tanım)

ISAD(G) arşiv belgelerinin tanımlanmasını standart hale getiren uluslararası bir standarttır. Bu standart, arşivlerdeki belgelerin yapılandırılmış bir şekilde tanımlanmasını sağlayarak, belgelerin erişilebilirliğini ve kullanılabilirliğini artırır.

Arşiv tanımlamalarının belirli bir şekilde yapılması, farklı arşivlerin birlikte çalışabilirliğini kolaylaştırır. Bu standart, özellikle büyük arşiv koleksiyonlarının yönetiminde ve belgelerin bulunabilirliğinin artırılmasında önemli bir rol oynar. Bu standartta zorunlu olarak tanımlanması gereken referans kodu, başlık, oluşturan adı, oluşturma tarihleri, tanımın kapsamı ve tanım seviyesi olmak üzere 6 alan bulunmaktadır. Bunlara ek olarak erişim koşulları, belgenin dili ve kapsam ve içerik alanları da bulunmaktadır (“Jisc Archives Hub”, t.y.).

3.2.2.38. EAD (Encoded Archival Description - Kodlanmış Arşivsel Tanımlama)

EAD (Kodlanmış Arşivsel Tanım) arşiv koleksiyonlarını tanımlamak için kullanılan XML tabanlı bir standarttır. Bu standart arşiv materyallerinin elektronik ortamda düzenlenmesini ve tanımlanmasını sağlar. EAD da ISAD(G) gibi arşivlerin dijital erişimini artırarak, kullanıcıların bu koleksiyonları daha kolay keşfetmesine böylece koleksiyonların daha fazla kullanılmasına olanak tanır (“University of Nebraska”, t.y.).

Bu mevzuatlar ve standartlar hem kurumsal belge yönetimi hem de dijital arşiv yönetimi için kapsamlı bir çerçeve sunar ve bilgi güvenliği, erişilebilirlik ve sürdürülebilirlik ilkelerine uygun olarak belgelerin ve verilerin yönetilmesini sağlar.

3.2.2.39. Prosci Change Management (Prosci Değişim Yönetimi)

Prosci Change Management, bireylerin kurumsal değişim süreçlerine adaptasyonunu önceliklendiren ve bu sürecin başarılı bir şekilde yönetilmesini hedefleyen bir değişim yönetimi metodolojisidir. 1994 yılında Prosci (Professional Science) adlı ABD merkezli bir araştırma şirketi tarafından geliştirilmiştir. Bu metodoloji, kapsamlı bir araştırma temeline dayanmaktadır ve değişim süreçlerini hem bireysel hem de kurumsal düzeyde ele alarak organizasyonların hedeflerine ulaşmasını desteklemektedir.

Prosci'nin temel aracı olan ADKAR modeli, bireylerin değişim yolculuğundaki ihtiyaçlarını anlamayı ve yönetmeyi sağlamaktadır. Model, bireylerin değişimi benimsemesini etkileyen beş kritik aşamayı tanımlamaktadır: Farkındalık (Awareness), istek (Desire), bilgi (Knowledge), yetenek (Ability) ve pekiştirme (Reinforcement). Bu model, değişim sürecindeki bireysel engelleri tespit etmeye ve bu engelleri aşmaya yönelik çözümler geliştirmeye odaklanmaktadır (Hiatt, 2006).

Prosci değişim yönetimi yaklaşımı, kurumların değişim projelerinde insan faktörünü başarıyla yönetmesine yardımcı olmaktadır. Özellikle veri yönetimi gibi karmaşık teknolojik dönüşüm süreçlerinde, çalışanların yeni sistemlere ve süreçlere adaptasyonunu kolaylaştıracaktır (Hiatt, 2006; Prosci, 2020). Mevcut veri yönetim sistemlerinden veri gölü gibi yenilikçi altyapılara geçişte Prosci'nin metodolojisi, değişim sırasında ortaya çıkabilecek dirençleri yönetmek ve gerekli eğitim programlarını planlamak için etkili bir çerçeve sunmaktadır.

3.2.2.40. DMBOK (Data Management Body of Knowledge - Veri Yönetimi Bilgi Kitaplığı)

DMBOK, veri yönetimi alanında standart bir rehber ve referans kaynağı olarak DAMA International (Data Management Association) tarafından geliştirilmiştir. İlk olarak 2009 yılında yayımlanan DMBOK 1.0, veri yönetimi disiplinlerini tanımlayarak veri profesyonellerine rehberlik etmek için tasarlanmıştır. Daha sonra, teknolojideki gelişmeler ve artan veri hacimleri nedeniyle, 2017 yılında DMBOK 2.0 sürümü yayımlanmıştır. Bu güncel sürüm, büyük veri, veri gölleri, veri analitiği ve veri yönetişimi gibi modern veri yönetimi kavramlarını kapsamaktadır (DAMA International, 2017).

DMBOK, organizasyonların verilerini daha etkili bir şekilde yönetmelerine, stratejik kararlar almalarına ve uyumluluk gerekliliklerini karşılamalarına yardımcı olmaktadır. Veri yönetişimi, veri kalitesi, veri mimarisi, üst veri yönetimi, veri

yönetimi ve veri güvenliği gibi kritik alanları kapsamaktadır. Aynı zamanda, veri yönetimi terminolojisi ve metodolojileri için bir standart sağlamaktadır (Ladley, 2020).

3.2.2.41. Uluslararası Arşivler Konseyi (International Council on Archives - ICA)

Uluslararası Arşivler Konseyi (International Council on Archives - ICA), arşivlerin ve verilerin etkin yönetilmesini, kullanılmasını ve insanlığın kültürel mirası olarak korunmasını amaçlayan uluslararası bir organizasyondur. ICA, arşivlerin gelişimini teşvik etmeyi, en iyi uygulamaların ve standartların geliştirilmesini desteklemeyi ve arşivciler ile bilgi yöneticileri arasında ilişkiler kurmayı hedeflemektedir. Ayrıca arşiv üzerine çalışan kurumlarının çalışmalarını destekleyerek arşivlerin kullanımını ve erişilebilirliğini artırmaya çalışmaktadır. ICA'nın 199 ülkeden 1400'ün üzerinde üyesi bulunmaktadır. ICA, arşivlerin hesap verebilirlik ve iyi yönetişim için temel oluşturduğunu vurgulamaktadır. Bu nedenle uluslararası işbirliğini teşvik ederek, profesyonel deneyimleri ve araştırmaları paylaşarak arşivlerin ve arşiv üzerine çalışan kurumların gelişimine katkıda bulunmaktadır ("ICA" , t.y.).

3.2.3. Ulusal Veri Sözlüğü Sistemi

Türkiye Cumhuriyeti Cumhurbaşkanlığı, Strateji ve Bütçe Başkanlığı tarafından 2019 yılında onaylanan "On Birinci Kalkınma Planı (2019-2023)" kapsamında ulusal veri sözlüğü ile ilgili olarak "e-Devlet Uygulamaları - Politika ve Tedbirler" başlığı altında "Ulusal veri envanteri ve kamu veri sözlüğü hazırlanacaktır" ifadesi bulunmaktadır (Türkiye Cumhuriyeti Cumhurbaşkanlığı, 2019, s. 183). Bu bağlamda 2019 yılında Türkiye Cumhuriyeti Cumhurbaşkanlığı, Dijital Dönüşüm Ofisi tarafından çalışmalara başlanmıştır ("Türkiye Cumhuriyeti Cumhurbaşkanlığı", t.y.b).

Veri sözlükleri kısaca, veriye ilişkin üst verinin tanımlı ve yönetilebilir olmasını sağlamak amacıyla "Üst Veri Kayıt Defteri" oluşturulmasıdır (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2019a, s. 13). Veri sözlükleri, bilgi sistemleri ile ilgili entegrasyon, mükerrer/çelişen veri, veri sahibinin belirsizliği vb. problemlerin çözümü için kritik öneme sahiptir ("Türkiye Cumhuriyeti Cumhurbaşkanlığı", t.y.b).

Ülkemizde Ulusal Veri Sözlüğü Sistemi, kurumların veri standardizasyonunun sağlanması, veri paylaşımının kolaylaşması ve ulusal veri envanterinin oluşturulması, böylece daha büyük projeler için gerekli altyapının sağlanması amacıyla tanımlanmıştır.

Bu kapsamda "ulusal veri envanterinin çıkarılması", "veri sahipliğinin belirlenmesi", "ulusal veri entegrasyon mimarisi ile yönetim ve izleme süreçlerinin yapılandırılması", "terminoloji birliğine varılması ve kurumsal hafızanın oluşturulması", "ulusal veri modellerinin belirlenmesi", "yeni yazılımlar için süreç iyileştirmeleri" ve "merkezi servis tasarım platformunun oluşturulması" hedeflenmiştir ("Türkiye Cumhuriyeti Cumhurbaşkanlığı", t.y.b). Bu bağlamda, veri sözlüğü oluşturmak amacıyla dünya genelinde farklı organizasyonlar tarafından kullanılan ISO/IEC 11179 standardı, ülkemizde TS ISO/IEC 11179 adı altında Türkçe'ye çevrilerek yayımlanmış (Veri Sözlüğü Oluşturma Metodolojisi, 2019) ve bu standarda uygun bir veri sözlüğü portalı açılmıştır ("Türkiye Cumhuriyeti Cumhurbaşkanlığı", t.y.b).

Bu standarda uygun olarak kurumların kendi veri sözlüklerini hazırlamaları istenmiş, veri sözlüğünün nasıl oluşturulacağı ile ilgili olarak "Veri Sözlüğü Hazırlama Metodolojisi" dokümanı oluşturulmuş ve kurumlarla paylaşılmıştır (*T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi*, 2019a).

Ulusal Veri Sözlüğü ile verilerin tekil hale gelmesi, böylece entegrasyonun kolaylaşması, tekrarlı, çelişen ve/veya tutarsız verilerin engellenmesi, üçüncü kişilere veya kurumlara bağımlılığın giderilmesi ve bilgiye erişim kolaylığı sağlanmış olur ("Türkiye Cumhuriyeti Cumhurbaşkanlığı", t.y.b). Ulusal Veri

Sözlüğü ile verinin anlam ve ilişkilerinin ortaya konması sağlanır böylece verinin standartlaşması, paylaşımı, hizmetlerin bütünleşmesi ve merkezi hizmet platformunun oluşturulması için gerekli temel sağlanmış olur (T.C. *Cumhurbaşkanlığı Dijital Dönüşüm Ofisi*, 2019a, s. 13).

Veri sözlüğü oluşturulması için öncelikle kurumlarda bu anlayışın oluşturulması, gerekli süreçlerin belirlenmesi ve ilgili ekiplerin oluşturulması, Üst veri Kayıt Defteri ile ilgili kuralların belirlenmesi, sonrasında üst verilerin toplanarak kayıtların oluşturulması ve yönetilmesi gerekir. Tüm bu süreç ve işlemlerin bir yazılım desteği ile yürütülmesi ile farklı kurumlarda aynı kurallara uyulmasına ve süreçlerin daha kolay yürütülmesine katkı sağlayacaktır (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2019a, s. 14). Bu nedenle “Veri Sözlüğü Oluşturma Metodolojisi” dokümanında detayları ile bilgi sistemlerinde karşılaşılabilecek problemler ve bu problemlerin çözümüne yönelik öneriler, üst veri yönetim standartları, veri sözlüğü bileşenleri ile ilgili tanımlar, veri sözlüğü oluşturma adımları, Üst veri Kayıt Defterinin oluşturulması ve ilgili yazılımın özellikleri konuları açıklanmıştır (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2019a, s. 14).

4. BÖLÜM
VERİ MİMARİSİ GELİŞTİRME

Veri mimarisi, günümüzün bilgi çağında kurumlar için kritik bir stratejik unsurdur. Genel olarak bir veri stratejisi çerçevesi sunar. Verilerin nasıl elde edildiği, taşındığı, depolandığı, güvenliğinin sağlandığı ve veri üzerinde sorgulama yapıldığı gibi süreçleri kapsar. Veri mimarisi, verinin daha iyi anlaşılmasına katkıda bulunur ve verinin oluşumundan kullanımına kadar olan veri yönetim süreçlerine ilişkin talimatları içerir. Böylece, veri yönetiminin geliştirilmesine yönelik bir yapı sağlar. Veri yönetimine yönelik süreçler, verilerin toplanmasını, kullanılmasını, işlenmesini, depolanmasını ve farklı sistemlerle entegrasyonunu kapsamaktadır (Hernandez ve Roberts, 2022).

Veri yönetimi teknolojilerinde son yıllarda büyük gelişmeler olmuştur ancak bu gelişmeleri kısıtlayan depolamanın çok pahalı olması veya büyük miktarda veri depolayacak kapasitenin bulunmaması gibi problemler ortaya çıkmıştır. Büyük veri teknolojileri bu problemlerin aşılmasında önemli bir adım olmuştur. Hadoop Dağıtılmış Dosya Sistemi (HDFS), birçok bilgisayar kümesi arasında büyük veri kümelerinin dağıtılarak depolanıp işlenmesi amacıyla tasarlanmıştır. Böylece veri depolama maliyet ve teknik kısıtlılıkları büyük ölçüde azalmış, veri işlemeye yönelik yeni fırsatlar oluşmuştur (Inmon, 2016, s. 15). Büyük veri teknolojilerinin sunduğu çözümler, sadece depolama ve işleme süreçlerini kolaylaştırmakla kalmamış aynı zamanda bu verilerin etkin bir şekilde yönetilmesi ve stratejik olarak kullanılabilmesi için kapsamlı bir veri mimarisi oluşturma gerekliliğini de beraberinde getirmiştir (Inmon, 2016). Literatür ile desteklenen bu gerekliliğin ortaya çıkması, tezdeki hipotezlerden biri olan "Kamu kurumlarının sistematik bir veri mimarisi politikası dahilinde veri yönetimi süreçlerini uygulamaması, kurumsal verilerin etkili kullanımını engellemektedir" ifadesini destekler niteliktedir. Bu hipotez, kamu kurumlarının veri yönetimi politikalarında ortaya çıkan eksikliklerin, büyük veri çağında stratejik karar alma süreçlerini sınırladığına işaret etmektedir.

Veri mimarisi modeli, kısaca bir organizasyonun uzun vadeli veri saklama ve veriye erişim sağlama ihtiyaçlarını karşılamak için kullandığı, verinin güvenli bir şekilde depolanmasını sağlayan, veri kaynakları, depolama altyapısı, veri kategorizasyonu, yedekleme, veri erişimi ve imha gibi bileşenleri olan bir model olarak tanımlanabilir. Veri mimarisi modeli ile organizasyonların büyük miktarda veriyi yönetmesi, uzun vadeli saklanması ve gerek duyulduğunda bu verilere erişimi, böylece verinin karar verme süreçlerinde kullanımı sağlanır.

Günümüzde çok fazla veri üretilmekte ve şirketler çok fazla veri depolamakta, büyük veri gölleri (data lake) oluşturmaktadırlar. Ancak bu büyük veri göllerinden bilgi elde etmek her zaman mümkün olmamakta veya küçük bir bilgiye ulaşmak için yoğun çaba gösterilmesi gerekebilmektedir. Kurumlar büyük bütçe ve emek harcayarak oluşturdukları veri göllerinin tek yönlü olduğunu yani göle sadece veri aktardıklarını ama gölden hiçbir şey elde edemediklerini fark ettiklerinde veri gölleri birer büyük çöplük alanına dönüşmüş olmaktadır. Verinin değeri büyüktür ancak kullanılabilmesi için veri göllerinin belirli kurallar dahilinde oluşturulması önemlidir (Inmon, 2016, s. 13, 14). Bu yüzden, veri sağlama aşamasında mümkün olduğunca fazla üst veri çıkarılması önemlidir. Eğer kaynaklardan yeterince üst veri otomatik olarak çıkarılamıyorsa, bir uzman bu kaynaklar hakkında ek bilgi sağlamak zorunda kalabilir (Hai, Quix ve Jarke, 2021, s. 4). Verinin etkin bir şekilde kullanılamamasının, rastgele depolama, entegrasyon eksikliği ve kolayca analiz edilemeyecek şekilde kaydedilmesi gibi çeşitli nedenleri bulunmaktadır. Ancak veri göllerinin şekillendirilmesi için gerekli özen gösterilirse, veri gölleri önemli varlıklara dönüşebilir (Inmon, 2016, s. 13, 14).

4.1. VERİ MİMARİSİ MODELİ

Veri mimarisi modelleri genel olarak, verilerin nasıl depolandığını, yönetildiğini ve kullanıldığını açıklayan planlardır. Veri mimarisi modellerinin oluşturulmasıyla kurumların veri yönetim süreçlerinin net bir şekilde anlaşılması sağlanarak veri yönetim ve karar verme süreçlerine katkı sağlanır.

Veri mimarisi modelleri ve veri yönetim planları ile kurumlarda oluşan verilerin uygun bir şekilde saklanması, yedeklenmesi, güvenliği ve gerektiğinde yeniden kullanımı sağlanır. Kurumların veri yönetim planları geliştirilirken öncelikle toplanan ve süreçlerde oluşan verilerin türleri ve kaynakları belirlenmelidir. Daha sonra bu verilerin nasıl toplanacağı ve arşiv için nasıl hazırlanacağı (dönüşüm ve entegrasyon), kategorizasyonun (düzenleme ve indeksleme) nasıl yapılacağı, depolama altyapısı, yedeklemenin nasıl yapılacağı, veriye nasıl erişim sağlanacağı, nasıl sorgulama yapılacağı, veri muhafaza ve imha süreçlerinin hangi durumlarda yapılacağı ve bu süreçlerin nasıl yürütüleceği belirlenmelidir (Tupper, 2011).

Veri mimarisi modellerinin geliştirilmesi ile veri yönetiminin iyileştirilmesi, veri kalitesinin iyileştirilmesi, veri paylaşımı ve iş birliğinin geliştirilerek veri tekrarının önüne geçilmesi, veriye daha kolay ve hızlı erişim sağlanması, verinin karar verme süreçlerinde kullanımının artması sağlanabilir.

Kamu kurumları tarafından toplanan ve kullanılan verilerin düzenlenmesine ve yönetilmesine yardımcı olacak temel bir çerçevenin belirlenmesi bu verilerin etkili bir şekilde kullanımı için önemlidir (Eroğlu ve Külcü, 2014). Genel olarak toplanan verilerin depolanma ve kullanım şeklinin belirlenmesi gerekir. Bu süreçte veri yönetişimi (veri süreçlerini yönetecek politika ve prosedürler), veri güvenliği (verilerin yetkisiz erişim, kullanım, değişiklik, ifşa ve imhaya karşı korunması), veri kalitesi (verilerin doğru, güncel ve eksiksiz olması), veri erişimi (veri okuma, yazma, güncelleme vb.) ve veri paylaşımı (kurum içi veya farklı kurum ile) önemli unsurlardır.

Giebler ve diğerlerinin (2021) geliştirdiği veri gölü mimari çerçevesi, veri süreçleri (data processes), veri organizasyonu (data organization), veri modelleme (data modeling), veri akışı (data flow), veri depolama (data storage), altyapı (infrastructure), üst veri yönetimi (metadata management), veri güvenliği ve gizliliği (data security & privacy), veri kalitesi (data quality) olmak üzere dokuz bileşenden oluşmaktadır. Bu bileşenler aşağıda kısaca açıklanmıştır:

Altyapı (Infrastructure): Veri gölünün fiziksel olarak nasıl uygulanacağını ele alır. Dağıtılmış dosya sistemi HDFS veya ilişkisel veri tabanı MySQL gibi depolama sistemlerine ve bu sistemlerin bulutta, şirket içi veya hem bulut hem de şirket içi ortamları birleştiren hibrit olarak nasıl dağıtılacağına odaklanır (Giebler ve diğerleri, 2021, s. 354).

Veri Depolama: Veri depolama ve işleme için kullanılan dağıtılmış dosya sistemleri veya NoSQL veri tabanları gibi yapıları içerir. Hangi aracın kullanılacağı değil, tek bir dosya sisteminin mi yoksa farklı depolama sistemlerinin mi kullanılacağı belirlenir (Giebler ve diğerleri, 2021, s. 355).

Veri Akışı: Genellikle büyük hacimli toplu verilerin ve hemen işlenmesi gereken gerçek zamanlı verilerin akışını yönetmek için kullanılan mimari ve yöntemleri ele alır.

Veri Modelleme: Verinin özelliklerine ve kullanım amacına uygun olarak nasıl modelleneceğinin belirlenmesidir. Veri göllerinde veri damlaları (data droplets) veya veri kasası (data vault) modelleri kullanılabilir.

Veri Organizasyonu: Veri organizasyonu, bir kuruluşun verilerini etkili bir şekilde yönetmesi, depolaması ve kullanması için yapılandırma sürecidir. Bu süreç, verilerin anlamlı ve kullanışlı bir şekilde düzenlenmesini, sınıflandırılmasını ve ilişkilendirilmesini içerir. Veri organizasyonu için çeşitli yaklaşımlar ve mimariler kullanılabilir. Etkili veri organizasyonu, veri kalitesini artırır, erişilebilirliği iyileştirir ve veri analitiği girişimlerinin başarısını destekler.

Veri Süreçleri: Verilerin taşınması ve işlenmesi ile ilgili süreci kapsar. Verileri oluşturma ve elde etme aşamasından elden çıkarma aşamasına kadar olan süreçleri kapsayan veri yaşam döngüsü yönetimi süreçleri ile çıkarma-dönüştürme-yükleme (ETL) işlemleri gibi verilerin alımına, taşınmasına ve işlenmesine odaklanan veri hattı işlemleri bu kapsamda değerlendirilir. Veri yaşam döngüsü yönetimi süreçlerinin veri kullanılabilirliğini sağlamak ve yasal uyumluluğu bakımından veri gölü için tanımlanması ve standartlaştırılması

önemlidir. Veri hattı işlemleri ise örneğin verilerin bölge mimarisi modelinde bölgeler arasında nasıl hareket ettiğini tanımlamak için kullanılır. “Veri Akışı” verinin özelliklerine odaklanırken, “Veri Süreçleri” verilerle ne yapıldığıyla ilgilidir.

Üst Veri Yönetimi: Üst verilerin iki alt yönünü içerir. Birincisi örneğin bir alan mimarisinde verinin hangi alana ait olduğunu veya veri yaşam döngüsünün hangi zamanında oluştuğunu gösteren etkinleştirici olarak üst veridir. Diğeri ise anlamsal açıklamalar veya iş sözlükleri gibi üst verinin etkinleştirici yeteneklerine ek olarak sağlayabileceği işlevleri kapsar. Constance gibi üst veri yönetim sistemleri veya HANDLE gibi üst veri modelleri örnek olarak verilebilir.

Veri Güvenliği ve Gizliliği: Yasalara ve iş hedeflerine uygunluğun sağlanmasını hedefleyen kavramsal bir alandır. Veri işleme sürecinde yasalara uyumluluğun kontrol edilmesi veya AMNESIA gibi araçlarla GDPR uyumlu makine öğreniminin uygulanması örnek gösterilebilir.

Veri Kalitesi: Veri gölünde verilerin kullanılabilirliğini sağlamak ve veri gölünün veri çöplüğüne dönüşmesini önlemek için önemli olan veri kalitesi kavramsal bir alandır. “Informatica data quality 7” gibi veri kalitesi araçları örnek gösterilebilir ancak bu bileşen üst veri yönetimi veya veri organizasyonu gibi diğer bileşenlerin uygulamalarına dayanır (Giebler ve diğerleri, 2021, s. 354-356).

4.2. VERİ GÖLLERİ

Veri gölleri, büyük veri çağında esnek bir veri depolama çözümü olarak ortaya çıkan ve veri yönetiminde önemli rol oynayan yapılardır. İlk olarak 2010 yılında James Dixon tarafından ortaya atılan “veri gölü” kavramı (Dixon, 2010), her yapıda veriyi - yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış - ham haliyle ve şemasız bir şekilde depolayabilen geniş veri alanları olarak ifade edilmektedir (Dixon, 2010; Inmon, 2016). Son yıllarda dijital dönüşümle birlikte, dünya genelinde üretilen veri miktarında büyük bir artış gözlenmektedir. Nesnelerin interneti (IoT), sosyal medya ve diğer dijital platformlardan gelen

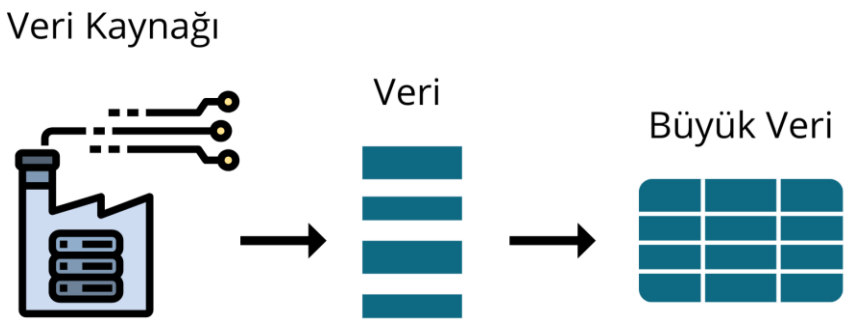
büyük veri, hızla artan hacmi, çeşitli formatları ve yüksek üretim hızıyla geleneksel veri yönetim sistemlerinin sınırlarını zorlamaktadır (Sawadogo ve Darmont, 2021).

Geleneksel veri ambarları, büyük verinin hacim, hız, çeşitlilik, doğruluk ve değer özniteliklerinin yönetilmesinde yetersiz kalmakta (Miloslavskaya ve Tolstoy, 2016), özellikle yapılandırılmamış ve yarı yapılandırılmış veriler için yeterli çözümü sağlayamamaktadır. Bu bağlamda, büyük veri sorunlarını çözmek amacıyla veri gölü kavramı önem kazanmıştır. Veri gölleri, farklı formatlardaki ham veriyi depolama ve yönetme kapasitesine sahip, geniş ve esnek veri depolama sistemleridir (Sawadogo ve Darmont, 2021, s. 97). Bilişim teknolojilerindeki gelişmeler ve veri depolama maliyetlerinin azalmasıyla birlikte veri gölleri, kuruluşların ürettikleri veya topladıkları büyük veriyi depolamak için tercih ettikleri temel yapı haline gelmiştir. Veri gölleri, depoladıkları verileri daha sonra işlenmek üzere ham halde bekletilebilen esnek yapıları sayesinde büyük veri analitiği, makine öğrenimi ve veri bilimi gibi veri yoğun uygulamalarda ideal bir çözüm sunmaktadır. Literatürde veri gölü kavramı ile eş anlamlı olarak "veri rezervuarı (data reservoir)" (Chessell, Scheepers, Nguyen, van Kessel ve van der Starre, 2014) ve "veri merkezi (data hub)" ("Esds", 2021) kavramları da kullanılmaktadır.

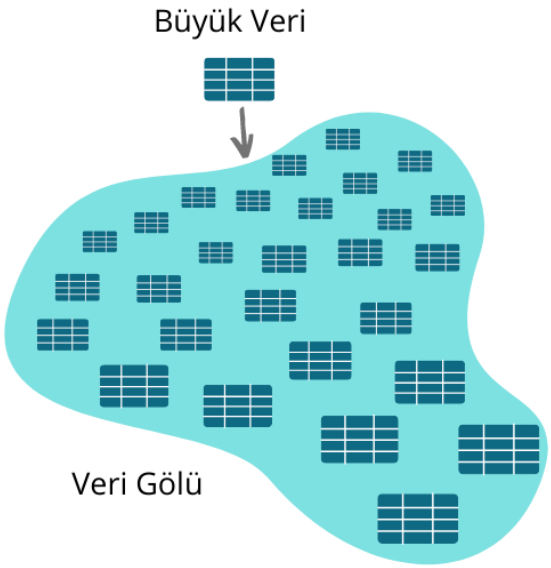
Veri gölleri geleneksel veri ambarlarının aksine, yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verileri bir arada tutarak, verilerin ham haliyle depolanmasına ve daha sonra analiz için uygun hale getirilmesine olanak tanımaktadır (Madera ve Laurent, 2016). Bu esneklik, veri bilimcilerin daha geniş veri setlerine erişimine olanak sağlayarak onlara veri üzerinde daha derinlemesine analiz yapma imkânı sunmaktadır. Veri gölleri, schema-on-read prensibine dayandığı için, verilerin depolama sırasında dönüştürülmesine gerek kalmadan saklanmasını sağlamakta, böylece büyük veri yönetiminde önemli bir avantaj sunmaktadır (Mathis, 2017). Literatürden destek alınarak oluşturulan bu bilgiler, tezde yer alan “Veri gölü mimarilerinin uygulanması, kamu kurumlarının veri analiz kapasitesini artırabilir” hipotezini desteklemektedir.

Veri göllerine geçiş, veri yönetimi ve analizi alanında yeni fırsatlar yaratırken aynı zamanda kurumların geçiş süreçlerini verimli yönetebilmesi için yeni zorluk ve gereklilikleri de beraberinde getirmektedir. Veri gölü mimarilerinin, özellikle karmaşık yapılı ve büyük hacimli verileri yönetmesi geçiş sürecinde dikkate alınması gereken en önemli unsurlardan biridir. Birçok araştırmacı ve uygulayıcı, veri göllerinin Hadoop teknolojisiyle özdeşleştirilmesi nedeniyle tam anlamıyla anlaşılmadığına dikkat çekmektedir (Grosser, Bloeme, Mack ve Vitsenko, 2016). Büyük verilerin dağıtılmış bir şekilde işlenmesi için yaygın olarak kullanılan bir teknoloji olan Hadoop, veri gölü uygulamalarında sıkça tercih edilmektedir. Ancak, veri gölleri yalnızca Hadoop altyapısıyla sınırlı değildir. Hem açık kaynaklı platformlar (örneğin Apache Hadoop) hem de ticari bulut tabanlı çözümler (örneğin Microsoft Azure ve IBM) kullanılarak uygulanabilir. Böylece belirli bir teknolojiye bağımlı kalmaksızın esnek bir yapı sunmaktadır (Madera ve Laurent, 2016). Ayrıca veri göllerinin yönetimi, özellikle üst veri yönetimi ve veri yönetişimi gibi alanlarda büyük önem taşımaktadır. Doğru yönetilmeyen veri gölleri, veri bataklığına (data swamp) dönüşme riski taşımaktadır. Veri bataklığına dönüşmüş bir veri gölündeki verileri yönetmek mümkün olmayacaktır (Inmon, 2016; Suriarachchi ve Plale, 2016). Literatürle desteklenen bu bilgiler, tezde yer alan "Üst veri yönetimi ve veri yönetişimindeki eksiklikler, veri mimarilerinin etkinliğini sınırlandırmaktadır" hipotezini desteklemektedir.

Şekil 13'te büyük verinin oluşumunu gösteren model, Şekil 14'te ise veri gölü oluşumunu gösteren model verilmiştir.



Şekil 13. Büyük veri oluşumunu gösterir şema



Şekil 14. Veri gölü oluşumunu gösterir şema (Inmon, 2016)

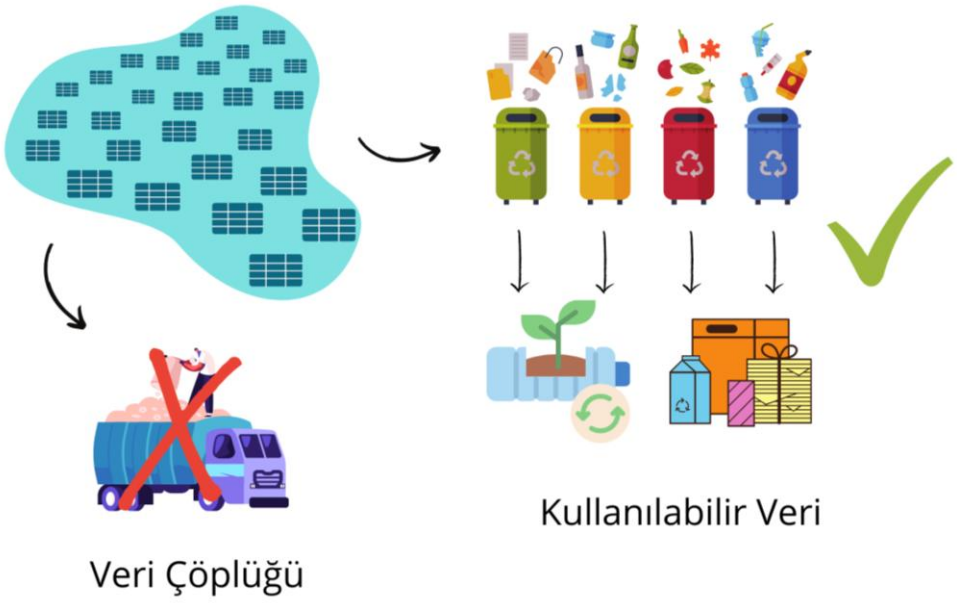
Veri gölleri endüstriyel veri analizi çözümlerinde kritik bir rol oynamaktadır ve araştırmalarda giderek daha fazla önem kazanmaktadır. Özellikle araştırma kurumları için merkezi veri yönetim sistemleri ve makine öğrenimi süreçlerinin temel bileşenleri olarak kullanılmaktadır (Wieder ve Nolte, 2022, s. 1).

Kuruluşlar, süreçlerini yürütürken farklı türde ve büyük miktarda veri oluşturur ve toplarlar. Bu veriler kuruluşun temel kaynakları olan kişiler, kuruluşlar ve varlıklar hakkındaki veriler; yeni iş istemleri, pazarlama kampanyaları gibi günlük faaliyetler hakkındaki veriler; web siteleri veya çağrı santralleri tarafından oluşturulan günlükler gibi bir sistem veya cihaz çalıştığında oluşturulan günlük verileri; referans verileri, sınıflandırmalar, temel performans göstergeleri (KPI'ler) veya haber akışları gibi türetilmiş veya harici verilerden oluşabilir (Chessell ve diğerleri, 2014, s.11).

Depolanan bu veriyi kullanarak faydalı sonuçlar elde etmek veri depolamak kadar kolay değildir. Şirketler bu konuda veri bilimciler ile çalışmaya başlamıştır fakat büyük veri, şirketler kadar veri bilimciler için de çok yeni ve keşfedilmemiş bir kavram olmuştur. Bu nedenle büyük veri analiz edildiğinde yanlış sonuçlar çıkarılmakta veya farklı sorunlarla karşılaşılabilmektedir. Şirketler veri göllerindeki

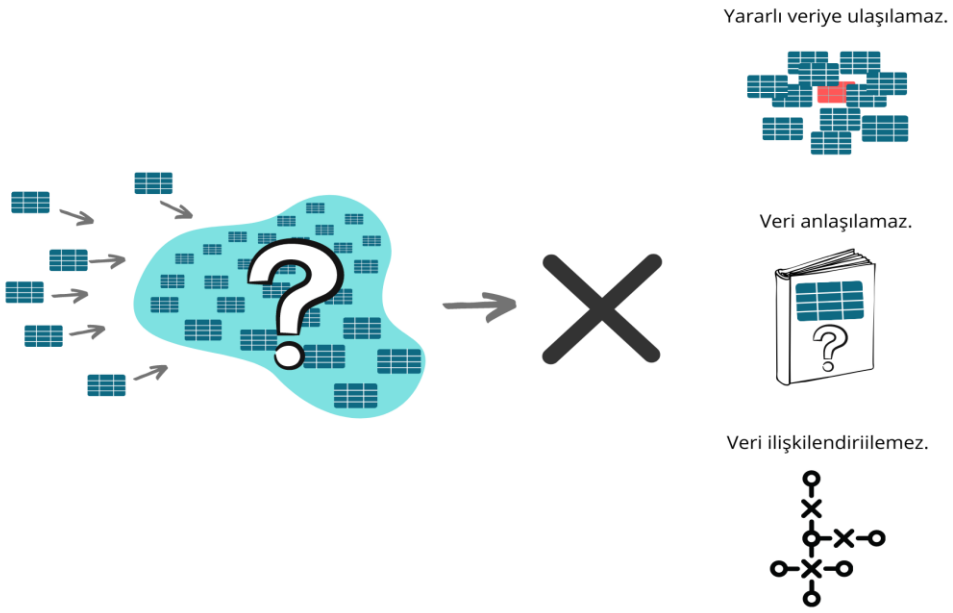
hazineyi değere dönüştürememenin hayal kırıklığını yaşamaya başlamıştır (Inmon, 2016, s. 16-18).

Veri göllerinin ana prensibi, veriyi orijinal formatında tutarak, geniş bir kullanım alanı sunması ve verinin yeniden kullanılabilirliğini artırmasıdır. Bu özellik, veri ambarlarında kullanılan ve veriyi depolamadan önce belirli bir şemaya dönüştüren *schema-on-write* yaklaşımına kıyasla büyük bir avantaj sağlamaktadır. Burada temel problem veri gölü büyüdükçe verileri analiz etmenin de daha zor olmasıdır. Büyük miktarda ham verinin depolanması, veri modelleme, indeksleme ve uygun hesaplama yeteneklerinin entegrasyonuna kadar çeşitli zorluklar barındırmaktadır (Wieder ve Nolte, 2022, s. 1). Veri göllerine sürekli bir veri akışı olmasına karşın bu verinin analiz edilerek bilgi elde edilmesi mümkün olamadığından bu durum veri göllerinin genellikle “tek yönlü göl” olarak anılmasına neden olmuştur. (Inmon, 2016, s. 16-18). Şekil 15‘te içine kontrolsüz veri iletilen veri gölünün veri çöplüğüne dönüşmesinin önüne geçilmesi için verinin bir plan ve düzen dahilinde gönderilmesi gerektiği, çöplerin ayrıştırılmadan atılarak bir çöplüğün parçası olması veya bir plan ve düzen dahilinde ayrılarak toplanması durumunda kullanılabilir ürüne dönüştürülebilmesine benzetilerek modellenmiştir.



Şekil 15. Veri çöplüğü (bataklığı) kavramını açıklayan görsel

Bu noktada, veri göllerinin etkin bir şekilde yönetilmesi ve sürdürülebilir hale getirilmesi için veri yerleştirme stratejilerinin ve analiz süreçlerinin dikkatle tasarlanması gerektiği ortaya çıkmaktadır. Veri gölleri ile, kurumsal veri süreçlerinde Bilgisayar Teknolojileri ekiplerinin çok az yardımı ile veya hiç yardımı olmadan kullanıcıların ihtiyaç duydukları verilere erişimini sağlamak amaçlanmaktadır (Chessell ve diğerleri, 2014, s.5). Bu sağlanamadığında veri gölleri büyüdükçe kuruluşlar için daha da işlevsiz hale gelmektedir. Veri göllerinin verilerin gelecekteki kullanımları için gerekli düzenlemeler yapılmadan yalnızca "döküldüğü" bir alan haline gelmesi önlenmelidir (Inmon, 2016, s. 19). Bu doğrultuda, veri göllerinin etkin bir şekilde yönetilebilmesi için yalnızca verilerin depolanması değil, aynı zamanda bu verilerin düzenlenmesi, ilişkilendirilmesi ve analiz edilebilir hale getirilmesi süreçlerine odaklanılması gerekmektedir. Bu tür bir yaklaşım, veri göllerinde yaşanan problemlerin önlenmesi ve kullanıcıların ihtiyaçlarına uygun çözümler sunulması açısından kritik bir adım olacaktır. Bu tür bir yaklaşım benimsenmediği takdirde, veri gölleri ile ilgili birtakım problemler ortaya çıkacaktır. Şekil 16'da mimari dikkate alınmadan oluşturulan veri gölleriyle ilgili temel problemlerin görselleştirildiği bir şema sunulmuştur.



Şekil 16. Mimari dikkate alınmadan oluşturulan veri gölleri ile ilgili temel problemler (Inmon, 2016)

Veri gölleriyle ilgili problemlerden biri, yararlı verilerin ilgili olmayan diğer büyük veri yığınlarının arkasında kalmasıdır; bu durum analistlerin ihtiyaç duydukları verilere ulaşmalarını zorlaştırmaktadır. Bir diğer önemli problem ise, veri gölündeki verilere ait üst verilerin erişilebilir olmamasıdır. Bu eksiklik, analistlerin ulaştıkları verinin anlamını ve kaynağını bilememesine yol açarak veri analizini riskli bir süreç haline getirmektedir (Sawadogo ve Darmont, 2021). Literatürle desteklenen bu bilgi, tezde yer alan "Üst veri yönetimi ve veri yönetişimindeki eksiklikler, veri mimarilerinin etkinliğini sınırlandırmaktadır" hipotezini destekler niteliktedir.

Veri gölleriyle ilgili bir başka sorun ise, veri ilişkilerinin kaybolması ya da bu ilişkilerin hiç tanımlanmamış olmasıdır. Veri göllerinin büyüklüğü nedeniyle, veri ilişkilerinin tanımlanması ve veri gölüne aktarılması çoğu zaman külfetli bir süreç olarak değerlendirilmektedir (Inmon, 2016, s. 21). Veri gölleri uygun şekilde oluşturulduğunda, analistlere, veri bilimcilerine ve iş ekiplerine veri sağlamada esneklik sunar; büyük verilerin çıkarılması ve depolanmasında verimlilik sağlar; verilerin korunması ve yönetiminde güvenilirlik sağlar; ayrıca performansı ve etkiyi artırmak için gelişmiş analitik seçenekler sunarak veri yönetimini kolaylaştırır (Chessell ve diğerleri, 2014, s.22).

4.2.1. Veri Göllerinin Yapısı

Veri gölleri, schema-on-read yapısını kullanarak verilerin ham halde saklanmasına ve analiz sırasında şemalandırılmasına olanak tanır. Buna karşılık geleneksel veri ambarlarında kullanılan schema-on-write yaklaşımı, verilerin depolama öncesinde şemalandırılmasını gerektirir ve bu nedenle aynı esnekliği sunmaz (Madera ve Laurent, 2016). Veri gölleri, farklı veri türlerini esnek bir şekilde depolayabilme özelliği sayesinde, veri bilimcilerin aynı veri seti üzerinde çeşitli analizler yapmasını ve yeni veri kaynaklarını sisteme kolaylıkla entegre etmesini mümkün kılmaktadır.

Veri gölleri; yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verileri bir arada tutabilmektedir. Örneğin, IoT sensör verileri, sosyal medya verileri veya metin tabanlı veriler gibi yapılandırılmamış veriler, veri göllerinde herhangi bir dönüştürme yapılmadan depolanabilir ve daha sonra analiz sırasında şemalandırılabilir (Sawadogo ve Darmont, 2021). Bu özellik, veri göllerine gerçek zamanlı veri akışına olanak tanımaktadır ve IoT verileri veya sosyal medya verileri gibi sürekli veri üreten sistemlerle entegrasyon açısından önemlidir (Miloslavskaya ve Tolstoy, 2016). Veri gölleri, geniş depolama kapasiteleri ile büyük veri kümelerini verimli bir şekilde saklayabilmektedir ve süreç genellikle Hadoop ve bulut tabanlı çözümlerle desteklenmektedir. Hadoop'un dağıtık dosya sistemi (HDFS - Hadoop Distributed File System), veri gölleri için ölçeklenebilir ve maliyet etkin bir depolama altyapısı sağlamaktadır (Grosser, Bloeme, Mack ve Vitsenko, 2016). Bu altyapı, veri göllerine depolanan verilerin büyük ölçeklerde saklanmasını mümkün kılmakta, ayrıca verilerin işlenmesi için paralel işlem gücü sağlamaktadır. Hadoop veya bulut tabanlı altyapılar sayesinde veri gölleri, yüksek hacimli verileri düşük maliyetle saklayabilmektedir. Böylece geleneksel veri ambarlarına kıyasla büyük veri kümelerini depolamak için daha uygun maliyetli bir çözüm sunmaktadır (Walker ve Alrehamy, 2015).

Ancak veri göllerinin bu avantajlarının yanı sıra bazı zorlukları da bulunmaktadır. En önemli zorluklardan biri, veri bataklığı riskidir. Veri göllerinin iyi yönetilememesi durumunda, verilerin karmaşık hale gelme ve kullanılmaz bir veri yığınına dönüşme riski bulunmaktadır. Bu durum etkili bir üst veri yönetimi ile önlenebilir (Sawadogo ve diğerleri, 2019). Üst veri, verilerin kaynağını, yapısını ve işleme amacını belirleyerek veri keşfi ve yönetimi sağlamaktadır. Üst veri yönetimi, veri göllerinde saklanan büyük miktardaki verilerin yönetilebilmesi için kritik öneme sahiptir. Üst veri yönetiminin ihmal edilmesi, veri gölünün verimsiz hale gelmesine neden olmaktadır (Suriarachchi ve Plale, 2016). Literatürle desteklenen bu bilgiler, tezde öne sürülen “Üst veri yönetimi ve veri yönetişimindeki eksiklikler, veri mimarilerinin etkinliğini sınırlandırmaktadır” hipotezini güçlendiren ve doğrulayan ek unsurlar sunmaktadır.

Veri güvenliğinin sağlanması da veri gölleri için büyük bir zorluktur. Veri gölleri büyük miktarda veri depoladığı için güvenlik riskleri de barındırmaktadır. Özellikle yapılandırılmamış verilerin güvenliği, geleneksel veri ambarlarına göre daha karmaşıktır. Bu nedenle veri göllerinde güvenlik politikalarının uygulanması kritik önem taşımaktadır (Terrizzano ve diğerleri, 2015). Bu güvenlik politikaları, erişim kontrolü, veri şifreleme ve sürekli izleme gibi önlemleri içermelidir.

Veri gölleri, daha önce mümkün olmayan analitik işlemlerin gerçekleştirilmesi, karmaşık analizler ve ileri düzey tahmin süreçleri için büyük potansiyel sunmaktadır (Fang, 2015). Ancak bu potansiyelin tam anlamıyla hayata geçirilebilmesi için veri göllerinin yalnızca veri bilimciler tarafından değil, organizasyon genelindeki geniş bir kullanıcı kitlesi tarafından da erişilebilir ve kullanılabilir olması gerekmektedir. Veri gölleri, entegre edildiklerinde ve genel kullanıma uygun hale getirildiklerinde, yöneticiler, sistem analistleri, muhasebe ve pazarlama gibi farklı uzmanlık alanlarından kullanıcılar tarafından kullanılabilir. Bu durum, veri gölünün kurumsal değerini ve faydasını önemli ölçüde artırmaktadır (Inmon, 2016, s. 31–32).

Veri gölünün kurumsal anlamda etkin bir kaynak haline dönüşmesi için, temel bazı bileşenlerin sağlanması gerekmektedir. Bu bileşenler arasında güçlü bir üst veri yönetimi, etkili güvenlik politikaları ve kullanıcı dostu araçların geliştirilmesi bulunmaktadır. Aksi takdirde, veri gölü potansiyel bir değer kaynağı olmaktan çıkıp yönetilemez bir yapıya dönüşebilir (Inmon, 2016, s. 24).

4.2.2. Veri Göllerinin Bileşenleri

Veri gölleri, farklı işlevsel bileşenlerden oluşmaktadır ve bu bileşenler, verilerin işlenmesi, depolanması ve analiz edilmesine hizmet etmektedir. Veri göllerinin temel bileşenleri şunlardır:

Veri Sağlama: Veri gölleri, birçok farklı kaynaktan veri alabilmektedir. Bu kaynaklar, yapılandırılmış veri tabanlarından gelen veriler, sensörlerden gelen

gerçek zamanlı veriler, sosyal medya verileri, IoT cihazlarından gelen veriler vb. farklı veri türlerini içerebilmektedir. Verilerin alındığı bu süreç, verilerin bir araya getirilmesini ve ham haliyle depolanmasını sağlamaktadır (Terrizzano ve diğerleri, 2015).

Depolama: Veri gölleri, büyük miktarlarda veriyi ölçeklenebilir bir şekilde depolamak için genellikle HDFS veya bulut tabanlı çözümler gibi geniş çaplı depolama sistemlerini kullanmaktadır (Bian ve Ailamaki, 2022). Veriler, bu depolama sistemlerinde ham veri olarak saklanmaktadır ve ihtiyaç duyuldukça işlenmektedir. Depolama altyapısı, veri gölünün esnek ve ölçeklenebilir olmasını sağlamaktadır.

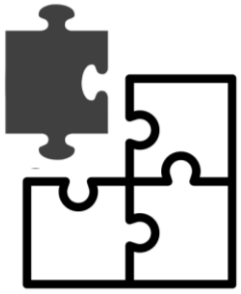
Veri İşleme ve Analitik: Veri göllerinde saklanan veriler, daha sonra ihtiyaç duyulduğunda işlenmekte ve analiz edilmektedir. Apache Spark, Hive veya Presto gibi büyük veri işleme araçları, veri gölleri içinde saklanan ham verileri hızlı bir şekilde işleyip analiz edebilmektedir. Bu süreçte, veri bilimi ve makine öğrenimi algoritmaları sıklıkla kullanılarak verilerin anlamlı bilgilere dönüştürülmesi sağlanmaktadır (Hai ve diğerleri, 2016).

Veri Güvenliği ve Yönetişimi: Veri göllerinde güvenlik ve veri yönetişimi, verilerin bütünlüğünü, gizliliğini ve uyumluluğunu sağlamak için kritik öneme sahiptir. Bu bileşen, erişim kontrolü, veri şifreleme, denetim izleri ve veri yaşam döngüsü yönetimi gibi unsurları içermektedir. Veri yönetişimi, veri kalitesini artırmak ve veri gölünün veri bataklığına dönüşmesini önlemek için gereklidir (Madera ve Laurent, 2016).

Veri Kullanımı: Bu bileşen, veri gölünden veri çıkışı ve kullanımı için arayüzler ve araçlar sağlamaktadır. Veri kullanımı; iş zekâsı araçları, veri görselleştirme platformları veya özel uygulamalar aracılığıyla gerçekleştirilebilmektedir. Bu bileşen, veri gölünden elde edilen içgörülerin ve analizlerin son kullanıcılara veya sistemlere iletilmesini sağlamaktadır.

Entegrasyon Haritalama: Entegrasyon haritası ile farklı uygulamalardan gelen verilerin birbirleriyle nasıl ilişkili olduğu, bu verilerin anlamlı bir şekilde nasıl

birleştirilebileceği açıklanır. Veri gölleri için üst veri oldukça kritik olmakla birlikte verilerin efektif kullanılabilmesi için yeterli değildir. Entegrasyon haritaları ile veri göllerindeki verilerin birbirleriyle nasıl entegre edilebileceği ayrıntılı olarak gösterilir, böylece verilerin veri gölünde izole olmasının önüne geçilir. Bu işlem olmadan farklı kaynaklardan gelen verilerin birbiriyle bütünleşmesi sağlanamaz. Bu durumda verilerin okunması ve yorumlanması çok zor bir hale gelir (Inmon, 2016, s. 25). Şekil 17'de veri entegrasyonu yapılmadan bir araya getirilmeye çalışılan verilerin izole olması, bir yapbozla uyumlu olmayan bir parçanın yerleştirilmesine benzetilerek modellenmiştir.



Şekil 17. Veri entegrasyonu yapılmadan veri gölüne veri eklenmesini açıklayan görsel

Entegrasyon haritalama ile, farklı kaynaklardan gelen verilerin birbiriyle bütünleşmesi sağlanır ve verilerin okunması, yorumlanması ve kullanılması kolaylaşır. Entegrasyon haritaları, verilerin anlamlı bir şekilde kullanılmasını sağlayarak veri göllerinin kurumsal değerini artırır. Bu süreç, veri tutarlılığını ve bütünlüğünü sağlamak için önemlidir (Fang, 2015). Entegrasyon haritaları, verilerin anlamlı bir şekilde kullanılmasını sağlayarak veri göllerinin kurumsal değerini artırır.

Bağlam: Veri göllerinin efektif olarak kullanılabilmesi için gerekli olan bileşenlerden bir diğeri ise verinin bağlamıdır. Bağlam verinin anlamını, önemini ve nasıl yorumlanması gerektiğini belirleyen çevresel ve durumsal bilgileri ifade eder. Bağlamı bilinmeyen veri anlamsızdır, hatta kimi durumlarda yanlış yorumlamalara yol açarak tehlikeli olabilir (Inmon, 2016, s. 26-27).

Veri göllerinde bağlam, üst veriden farklı ancak onunla yakından ilişkili bir kavramdır. Üst veri verinin yapısı, formatı ve temel özellikleri hakkında bilgi verirken, bağlam verinin içeriğini ve kullanım amacını anlamlandırır (Sawadogo, Kibata ve Darmont, 2020). Örneğin, "saha" sözcüğü ile bir futbol sahasından mı, yoksa bir inşaat sahasından mı bahsedildiğinin anlaşılabilmesi için metnin bağlamının bilinmesi gerekir (Inmon, 2016, s. 26-27). Bu bağlamsal bilgi, verinin doğru yorumlanmasını ve kullanılmasını sağlar.

Veri göllerinde bağlamın önemi, özellikle farklı kaynaklardan gelen çeşitli veri türlerinin bir arada bulunması nedeniyle daha da artmaktadır. Bağlamsal bilgiler, veri bilimcilerin ve analistlerin veriyi doğru şekilde analiz etmelerine ve anlamlı içgörüler elde etmelerine yardımcı olur (Giebler, Gröger, Hoos, Schwarz ve Mitschang, 2019). Ayrıca, bağlam bilgisi, veri gölündeki verilerin farklı iş birimleri ve kullanıcılar tarafından doğru şekilde anlaşılmasını ve kullanılmasını sağlar. Veri gölüne aktarılan verinin bağlamı ve bağlamının nasıl bulunacağı da veri ile birlikte veri gölüne aktarılmalıdır (Inmon, 2016, s. 26-27).

Üst (Meta) İşlem: Üst işlem bilgisi veri gölündeki bilgilerin nasıl işlendiği veya işleneceği ile ilgili bilgidir. Verilerin ne zaman, nerede, kim tarafından üretildiği, ne kadar veri üretildiği, veri gölüne yerleştirilen verilerin nasıl seçildiği, veri gölüne iletildikten sonra verilerin işlenip işlenmediği gibi bilgiler veri gölündeki verinin analiz edilme süreci için önemlidir. Bu özelliklerin veri ile birlikte veri gölüne dahil edilmesi gerekir. Veri aktarıldıktan sonra bu bilgilerin dahil edilmesi her zaman mümkün olmamaktadır. Tüm bileşenleri ile birlikte veri, veri gölüne aktarıldığında potansiyel bir altın madeni elde edilmiş olur.

Veri Kataloglama ve Üst Veri Yönetimi: Veri göllerinde saklanan büyük veri setlerinin yönetilebilmesi için etkili bir üst veri yönetim sistemi gereklidir. Üst veri, verilerin kaynağı, yapısı, işleme amacı vb. hakkında bilgi sağlamaktadır. Bu, veri gölleri içinde verilerin anlamlandırılmasını ve analiz için uygun hale getirilmesini kolaylaştırmaktadır (Suriarachchi ve Plale, 2016). Veri kataloglama, verilerin kolayca bulunabilmesi ve erişilebilmesi için kritik bir adımdır.

Üst veri, verinin tanımlanmasını, anlamlandırılmasını sağlayan, verinin kaynağını, yapısını ve diğer önemli özelliklerini tanımlayan veri hakkındaki veridir (“Türkiye İstatistik Kurumu”, t.y.).

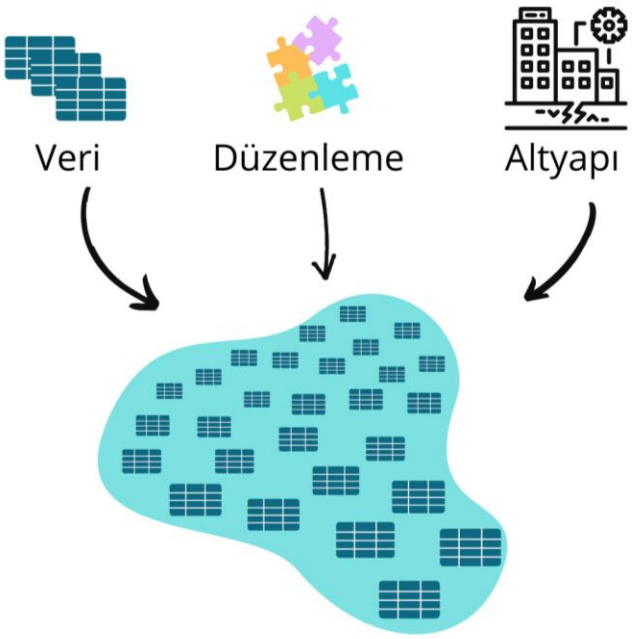
Üst veri kısaca verilerin açıklaması veya her veri koleksiyonunun kendisiyle ilgili temel yapısal bilgi olarak da ifade edilebilir. Üst veri kaydın tanım, dizin ve niteliklerini ayrıca verilerin farklı nitelikleri arasındaki ilişkileri içerir. Örneğin bir web sitesi ziyaretiyle ilgili üst veri ziyaret eden bilgisayarın IP adresini ve coğrafi konumunu içerir. Veri analistleri ancak verilere ait üst verileri kullanarak verimli analiz yapabilirler (Inmon, 2016, s. 24). Ham veriler üst verileri olmadan analiz edilmeye uygun değildir. Basit bir örnekle Wikipedia’daki makalelerin başlıkları olmadan arama yapmaya çalıştığımızı düşünebiliriz. Veriler ancak üst verileri ile etiketlendiğinde ve gölünde veri üst verisi ile birlikte depolandığında veriden fayda sağlanması mümkün olur (Inmon, 2016, s. 25).

Literatürde veri gölleri için üst veri fonksiyonel ve yapısal olarak ikiye ayrılmaktadır.

Fonksiyonel Üst Veri: İş süreç verileri; veri kaynağı, dosya boyutu, kayıt sayısı vb. operasyonel veriler ve teknik verileri içerir.

Yapısal Üst Veri: Verinin içeriğine dair bilgiler, verinin hangi nesnelerle bağlantılı olduğu gibi detayları içerir. Yapısal üst veri, verinin başka nesnelerle olan ilişkilerini ve benzerliklerini de kaydeder (Sawadogo ve diğerleri, 2019).

Şekil 18‘de bir veri gölünün altın madenine dönüşmesi için, içine gönderilen veri kadar veri gölünün düzeni ve altyapısının da önemli olduğunu gösteren model verilmiştir (Inmon, 2016, s. 27-28). Ayrıca verinin bütün bileşenlerinin bulunması verinin farklı kullanıcı toplulukları tarafından da kullanılabilmesine olanak sağlar (Inmon, 2016, s. 29). Veri ile birlikte üst işlem bilgisinin de veri gölüne dahil edilmesi, veri analizi sürecini daha verimli hale getirir ve doğru kararlar alınmasına yardımcı olur.



Şekil 18. Verinin tüm bileşenleri ile birlikte belirli bir düzen dahilinde veri gölüne aktarılmasını gösteren şema (Inmon, 2016)

4.2.2.1. Veri Gölü Odaklı Üst Veri Yönetim Sistemleri

Veri gölleri odaklı üst veri türlerini destekleyen bazı önemli sistemler şunlardır:

MEDAL (Metadata-driven Automation and Learning): MEDAL sistemi, veri göllerinde üst veri yönetimini merkezi hale getirir ve verilerin etkin bir şekilde yönetilmesine olanak tanır. Bu sistem, veri keşfi, veri ilişkileri ve analiz için uygun üst veri yapıları sunar. MEDAL, verilerin merkezi olarak izlenmesini ve ilişkilendirilmesini sağlar, böylece veri göllerinin karmaşıklığını azaltır. Bu sistem aynı zamanda, makine öğrenimi ve otomasyon destekli üst veri yönetimi ile verilerin keşfini kolaylaştırır (Sawadogo ve diğerleri, 2019).

GOODS: Google tarafından geliştirilen GOODS sistemi, veri gölleri içinde büyük veri kümelerinin yönetimi ve keşfi için kullanılan bir üst veri yönetim sistemidir. GOODS, kullanıcıların büyük veri kümelerine kolayca erişmesini sağlar ve veri keşif süreçlerini hızlandırır. Bu sistem, veri yönetimini merkezileştirerek veri

bütünlüğünü sağlar ve büyük ölçekli veri kümelerinin etkin şekilde kullanılmasına yardımcı olur (Halevy ve diğerleri, 2016).

CoreKG (Core Knowledge Graph): CoreKG, veri gölleri içinde veri ilişkilerini yönetmek için kullanılan bir sistemdir. Bu sistem, veri kümeleri arasındaki ilişkileri keşfederek veri bütünlüğünü sağlar. CoreKG, veri göllerinde kapsamlı veri keşfi için optimize edilmiştir ve veri kümeleri arasında semantik bağlantılar kurarak analiz süreçlerini kolaylaştırır (Diamantini, Potena ve Storti, 2024).

Atlas: Apache Atlas, Apache Hadoop ekosistemine tümleşik olan bir üst veri yönetim sistemi olup, veri keşfi ve sınıflandırma süreçlerini yönetmek için kullanılır. Atlas, veri yönetişimi ve veri yaşam döngüsü yönetimini destekler. Bu sistem, veri göllerinde veri keşfini kolaylaştırmak için veri varlıklarının merkezi bir şekilde yönetilmesine olanak tanır. Ayrıca veri yönetişimi ve güvenlik politikalarını uygulamak için güçlü araçlar sunar (Stein ve Morrison, 2014).

Amundsen: Amundsen, Lyft tarafından geliştirilen bir üst veri yönetim platformudur. Bu sistem, veri keşfini ve analiz süreçlerini kolaylaştırmak amacıyla tasarlanmıştır. Amundsen, veri bilimi ekiplerinin veri gölleri içinde kolayca veri keşfi yapabilmesi ve verilerin etkili bir şekilde yönetilebilmesi için güçlü bir altyapı sağlar. Bu sistem, veri göllerinde farklı veri varlıklarını kategorize ederek kullanıcıların aradıkları verilere hızlıca ulaşmasını sağlar (“Amundsen-io”, 2024).

DataHub: LinkedIn tarafından geliştirilen DataHub, büyük veri ve veri gölleri için açık kaynaklı bir üst veri yönetim platformudur. DataHub, veri kümeleri arasında semantik bağlantılar kurarak, veri keşfi ve veri yönetişimini kolaylaştırır. Bu sistem, veri göllerinde karmaşık veri ilişkilerini anlamlandırmak ve veri bilimcilerin aradıkları verilere kolayca ulaşmasını sağlamak için güçlü bir araç sunar (“DataHub”, 2024).

Informatica: Informatica, veri entegrasyonu ve üst veri yönetimi için kullanılan popüler bir platformdur. Büyük veri projelerinde, verilerin yönetimini ve iş süreçlerine entegrasyonunu sağlayan kapsamlı bir sistemdir (“Informatica”, t.y.).

Collibra: İşletmeler için veri yönetişimi ve üst veri yönetimi sağlayan Collibra, veri göllerinde verilerin doğru bir şekilde sınıflandırılması ve izlenebilirliğini sağlamak için kullanılır. Verilerin uyumluluk süreçlerine uygun hale getirilmesi ve veri keşfi için önemli bir araçtır (“Collibra”, t.y.).

Üst veri yönetim sistemleri farklı veri gölü senaryolarına ve kullanım durumlarına göre değerlendirilmelidir.

4.3. VERİ GÖLLERİNİN UYGULAMA ALANLARI

Veri gölleri, büyük veri analitiği, yapay zekâ, makine öğrenimi ve gerçek zamanlı veri işleme gibi alanlarda esneklik ve ölçeklenebilirlik sağlaması nedeniyle birçok alanda yaygın olarak kullanılmaya başlanmıştır. Veri gölleri farklı kaynaklardan gelen verileri tek bir merkezde depolayarak veri entegrasyonu ve analizi için uygun bir altyapı sunar (Inmon, 2016).

Finans sektöründe veri gölleri, büyük hacimli verilerin yönetimi ve analizinde önemli bir rol oynar. Pisoni, Molnár ve Tarcsi (2021), veri göllerinin finansal işletmelerde bilgi yönetimini ve veri odaklı karar alma süreçlerini desteklediğini belirtmektedir. Benzer şekilde, Kamalakkannan, Yasmin, Arunkumar ve Kavitha (2023), Azure veri gölü altyapısını borsa analizinde kullanılarak yatırım tahminlerine katkı sağladığını vurgulamaktadır. Sağlık alanında, elektronik sağlık kayıtları, genetik veriler ve görüntüleme verileri gibi yapılandırılmamış verilerin analizini, veri gölleri ile daha verimli hale getirerek, kişiselleştirilmiş tedavi planları tasarlamak ve sağlık hizmetlerini iyileştirmek mümkündür (Fang, 2015). Perakende ve e-ticaret sektörü, müşteri davranışlarını analiz ederek alışveriş deneyimini kişiselleştirme ve tedarik zinciri süreçlerini optimize etme amacıyla veri göllerini kullanmaktadır. Müşteri verileri, web sitesi etkileşimleri ve sosyal medya verileri gibi büyük ve çeşitli veri kümeleri veri göllerinde saklanarak analiz edilir (Inmon, 2016).

Kamu kurumlarında, şehir planlaması, altyapı yönetimi ve vatandaş hizmetlerinde veri göllerinden yararlanarak büyük veri analizleri yapılmaktadır. IoT verileri ve coğrafi veriler analiz edilerek daha akıllı şehir çözümleri geliştirilebilir (Miloslavskaya ve Tolstoy, 2016). Eğitim ve araştırma alanında, veri gölleri akademik çalışmalar ve öğrenci performans analitiği için giderek daha fazla kullanılmaktadır. Eğitim kurumları, büyük veri kümelerini analiz ederek öğrenci performansını değerlendirebilir ve araştırma verilerini daha etkin bir şekilde yönetebilirler (El Assouri, 2024). Veri gölleri, Sanayi 4.0 uygulamalarında üretim süreçlerini optimize etmek için kritik bir rol oynamaktadır. Örneğin, sensör verileri kullanılarak ekipman arızaları önceden tahmin edilebilir, önleyici bakım planlaması yapılabilir ve enerji tüketimi optimize edilebilir. Bu teknoloji, üretim hatlarından toplanan büyük miktardaki veriyi depolayarak ve analiz ederek, verimliliği artırmakta ve maliyetleri düşürmektedir (Qi ve Tao, 2018).

Gerçek zamanlı veri akışıyla birleştirilmiş veri gölleri de kurumların hızlı karar alma süreçlerine önemli katkılar sunmaktadır. Örneğin, Coca-Cola anlık envanter takibi ve dağıtım planlamasını yönetmek için gerçek zamanlı veri gölü teknolojilerinden yararlanarak tedarik zinciri süreçlerini optimize etmeyi amaçlamaktadır. AstraZeneca ise klinik araştırma ve ilaç geliştirme süreçlerinde büyük hacimli verileri sürekli işleyerek araştırma sonuçlarına hızla tepki verebilmekte ve böylece ürün geliştirme döngüsünü kısaltmaktadır (Estuary, n.d.). Bu sayede kurumlar, veri göllerinin esnek yapısını sürekli veri akışıyla birleştirerek hem iş süreçlerinde verimliliği artırmakta hem de yeni ürün ve hizmetler oluşturma noktasında rekabet avantajı kazanmaktadır.

Bilimsel araştırmalarda ise veri gölleri, yalnızca büyük veri analitiğini kolaylaştırmakla kalmayıp aynı zamanda akademik üretkenliği artıran ve iş birliğini güçlendiren altyapılar sunmaktadır. Bu bağlamda Lin, Yin, Liu ve Wang (2023), NSF ve AFOSR gibi kamu programlarının desteğiyle geliştirilen SciSciNet adlı platformun, bilim camiasının kullanımına sunulduğunu ifade etmektedir. SciSciNet, 134 milyonun üzerinde bilimsel yayını içermekte ve bu yayınları NIH ve NSF fonları, patentler, klinik çalışmalar ve sosyal medya etkileşimleri gibi dış

bağlantılarla ilişkilendirmektedir. Bu platform, bilimsel bulguların yeniden üretilebilirliğini artırmayı ve “bilim üzerine bilim” araştırmalarında yenilikçi fikirlerin temsilini desteklemeyi hedeflemektedir (Lin ve diğerleri, 2023).

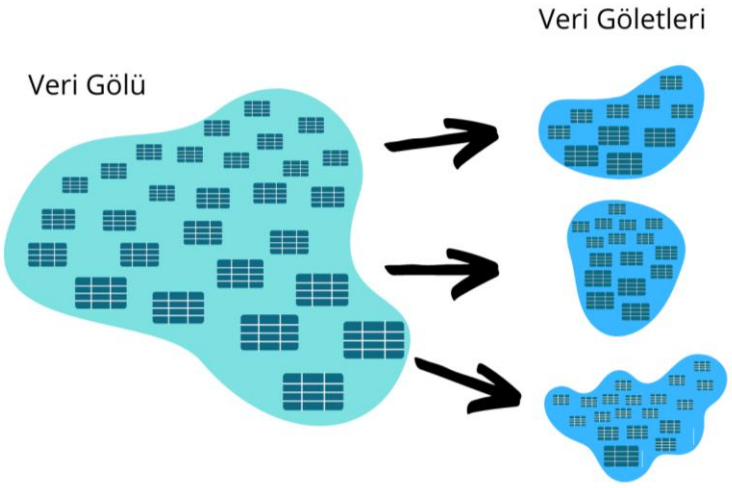
4.4. VERİ GÖLÜ MİMARİLERİ

Veri gölü mimarileri, veri göllerinin etkin ve verimli bir şekilde yönetilmesi için kritik bir rol oynar. Bu mimariler, verilerin nasıl depolanacağı, erişileceği ve işleneceği konusunda farklı yaklaşımlar sunarak, veri göllerinin esnekliğini ve ölçeklenebilirliğini artırır. Bu mimariler, verilerin ham halde saklanmasını, işlenmesini ve analiz edilmesini düzenler. Verilerin türüne, kullanıcının ihtiyaçlarına ve veri işleme gereksinimlerine göre farklı veri gölü mimarileri ortaya çıkmıştır. En yaygın kullanılan veri gölü mimarileri, gölet mimarisi (pond architecture) ve bölge mimarisi (zone architecture) olarak bilinir. Bunun yanı sıra, karma mimariler ile fonksiyonel ve olgunluk tabanlı mimariler de belirli kullanım senaryolarında etkili ve verimli çözümler sunabilir (Sawadogo ve Darmont, 2021).

4.4.1. Gölet Mimarisi

Veri göleti, Inmon'un (2016) veri gölü mimarisi çerçevesinde tanımladığı bir önemli bileşendir. Veri göletleri, veri gölünün daha organize ve yönetilebilir alt bölümlerini oluşturur. Inmon, veri gölünü verilerin türüne ve kaynağına göre çeşitli veri göletlerine ayırarak, verilerin daha etkili bir şekilde yönetilmesini ve analiz edilmesini amaçlamıştır. Gölet tabanlı veri gölü mimarisi, büyük ve çeşitli veri kümeleriyle çalışan organizasyonlar için esnek bir çözüm sunar, ancak verilerin tutarlı ve iyi yönetilmiş bir yapıda tutulması da kritik bir öneme sahiptir. Sawadogo ve Darmont (2021), veri göleti mimarisinin esnekliğini vurgulamakla birlikte, potansiyel veri yönetimi zorluklarına da dikkat çekmişlerdir.

Veri göllerindeki farklı veri türlerinin analiz edilebilmesi için veri gölleri içinde üst düzey veri yapısı oluşturmak gerekir. Ham veri gölleri veri tutma yeri olarak hizmet eder. Ham veri göllerindeki verinin analizi çok az yapılmaktadır. Veri analiz edileceği zaman ham veri gölündeki veri, verinin türüne (örneğin analog, uygulama veya metin verileri) göre bir veri göletine aktarılır. Şekil 19'da bu durumu gösteren model verilmiştir.



Şekil 19. Ham veri gölündeki verilerin türlerine göre farklı veri göletlerine aktarılması (Inmon, 2016)

Farklı türdeki veriler için kullanılan veri göletlerinin gereksinimleri farklıdır. Veri göletindeki verinin işlenmesi için önce koşullandırılarak (data conditioning) analize uygun hale getirilmesi gerekir. Veri, veri göletine girdiğinde analitik işleme sürecine hazırlanması gerekir. Çünkü veri göletindeki veri çoğu zaman analize uygun formatta değildir. Bu nedenle ham verinin işlenmesi için öncelikle koşullandırılması zorunludur. Farklı veri türleri için koşullandırma süreçleri büyük farklılık göstermektedir (Inmon, 2016, s. 48).

Veri göletlerindeki verinin kullanım ömrü sona erdikten sonra veri, bir arşivsel veri göletine taşınır (Inmon, 2016, s. 46).

4.4.1.1. Veri Göletlerinin Bileşenleri

Veri göletlerinin bazı ortak bileşenleri bulunmaktadır. Bunlar; gölet tanımlayıcı, gölet hedefi, gölet verisi, gölet üst verisi, gölet üst işlem ve gölet dönüşüm kriterleri başlıkları altında açıklanacaktır.

Gölet Tanımlayıcı (Pond Descriptor): Gölet tanımlayıcısı veri güncelleme veya yenileme sıklığı, kaynak açıklaması, veri hacmi, veri seçim kriterleri, veri özetleme kriterleri, veri organizasyon kriterleri ve veri ilişkileri gibi bilgilere sahiptir. Aşağıda bu bilgilerle ilgili kısa açıklamalar verilmiştir.

Güncelleme/yenileme sıklığı, verilerin veri göletine gönderildiği veya veri göletine iletilmeden önceki sıklığı veya döngüsünü ifade eder. Bu önceden planlanmış düzenli bir veri hareketi olabileceği gibi gerek duyulduğu takdirde yapılan bir güncelleme veya yenileme de olabilir (Inmon, 2016, s. 54).

Kaynak açıklaması, verilerin kökenini açıklar. Veriler genellikle birden fazla kaynaktan geçerek gölete ulaşacaktır. Veri analisti bu bilgiyi verilerin analiz için uygunluğunu belirlerken kullanır.

Veri hacmi, veri göletindeki verilerin hacmini ifade eder. Veri üzerinde yapılabilecek analizin türü ve derinliği üzerinde büyük etkiye sahiptir. Veri göletindeki kayıtlı veri sayısı ve byte olarak verinin kapladığı hacmi açıklar.

Seçim kriterleri, veri göletinde hangi verilerin bulunduğunu ve bu verilerin veri göletine hangi amaçla aktarıldığını açıklar.

Özetleme kriterleri, Veri göletine iletilmeden önce veriler genellikle özetlenir veya farklı bir modele dönüştürülerek işlenir. Özetleme kriterleri bu işlemler sırasında kullanılan algoritmaları açıklar. Veri analistinin yapacağı analizleri belirlemesinde fayda sağlar.

Veri ilişkileri, Farklı veri kümeleri arasındaki bağlantıları ifade eder. Veri ilişkilerinin yönetimi hem veri saklama hem de veri sorgulama süreçlerinde önemlidir (Inmon, 2016, s. 55).

Gölet Hedefi: Veri göletindeki veriyi şekillendirmek için kullanılan modeldir. Veri göleti hedefi veriyi kullanacak kuruluşun hedefleriyle ilgilidir. Satış kayıtları, hasta kayıtları, uçuş programı, ders programı, gönderi bilgileri, müşteri bilgileri, tıklanma akışı, yolcu kayıtları gibi öğeleri içerebilir. Analistin veri analizini nasıl yapacağını belirlemesi için çok önemlidir (Inmon, 2016, s. 56).

Gölet Verisi: Veri göletindeki verinin kendisiyle ilgili fiziksel görünümüdür. Gölet verisi, depolandığı altyapıya ve veri formatına bağlı olarak farklı şekillerde düzenlenebilir. Bu sayede veriler, sorgulama veya analiz aşamalarında ihtiyaçlara göre esnek biçimde işlenebilir.

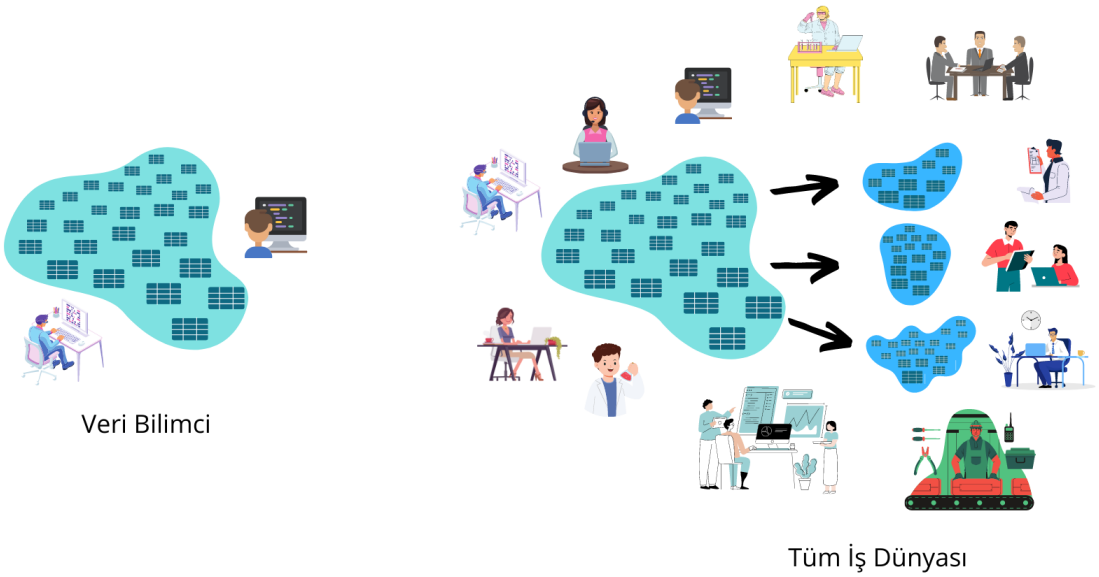
Gölet Üst Verisi: Veri göletindeki verinin fiziksel özelliklerini açıklayan üst veri, göletler için önemli bir bileşendir. Üst veri gölet dışındaki veri ve göletin kendi fiziksel organizasyonu ile ilgilidir. Farklı kaynak ve formatlardan gelen verilerin keşfi, izlenebilirliği ve kullanışlılığı için gölet üst veri yönetiminin doğru bir şekilde yürütülmesi şarttır (Inmon ve Linstedt, 2015). Bu sayede göletteki veriler, içinde bulunduğu veri gölü mimarisinin diğer bileşenleriyle entegre biçimde çalışır ve analiz süreçlerinde veriye hızlı, güvenilir erişim sağlanır.

Gölet Üst İşlem: Veri gölete ham veri olarak aktarıldıktan sonra koşullandırılarak kullanılabilir ve analiz edilebilir duruma getirilir. Bu sürece ait açıklama gölette bulunur. Farklı veri göletlerinin gerektirdiği koşullandırma işlemleri ve süreçleri farklıdır. Üst işlem bilgileri veri gölete ulaşmadan önce veri üzerinde yapılan işlemleri de içerebilir (Inmon, 2016, s. 58, 59).

Gölet Dönüşüm Kriterleri: Gölete uygun olarak göletteki verinin koşullandırılması sürecinde kullanılan kriterleri içerir. Her veri göletinin kendine ait özgün dönüşüm kriterleri bulunur. Gölet dönüşüm kriterleri analistin veride yapılan dönüşüm işlemlerinin nasıl yapıldığını anlamasını sağlar. Örneğin analog veri göletinin ölçümlerle ilgili belirli bir ölçümün belirli bir sayıdan büyük olması durumunda rekor olarak tutulması veya belirli bir makineden belirli bir dönemde gelen bütün ölçümlerin yakalanması gibi kriterleri olabilir. Uygulama veri göletinde ise örneğin “cinsiyet=1” ise cinsiyeti kadına, “cinsiyet=0” ise cinsiyeti erkeğe dönüştür veya

ölçüm inç cinsinden yapılıyorsa, santimetreye dönüştür biçiminde bir kriter bulunabilir. Metinsel veri göletinde ise “kelime=Ford” ise sınıflandırmaya araba eklenmesi veya “kelime=selvi” ise “ağaç” yazılması biçiminde kriterler bulunabilir (Inmon, 2016, s. 61).

Veri göletleri yukarıda açıklanan bileşenler doğrultusunda oluşturulduğunda sadece veri bilimciler tarafından değil, kuruluş yöneticileri, paydaşları, çalışanları vb. veriye ihtiyaç duyan çok daha geniş bir kitle tarafından kullanılabilirler. Ham veri göllerini ancak bu konuda uzman veri bilimciler kullanabilirler ve bu süreç hem çok maliyetli olmaktadır hem de çok fazla zaman gerektirmektedir. Şekil 20‘de bu durumu gösteren şema verilmiştir.



Şekil 20. Ham veri gölü ile veri göleti mimarisi ile oluşturulmuş veri gölünün kullanıcı karşılaştırılmasını açıklayan görsel (Inmon, 2016)

4.4.1.2. Veri Türüne Göre Oluşturulan Veri Göletleri

4.4.1.2.1. Ham Veri Göleti

Veri gölü, farklı kaynaklardan gelen ham veriyi orijinal formatında depolayan, esnek ve ölçeklenebilir bir veri saklama ve yönetim sistemidir (Hai ve diğerleri, 2021, s. 4). Ham veri göletleri birçok kuruluşun veri gölü dediği yapılardır. Kuruluşlar verilerini veri gölünde biriktirirler ancak çoğunlukla veri üzerinde analitik işlem yapamazlar. Veri gölündeki ham verilere de analitik işlem yapılması mümkündür ancak bu süreç bu konuda uzmanlık gerektirir, yani ham verinin işlenmesi için veri bilimcisine ihtiyaç duyulur. Ancak veri koşullandırıldıktan sonra analiz edilmesi kolaylaşır. Veri ham veri göletinden veri göletine aktarıldıktan sonra ham veri göletindeki verinin kaldırılması gerekir. Veri ham veri göletinden analog veri, uygulama verisi veya metin verisi göletine aktarıldıktan sonra veri amacına hizmet etmiş olacaktır. Ham veri göletinde verinin işlenmesi çok nadir yapılmaktadır, bu nedenle veri aktarıldıktan sonra veriyi ham veri göletinde de tutmaya devam etmek veri karmaşası yaratır ve ham veri gölü sadece verinin tutulduğu bir ortam haline gelir (Inmon, 2016, s. 48). Ham veri göletleri için kalite ölçüsü, göletten verinin ne kadar hızlı geçtiği yani verinin ne kadar hızlı veri göletine aktarıldığı ve ham veri göletinin küçüklüğüdür. Bu nedenle ham veri göletindeki veri hızlıca veri göletlerine iletilmelidir (Inmon, 2016, s. 49).

Verinin veri göletine her zaman ham veri göletinden aktarılması gerekmez ancak uygulamada çoğunlukla önce ham veri göletine oradan veri göletine aktarılır. Eğer geliştirici yeterince bilgiliyse verinin direk olarak analog, uygulama veya metin veri göletine gönderilmesi mümkündür. Verinin yaşam döngüsünün sonunda veri analog, uygulama veya metin veri göletinden arşivsel veri göletine aktarılır (Inmon, 2016, s. 51).

4.4.1.2.2. Analog Veri Göleti

Analog veri göletlerinde mekanik olarak oluşturulmuş ölçümlerden oluşan veri bulunur. Üretim kontrol makineleri, periyodik ölçümler gibi çeşitli analog veri kaynakları bulunmaktadır. Analog veri genellikle belirli aralıklarla ölçüm yapılarak elde edilir. Bu sıklık uzunluk veya mesafe cinsinden belirlenmektedir. Örneğin ürünler dizilir ve belirli bir mesafe aralıklarla anlık görüntü alınır veya her milisaniyede bir ürün ölçülerek veri elde edilir (Inmon, 2016, s. 63).

Genel olarak analog veri göletleriyle ilgili iki temel problem bulunmaktadır. Bunlardan ilki, veri hacminin çok büyük olmasıdır. Makineler sürekli olarak veri üretirler ve bu verilerin büyük çoğunluğu tekrarlardan oluşmaktadır. Örneğin her milisaniyede görüntü alan bir makinenin kaydettiği görüntülerin çok küçük bir kısmı iş değeri taşımaktadır. Veri hacminin çok büyük olması nedeniyle farklı olan ve değer taşıyan verilerin diğer veriler arasından bulunamaması da önemli bir problemdir. Analog veri göletleriyle ilgili ikinci problem ise verinin üretilmesi ile ilgili verinin kaybolmasıdır. Analog veri analistleri genellikle yalnızca verilerle ilgilenmektedirler ve çoğunlukla verilerle ilişkili veri tanımlayıcıları göz ardı etmektedirler (Inmon, 2016, s. 63). Ancak veri tanımlayıcılar en az verinin kendisi kadar değerlidir. Analistler için analog veriyi analiz edilebilir duruma getirmek önemli bir zorluktur. Verilerin düzenlenerek ana hatlarının belirlenmesi ve dönüştürme/koşullandırma işlemlerinin uygulanarak analize hazır hale getirilmesi gerekir (Inmon, 2016, s. 64).

Analog veri göletlerindeki verilerin seçim kriterleri, kaynağı, hacmi, veri göletine taşınma sıklığı, tarihi ve saati gibi açıklayıcı bilgiler analistin verileri anlayabilmesi için kritik öneme sahiptir. Bu bilgiler genel olarak Veri Açıklayıcı (Data Descriptor) olarak adlandırılır (Inmon, 2016, s. 64).

Analog veri göletlerine veri taşınırken öncelikle verinin yakalanması ve gölete taşınması, sonrasında ise verinin son kullanıcı tarafından kolaylıkla analiz edilebilecek yapı ve forma dönüştürülmesi gerekir (Inmon, 2016, s. 66).

Ham veri göletinden analog veri göletine aktarılan analog veri için koşullandırma işlemi büyük ölçüde veri azaltma işleminden oluşur. Analog veri göletindeki veri hacminin yönetilebilir ve anlamlı bir veri hacmine indirgenecek şekilde koşullandırılması gerekir (Inmon, 2016, s. 50). Veriyi dönüştürme yani koşullandırma süreci tamamen veri göletinin içinde gerçekleşir (Inmon, 2016, s. 66).

4.4.1.2.2.1. Analog Veri İş Değeri

Analog veri kullanılarak elde edilebilecek kazanca aşağıdaki örnekler verilebilir:

Bir şirket tarafından üretilen bir bisküvide bakteri tespit edildiğini ve bu bakterinin ciddi hastalıklara sebep olabileceğini varsayalım. Bakteri bulaşan bisküviyi üreten şirket ile iletişime geçilerek bisküvinin nerede ve hangi tesiste üretildiği bilgisine ulaşılır. Daha sonra şirketin analog verileri aracılığıyla bu bisküvi ile aynı dönemde, aynı tesiste üretilen bütün bisküviler belirlenerek bisküvilerin toplatılması sağlanabilir ve böylece potansiyel hastalıklar önlenmiş olur.

Başka bir örnek olarak telefon görüşmelerinden bir suç teşkil eden olayların tespit edilmesi verilebilir. Telefon konuşmalarının doğrudan insanlar tarafından tek tek dinlenerek şüpheli durumların tespit edilmesi mümkün değildir. Ancak bu işlem gelişmiş veri analizi teknikleri kullanılarak makinelerce yürütülebilir. Böylece toplum için potansiyel tehlike oluşturabilecek olayların önceden tespit edilerek, önlem alınması ve engellenmesi mümkün olabilir.

4.4.1.2.3. Uygulama Veri Göleti

Uygulama veri göletine uygulama veya uygulamaların yürütülmesi sonucu oluşan veri aktarılır, bu verilerin çoğunluğu işlemler sonucu oluşan verilerdir. Bir uygulamada bir işlem gerçekleşir ve bununla ilgili elektronik kayıt oluşturulur. Bu kayıt kurumsal sistemlerde saklanır ve iş süreçlerini yürütmede kullanılır. Bu

verinin aktif kullanımı sona erdikten sonra veri uygulama veri göletine aktarılır. Ürün katalogları, telefon görüşme kayıtları, sevkiyat çizelgeleri vb. veriler de operasyonel uygulama verileri olarak uygulama veri göletlerine gönderilebilir.

Uygulama verilerini şekillendiren önemli unsurlardan bir tanesi operasyonel sistemin altyapısıdır. Uygulama veri bankasına gönderilen veri için verinin operasyonel uygulamaya yerleştiği andaki kaydedilme şekli derin bir etkiye sahiptir. Verinin bu ilk kayıt şekli ve özellikleri bir DNA gibi verinin tüm yaşamını etkiler. Veri organizasyonu, verilerin şekillenmesi, saklama biçimi, içeriği vb. bütün süreçler uygulama altyapısından ve buradaki ilk kayıttan etkilenir. Yani bu ilk kayıt verinin tanımlayıcı özelliklerinden biri olmaktadır. Uygulama verilerinin kaynağı, yaklaşık hacmi, toplanma sıklığı vb. özellikleri uygulama verilerinin tanımlayıcıları olarak adlandırılır. Bu bilgiler uygulama veri göletine aktarılan verilerin nasıl işleneceğinin belirlenmesi konusunda analiste yol gösterir.

Bir uygulama tarafından oluşturulan veri tek tiptedir bu nedenle diğer veri türlerine göre daha temizdir. Ancak uygulama veri göletleri genellikle birçok farklı uygulamadan veri içerir bu nedenle bu tür veri çoğunlukla entegre değildir. Büyük kuruluşlar genellikle birden fazla uygulama kullanırlar ve farklı uygulamalara ait veri tek bir veri göletine gönderilir. Bu nedenle bu göletteki verinin analiz edilmesi zordur. Bir uygulama veri göletine tek bir uygulamadan veri gönderiliyorsa veri entegre olabilir ancak bu yaygın olarak görülen bir durum değildir (Inmon, 2016, s. 85-87).

4.4.1.2.3.1. Uygulama Verisi İş Değeri

Örneğin bir yemek sipariş uygulamasının verileri kullanılarak hangi bölgelerde, hangi saatlerde siparişlerin daha yoğun olduğu belirlenebilir ve böylece bu bilgi kullanılarak sunulan hizmet geliştirilebilir.

Yine kişilerin kredi kartı harcamaları ve ödemeleri takip edilerek, bankalar tarafından kişiye özel kredi kartı limitinin artırılması önerileri, uygulama verilerden değer elde edilmesine örnek gösterilebilir.

Başka bir örnek olarak, çevrimiçi alışveriş uygulamaları üzerinden verilen siparişler dikkate alınarak, şirketler tarafından kişiye özel ürün önerileri yapılması, verilebilir.

Uygulama veri göletlerinin kısaca özetlemek gerekirse, bu göletlere bir uygulamanın oluşturduğu veriler aktarılır. Bu verilerin analiz edilebilmesi için öncelikle veri göleti modeline göre entegre edilmesi gerekir. Bu işlem analog veri göletlerindeki verilerin koşullandırılmasına benzer ancak yapılan işlemler çok farklıdır. Bir verinin uygulamada oluştuğu ilk şekli uygulama veri göletine saklandığında dahi korunur. Uygulama veri göletlerinde ilişkisel bir veri tabanı yönetim sistemi kullanılmasa dahi veriler ilişkisel kökenlerini yansıtırlar. Veri ile ilgili tanımlayıcılar analist için çok önemlidir. İşletim ortamındaki veriler ilişkisel bir biçimde saklanır. İlişkisel format kayıtları, kayıtların niteliklerini, anahtar, dizin vb. bilgileri içerir (Inmon, 2016, s. 102).

4.4.1.2.4. Metinsel Veri Göleti

Metinsel verilerin gönderildiği veri göletleridir. Metin verileri yapılandırılmamıştır, bu nedenle verilerin yüzeysel analizi yapılabilir ancak derinlemesine analiz edilmesi zordur. Kurumsal dünyada çok fazla metinsel veri bulunmakla birlikte bu veriler yeterince analiz edilememekte ve karar aşamalarında kullanılmamaktadır. Ancak metinsel verinin de koşullandırılarak analiz süreçlerinde kullanılması ve sahip olduğu büyük potansiyelin açığa çıkarılması mümkündür (Inmon, 2016, s. 103).

Bilgisayarlar tekdüze verileri daha kolay ve hızlı şekilde okuyup işleyebilirler, çünkü sistem aynı işlemin tekrarı üzerine çalışır. Ancak bilgisayarlar için metinsel verinin her kelimesi ayrı bir evrendir ve bilgisayarın yapacağı işlem farklı olacaktır.

Bu nedenle metin verileri çok yüzeysel olarak analiz edilmektedir (Inmon, 2016, s. 103).

Metin verilerinin yönetsel kararlarda kullanılabilecek birçok alanı bulunmaktadır. Örneğin kurumsal sözleşmeler, çağrı merkezi konuşmaları, müşteri geri bildirimleri, sağlık kayıtları, sigorta kayıtları, insan kaynakları kayıtları vb. gibi analiz edilerek kurumsal fayda sağlanabilecek potansiyele sahip çok farklı veriler bulunmaktadır. Ancak birçok kuruluş metinsel verileri toplar ve dosyalayarak kaldırır; analiz etmek için çaba harcamaz (Inmon, 2016, s. 104).

Derinlemesine analiz yapılabilmesi için öncelikle metin verisinin belirsizliğinin giderilmesi gerekir. Belirsizlik giderme ile metin tek tipte yeniden yapılandırılmış, bağlamı tanımlanmış ve metne eklenmiş olur (Inmon, 2016, s. 50).

4.4.1.2.4.1. Metinsel Veri İş Değeri

Metinsel veri kullanımına Culotta (2010) tarafından gerçekleştirilen, Twitter verilerinin kullanıldığı, çalışma örnek olarak gösterilebilir. Bu çalışmada influenza ile ilgili tweetler regresyon yöntemi ile CDC (Hastalık Kontrol ve Önleme Merkezleri) istatistikleri ile ilişkilendirilmiş ve influenza salgınını tahmin eden bir yöntem geliştirilmiştir. Bir başka çalışmada ise yine Twitter mesajları üzerinden ilaç deneylerinde gönüllü olan bireylerin yaşadıkları yan etkiler tespit edilmeye çalışılmıştır (Jiang ve Zheng, 2013). Yine kuruluşlar e-posta verilerini kurumsal sorunların belirlenmesi, iş süreçleriyle ilgili zorlukların belirlenmesi, müşterilerin yaşadığı sıkıntıların belirlenmesi vb. amaçlarla analiz edebilirler.

Kısaca özetlemek gerekirse metinsel veri göleti, metin verilerin bulunduğu yerdir ve buradaki verinin etkin olarak kullanılabilmesi için dönüştürme ve koşullandırma işlemlerinin uygulanması gerekir. Bu işlemler sonucunda metin ve bağlamı belirlenerek standart bir veri tabanı biçiminde saklanır. Genel olarak bu işlemler metinsel belirsizliğin giderilmesi olarak adlandırılır. Metinsel verilerin

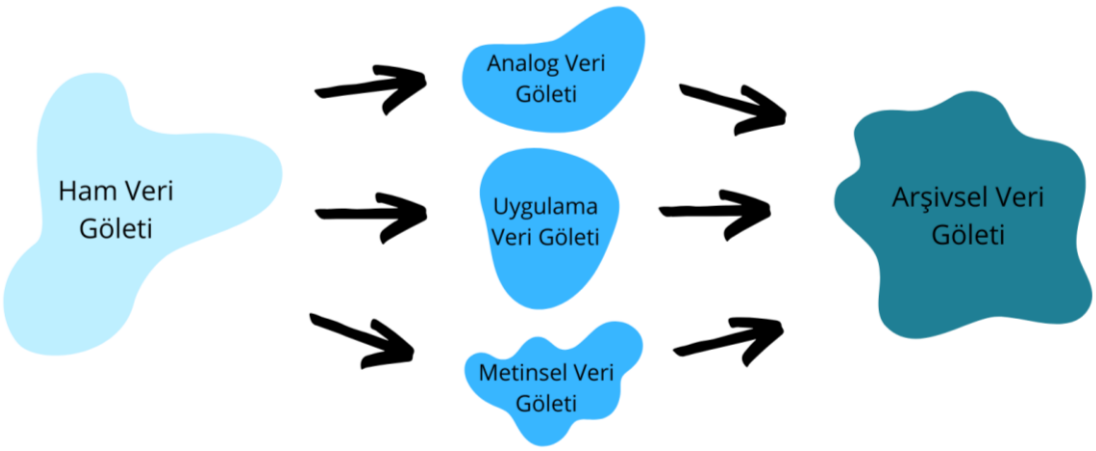
koşullandırılması uygulama verilerinin koşullandırılmasından oldukça farklıdır (Inmon, 2016, s. 123).

4.4.1.2.5. Arşivsel Veri Göleti

Veri göllerinin önemli bir bileşeni arşivsel veli göletleridir. Arşivsel veri göletleri aktif olarak analiz edilmesine ihtiyaç duyulmayan ancak gelecekte analiz edilebilecek verinin saklandığı göletlerdir. Analog veri göleti, uygulama veri göleti ve metinsel veri göletindeki kullanım ihtimali azalmış veriler arşivsel veri göletine gönderilir. Arşivsel veri göletinin amacı gelecekte kullanılma ihtimali olan bu verileri tutmak, böylece veri göletlerindeki gereksiz verilerin çıkarılmasına olanak sağlamaktır. Kullanım ihtimali azalan verilerin veri göletlerinden çıkarılmasıyla veri göletlerindeki verinin verimli bir şekilde analizine devam edilmesi, veri göletlerinin daha etkin ve verimli şekilde kullanılması hedeflenir (Inmon, 2016, s. 123, 204).

Veri göletlerindeki veriler arşivsel veri göletine gönderilirken verinin yapısı değiştirilir ve veriye ait üst veri ve üst işlem verisi fiziksel olarak veriye bitişik olarak kaydedilir. Böylece bu göletteki veriler ileride analiz edilmek istendiğinde bu bilgilerin korunması sağlanmış olur. Yine arşivsel veri göletlerindeki verilere ihtiyaç duyulduğunda verimli bir şekilde ulaşılabilmesi için buradaki verilerin bağımsız olarak indekslenmesi de faydalıdır (Inmon, 2016, s. 204, 206).

Kısaca özetlemek gerekirse, analiz edilecek veri öncelikle ham veri göletine girer. Buradan verinin türüne göre analog, uygulama veya metinsel veri göletine iletilir. Burada koşullandırılarak analiz edilir ve kullanım ömrü sona erdiğinde gelecekte ihtiyaç duyulabilecek analizlerde kullanılmak üzere arşivsel veri göletine gönderilir (Inmon, 2016, s. 53). Şekil 21’de Inmon’un (2016) önerdiği veri gölü mimarisini açıklayan görsel sunulmuştur.



Şekil 21. Inmon (2016)'un veri göleti tabanlı veri gölü mimarisi

Veri göletleri ilk bakışta benzer görünmekle birlikte bazı önemli yapısal farklılıklara sahiptir.

Veri göletlerinin birçok ortak yapısal özelliği bulunmaktadır:

- Bütün veri göletlerine çoğunlukla ham veri aktarılır.
- Ham veri analiz için uygun forma dönüştürülür veya koşullandırılır.
- Gölet içinde tek tip, entegre edilmiş ve analize uygun bir veri yapısı oluşturulur.
- Gölet verisi analiz edilerek sonuç iş süreçlerinin geliştirilmesi için kullanılır.
- Aktif kullanımı sona eren veriler arşivsel veri göletine gönderilir.
- Analitik işlemlere uygun veri barındırır.
- Analiste yardımcı olacak destekleyici altyapıya sahiptir (Inmon, 2016, s. 125).

Farklı türdeki veri göletlerinin yapısal farklılıkları bulunmaktadır:

- Farklı göletlere giren ham veri gölet türüne göre büyük farklılıklar içermektedir.

- Farklı göletlerdeki veriler için dönüşüm ve koşullandırma işlemleri çok farklıdır.
- Farklı veri göletlerindeki veriye uygulanan analizlerin türleri çok farklıdır (Inmon, 2016, s. 126).
- Her veri göletindeki verilerin ilişkisel formatta olması gerekmez.
- Veri göletlerinde aynı teknolojinin kullanılması şart değildir, ancak kurumlar birden fazla platformu kullanabilmek için aynı teknoloji olmasını tercih edebilirler.
- Veri göletlerinin toplam hacmi farklıdır. Gölet hacmi iş hedefleri ve göletlerindeki verinin doğasına bağlıdır. Bir endüstrinin daha fazla veri türü ve daha az verisi olabilir. Bir mühendislik firması daha fazla analog veriye, bir telefon şirketi daha fazla uygulama verisine, bir pazar araştırma firması ise daha fazla metinsel veriye sahip olabilir (Inmon, 2016, s. 126, 127).

4.4.1.3. Göletler Arası Veri Transferi ve Birden Fazla Göletten Analiz Yapılması

Veri göletleri arası veri transferi yapılması mümkündür ancak veri bulunduğu göletin üst veri, üst işlem ve tanımlayıcı gibi altyapı bileşenleriyle birlikte değerlidir ve veri taşınırken bu altyapının da taşınması çok zordur. Bu nedenle teknik olarak veri göletleri arası transfer mümkün olsa da uygulamada bu çok nadir karşılaşılan bir durumdur (Inmon, 2016, s. 127, 128).

Birden fazla göletteki verilerin aynı analizde kullanılması da mümkündür ancak yapılan analizler daha çok veri türü ile ilgili olduğu için genellikle tek bir veri göletindeki veriler için analiz yapılmaktadır. Yine de birden fazla göletteki veri birlikte analiz edilecekse öncelikle göletlerin üst veri altyapıları kullanılarak ilişkilendirilmesi gerekir (Inmon, 2016, s. 129, 131).

Burada önemli bir konu ise veri türünün standart olarak tanımlanan analog veri, uygulama verisi veya metinsel veri türüne uygun olmadığı durumda ne yapılması gerektiğidir. Bu tür standarda uymayan veriler ait olmadıkları veri göletlerine yerleştirilmemeli, ham veri göletinde ayrı bir kısımda tutulmalıdır. Bu tür verilerin analizlerde kullanılabilmesi için dönüştürülüp entegre edilerek koşullandırılması şarttır (Inmon, 2016, s. 133, 134).

Kısaca özetlemek gerekirse farklı veri göletleri yapısal olarak benzerlikler içermekle birlikte, türe özgün farklılıklar da içermektedir. Bu nedenle her veri göletindeki veriye uygulanan analiz yöntemleri de farklı olmaktadır. Standart olarak açıklanan analog veri göleti, uygulama veri göleti ve metinsel veri göleti türlerine uymayan veriler ham veri göletinde ayrı bir kısımda tutulmalıdır. Bu verilerin de koşullandırılarak analiz edilmesi mümkündür (Inmon, 2016, s. 134, 136).

4.4.2 Bölge Mimarisi (Zone Architecture)

Veri göletleri için kullanılan diğer bir mimari model ise bölge mimarisidir (zone architechture). Bölge mimarisinde veri işlenme düzeyine göre farklı bölgelerde saklanır (Giebler ve diğerleri, 2019, s. 4). Bölge mimarisi, verilerin yaşam döngüsüne göre farklı aşamalarda düzenlenmesini sağlayan bir veri gölü yapısıdır. Bu yapı, verilerin işlenme olgunluğuna ve kullanım amacına göre farklı "bölgelerde" depolanmasını öngörür. Geleneksel üç bölgeli yapının ötesinde, modern bölge mimarileri genellikle daha fazla bölge içerir ve her bir bölge belirli bir amaca hizmet eder (Sawadogo ve Darmont, 2021):

Ham Veri Bölgesi (Raw Data Zone): Verilerin ilk kez alındığı ve ham halde saklandığı bölgedir. Herhangi bir işlem yapılmamış veya temizlenmemiş veri bu bölgede bulunur. Ham veri, kaynak sistemlerden (IoT cihazları, sosyal medya, log dosyaları vb.) doğrudan alınır ve herhangi bir şemaya tabi tutulmadan veri gölüne aktarılır.

Geçici Bölge (Transient Zone): Verilerin işlenme sürecine hazırlandığı ara aşamayı temsil eder. Veriler, ham veri bölgesinden alındıktan sonra genellikle geçici bölgeye aktarılır. Bu aşamada, veri temizleme, veri doğrulama, dönüşüm ve birleştirme gibi temel işlemler gerçekleştirilir (Giebler ve diğerleri, 2019).

Temizlenmiş Veri Bölgesi (Cleansed Data Zone): Bu aşama, verilerin işlenmeye başlandığı, temizlendiği ve bir miktar işleme tabi tutulduğu aşamadır. Ham veri bölgesindeki veriler, güvenilir hale getirilir ve hatalı veya eksik veriler düzeltilir. Temizlenmiş veri, genellikle analiz veya daha ileri işleme hazır hale getirilir (Giebler, Gröger, Hoos, Schwarz ve Mitschang, 2020).

İşlenmiş Veri Bölgesi (Processed Data Zone): Bu bölge, daha önce temizlenmiş ve temel işlemlerden geçmiş verilerin ileri düzeyde işlenip analiz için optimize edildiği yerdir. Veriler, burada yapılandırılır ve analiz süreçlerinde kullanılmak üzere son haline getirilir. Bu aşamada, veriler analitik uygulamalara uygun formatlara dönüştürülür ve yüksek performanslı sorgulama veya raporlama için hazırlanır (Miloslavskaya ve Tolstoy, 2016).

Güvenilir Veri Bölgesi (Trusted Data Zone): Bu bölge, işlenmiş, temizlenmiş ve analiz için hazır hale getirilmiş verilerin bulunduğu yerdir. Veri bilimi ekipleri ve analistler, bu bölgede bulunan veriler üzerinde çalışarak derinlemesine analizler yapar. Güvenilir veri bölgesi, aynı zamanda raporlama ve iş zekâsı araçlarına veri sağlayan bir bölge olarak da kullanılır.

Keşif Bölgesi (Discovery Zone): Veri bilimcilerin ve analistlerin deneysel çalışmalar yapabildiği, yeni modeller geliştirebileceği ve hipotezler test edebileceği alandır (Hai ve diğerleri, 2016).

Arşiv Bölgesi (Archive Zone): Uzun süredir kullanılmayan veya yasal gereklilikler nedeniyle saklanması gereken verilerin depolandığı bölgedir.

Örneğin, Zaloni'nin mimarisi geçici yükleme bölgesi, ham veri bölgesi, güvenilir veri bölgesi, keşif alanı, tüketim bölgesi ve yönetişim bölgesi olmak üzere altı

bölge içerir. Bu bölge mimarisinin modeli Şekil 22'de verilmiştir (LaPlante ve Sharma, 2016'dan aktaran Sawadogo ve Darmont, 2021).

Şekil 22. Zaloni'nin bölge mimarisi modeli

Bölge mimarisi, veri güvenliğini artırır ve verilerin yaşam döngüsü boyunca belirli bir düzen içinde işlenmesine olanak tanır. Bu yapı aynı zamanda veri kalitesinin artırılmasına, veri soyağacının takip edilmesine ve veri erişim kontrollerinin daha etkili bir şekilde uygulanmasına olanak tanır (Giebler ve diğerleri, 2021). Modern bölge mimarileri, organizasyonların ihtiyaçlarına göre uyarlanabilir ve gerektiğinde mimariye yeni bölgeler eklenebilir. Örneğin bazı organizasyonlar, gerçek zamanlı veri işleme için ayrı bir "hız bölgesi" veya makine öğrenimi modelleri için özel bir "model bölgesi" ekleyebilir (Sawadogo ve Darmont, 2021).

Bölge mimarisine başka bir örnek olarak Lambda mimarisi gösterilebilir. Lambda mimarisi, yığın işleme ve akan veri işleme bölgelerini birleştirerek, verilerin hem toplu hem de gerçek zamanlı işlenmesini sağlayan bir yapı sunmaktadır. Bu modelde ham veri bölgesi, güvenilir veri bölgesi ve tüketim bölgesi gibi bölümler, yığın ve gerçek zamanlı işleme gereksinimlerine göre organize edilmiştir. Lambda mimarisi veri işleme süreçlerini destekleyerek modern veri yönetimi ihtiyaçlarını karşılayan bir bölge mimarisi çözümüdür (John ve Misra, 2017).

Bölge mimarisinin başarılı bir şekilde uygulanması, güçlü bir üst veri yönetimi, veri kataloglama ve veri yönetişimi stratejisi gerektirir. Bu yaklaşım, veri gölünün zamanla bir veri bataklığına dönüşme riskini azaltır ve verinin değerini maksimize eder.

4.4.3. Gölet ve Bölge Mimarisi Tabanlı Melez Mimariler

Karma mimariler, bölge ve gölet mimarilerinin avantajlarını bir araya getirerek, farklı veri kullanım durumları için optimize edilmiş bir çözüm sunar. Bu mimariler, verilerin yaşam döngüsüne göre bölgelere ayrıldığı bir yapıyı, verilerin işleme ve kullanım amaçlarına göre özelleştirilmiş göletlerle birleştirir. Sawadogo ve Darmont (2021), karma mimarilerin özellikle büyük ve karmaşık veri ekosistemlerinde etkili olduğunu vurgulamıştır. Bu mimariler, veri bilimi ekiplerinin farklı gereksinimleri olduğu durumlarda veya çeşitli analitik iş yüklerinin bir arada yönetilmesi gerektiğinde ideal bir çözüm sunar (Sawadogo ve Darmont, 2021). Giebler ve diğerleri (2019), karma mimarilerin özellikle veri gölü ve veri ambarı entegrasyonunda etkili olduğunu belirtmiştir. Bu yaklaşım, geleneksel veri ambarlarının yapılandırılmış veri yönetimi avantajlarını, veri göllerinin esnekliği ile birleştirir. Şekil 23'te örnek bir karma mimari uygulaması verilmiştir.

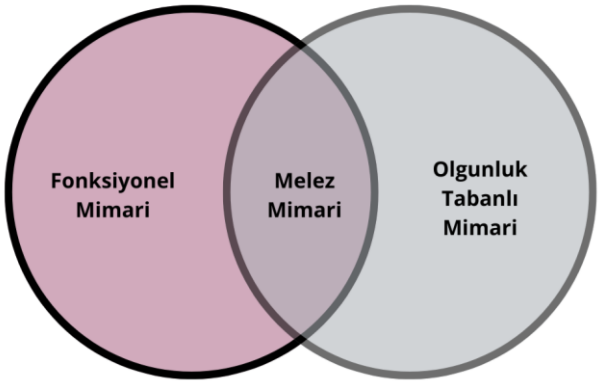
Şekil 23. Örnek bir (gölet-bölge) melez mimari uygulaması

Bu yapı hem verilerin yaşam döngüsünü yönetmeyi hem de farklı kullanım senaryolarına hizmet etmeyi mümkün kılar. Örnek bir karma mimaride, analitik göletler makine öğrenimi göleti ve iş zekâsı göleti gibi özelleştirilmiş göletler olabilir. Bu tür karma mimariler, büyük veri analitiği projelerinde ve çeşitli veri kullanım senaryolarının bir arada bulunduğu organizasyonlara önemli avantajlar sağlar.

4.4.4. Fonksiyonel ve Olgunluk Tabanlı Mimariler

Fonksiyonel ve olgunluk tabanlı mimari, geleneksel veri gölü mimarilerinin karşılaştığı bazı zorlukları aşmak ve organizasyonların değişen ihtiyaçlarına daha uygun çözümler sunmak amacıyla geliştirilmiştir. Sawadogo ve Darmont (2021),

veri gölü mimarilerini sınıflandırma tabanlı bir topoloji ile üç grupta ele almaktadır: Fonksiyonel mimariler, olgunluk tabanlı mimariler ve bu iki yapının birleşimi olan melez mimariler. Şekil 24'te bu sınıflandırmayı gösteren şema verilmiştir.



Şekil 24. Sawadogo ve Darmont (2021)'un önerdiği fonksiyonel X olgunluk tabanlı veri gölü mimarisi sınıflandırma topolojisi

Fonksiyonel mimari, veri gölünün farklı işlevlerini (veri alımı, işleme, analiz gibi) ayrı bileşenler olarak ele alarak her bir aşamanın optimize edilmesini hedeflemektedir. Bu mimaride her işlev için özelleşmiş süreçler tanımlanmakta, böylece veri işleme süreçlerinde verimlilik sağlanmaktadır. Öte yandan, olgunluk tabanlı mimari, organizasyonun veri yönetimi kapasitesinin zaman içinde gelişimini göz önünde bulundurarak veri gölünün olgunlaşma seviyelerine göre yapılandırılmasını sağlamaktadır. Bu yaklaşım, veri yönetimi becerilerinin zamanla gelişmesini destekleyen bir yapı kurmayı amaçlamaktadır.

Bu iki mimari yaklaşımın birleşimi olarak tanımlanan melez mimariler, organizasyonların hem işlevselliği optimize etmesine hem de olgunluk seviyelerine göre veri yönetim süreçlerini uyarlamalarına olanak tanımaktadır. Melez mimariler, esnek ve gelişime açık yapısıyla özellikle karmaşık veri ekosistemlerine sahip organizasyonların veri göllerini daha etkili kullanmalarına katkı sağlamaktadır (Sawadogo ve Darmont, 2021).

Bu sınıflandırmalar, veri göllerinin organizasyonun ihtiyaçlarına ve veri yönetimi olgunluk seviyesine göre daha verimli kullanılmasını sağlamaktadır.

4.4.4.1. Fonksiyonel Mimariler

Fonksiyonel mimari sınıflandırması içinde, veri gölü, genel veri gölü iş akışı boyunca veri üzerinde gerçekleştirilen işlemler açısından analiz edilir (Wieder ve Nolte, 2022). Bu yaklaşımda veri gölü, bu işlemler çerçevesinde çeşitli fonksiyonel katmanlara ayrılır. Bu mimari, her bir veri yönetim sürecini bağımsız bir modül olarak ele alarak, bu modüllerin birbirleriyle etkileşimini optimize eder. Fonksiyonel mimarilerin temel amacı, veri işleme, depolama, analiz ve erişim süreçlerini verimli ve yönetilebilir hale getirirken, bu süreçler arasındaki işlevsel ilişkiyi korumaktır (Sawadogo ve Darmont, 2021). Fonksiyonel mimari, veri gölünü oluşturan ana bileşenlerin entegrasyonu, esnekliği ve yönetimi açısından önemli avantajlar sunar. Bu bağlamda, fonksiyonel mimarinin tanımladığı katmanları incelerken, her bir bileşenin optimize edilmiş veri yönetim süreçlerine nasıl katkıda bulunduğunu değerlendirmek gerekir.

Veri Alımı: Fonksiyonel mimaride veri alımı çeşitli kaynaklardan veri almakla birlikte bu sürecin gerçek zamanlı veri işleme ve toplu işleme metodolojileri ile optimize edilmesini sağlar. Örneğin, gerçek zamanlı veri akışları, IoT cihazlarından gelen verilerin anında işlenmesine olanak tanır, böylece büyük ölçekli veri işleme süreçleri hızlanır (Terrizzano ve diğerleri, 2015).

Veri İşleme: Fonksiyonel mimaride veri işleme, veri göllerinden farklı olarak modüler bir yaklaşımla ele alınır. Her modül, spesifik bir veri işleme görevini üstlenir ve bu modüller birbirinden bağımsız olarak çalışabilir. Örneğin, veri temizleme işlemleri bir modül tarafından gerçekleştirilirken, veri analizine yönelik zenginleştirme işlemleri farklı bir modül tarafından yürütülür. Bu modüler yapı, büyük veri işleme teknolojilerinin (Apache Spark, Flink vb.) daha verimli kullanılmasına olanak tanır (Halevy ve diğerleri, 2016).

Veri Depolama ve Erişim: Fonksiyonel mimariler, veri göllerindeki depolama süreçlerini daha detaylı ve esnek bir yetkilendirme sistemiyle destekler. "İnce taneli erişim kontrolü" olarak adlandırılan bu yaklaşım, kullanıcılara veri setinin belirli bölümleri için farklı erişim seviyeleri tanımlar (Hu ve diğerleri, 2015). Bu

yöntem, veri yöneticilerine, kullanıcıların yalnızca belirli veri öğelerine, sütunlara, satırlara veya hatta hücrelere erişimini kontrol etme imkânı verir, böylece veri güvenliği ve gizliliği daha hassas bir şekilde sağlanabilir. Kullanıcılar, rol tabanlı veya öznitelik tabanlı erişim kontrolü gibi mekanizmalar aracılığıyla verilere güvenli bir şekilde erişebilirler (Gupta, Yamaguchi ve Agrawal, 2017). Bu süreç, veri güvenliğini artırırken erişim yönetimini de daha etkili hale getirir, böylece organizasyonlar yasal uyumluluk gereksinimlerini daha rahat karşılayabilir ve veri sızıntılarını önleyebilir (Bertino, 2016).

Veri Analitiği ve Makine Öğrenimi: Fonksiyonel mimarinin en önemli avantajlarından biri, analitik süreçlerin sistematik bir şekilde yürütülmesidir. Veri bilimciler, fonksiyonel mimari sayesinde veri işleme ve modelleme süreçlerini birbirinden bağımsız, ayrı modüller aracılığıyla yönetebilirler. Bu birbirinden ayrılmış modüller, her bir analitik görevin kendi özel ortamında çalışmasını sağlar, böylece farklı projeler birbirini etkilemeden ilerleyebilir. Örneğin, bir veri bilimci müşteri segmentasyonu üzerinde çalışırken, diğeri aynı veri gölü üzerinde satış tahmini yapabilir ve bu iki proje birbirinin kaynaklarını veya sonuçlarını etkilemez. Bu yaklaşım, veri gölü üzerinde eş zamanlı olarak yürütülen projelerin daha verimli ve güvenilir bir şekilde ilerlemesini sağlar. Ayrıca bu modüler yapı derin öğrenme ve makine öğrenimi algoritmalarının daha etkili bir şekilde eğitilmesine ve iyileştirilmesine olanak tanır (Abadi ve diğerleri, 2020).

Daha önce bölge mimarisi örneği olarak sunulan Lambda mimarisi, Sawadogo ve Darmont'un (2021) geliştirdiği topolojiye göre fonksiyonel mimari sınıfında yer almaktadır. John ve Misra (2017) tarafından geliştirilen bu mimari, yığın işleme, akan veri işleme ve sunum gibi işlevleri barındıran bileşenleriyle veri gölü işlevlerini etkin bir şekilde temsil etmektedir Sawadogo ve Darmont (2021), Lambda mimarisini depolama, işleme ve sunum süreçlerini sistematik bir şekilde yapılandırdığı için fonksiyonel mimari kategorisinde değerlendirmiştir.

Fonksiyonel mimariler, veri göllerinin sunmuş olduğu esneklik ve ölçeklenebilirlik avantajlarını daha sistematik ve yönetilebilir bir yapı altında toplamaktadır. Bu

sayede veri işleme süreçlerinin entegrasyonu kolaylaşmakta, veri yönetim süreçleri optimize edilmektedir. Böylece veri odaklı organizasyonların karar alma süreçlerine büyük katkı sağlanmaktadır (Sawadogo ve Darmont, 2021).

4.4.4.2. Olgunluk Tabanlı Mimariler

Olgunluk tabanlı mimari, veri gölünün zaman içinde nasıl olgunlaştığını ve veri yönetimi süreçlerinin nasıl geliştirildiğini gösteren bir yaklaşımdır. Bu mimari, organizasyonların veri yönetimi kapasitesinin aşamalı olarak gelişimini göz önünde bulundurarak, veri gölü altyapısının ve süreçlerinin kademeli olarak iyileştirilmesini sağlar (Sawadogo ve Darmont, 2021). Veri yönetiminde olgunluk, verilerin ham halden değerli bilgiye dönüştürülme süreci boyunca izlenir ve yönetim politikaları bu gelişime göre şekillendirilir. Bu yaklaşımda, veri gölünün olgunluk seviyesi arttıkça, uygulanan veri yönetimi pratikleri de gelişir. Örneğin, başlangıç aşamasında temel veri toplama ve depolama işlemleri yapılırken, ilerleyen aşamalarda veri kalitesi kontrolleri, üst veri yönetimi ve gelişmiş analitik uygulamalar devreye girer. Böylece veri gölü zamanla daha yapılandırılmış, güvenilir ve değer üreten bir sistem haline gelir.

Başlangıç Seviyesi (Initial Level): Bu aşamada, ham veriler doğrudan veri gölüne alınır. Verilerin doğruluğu veya kalitesi üzerinde düşük bir seviyede kontrol uygulanır. Veri yönetimi süreçleri genellikle önceden belirlenmiş bir plan olmaksızın, sorunlar ortaya çıktıkça müdahale edilerek yürütülür (Sawadogo ve Darmont, 2021). Bu seviyede, veri entegrasyonu sınırlıdır ve süreç genellikle insan müdahalesine dayalı şekilde yönetilir. Üst veri yönetimi yetersizdir veya hiç yoktur. Veri kalitesi kontrolleri oldukça sınırlıdır; veri güvenliği ile erişim kontrolü ise sadece en temel güvenlik önlemleriyle sağlanmaktadır.

Gelişen Seviye (Developing Level): İkinci aşamada, veri yönetimi süreçleri olgunlaşmaya başlar. Veriler temizlenir, düzenlenir ve belirli kategorilere ayrılır (Giebler ve diğerleri, 2019). Bu seviyede, veri entegrasyonu için otomatik süreçler

devreye girer. Temel üst veri yönetimi uygulanmaya başlanır ve veri kalitesi kontrolleri sistematik bir yapıya kavuşur. Veri güvenliği ve erişim kontrolü için daha gelişmiş politikalar uygulanır. Ayrıca basit analitik ve raporlama araçları kullanılmaya başlanır.

Olgunlaşmış Seviye (Mature Level): Bu seviyede, veri yönetimi süreçleri tam olarak olgunlaşmıştır. Veriler yüksek kalitededir ve iş zekâsı sistemlerine dahil edilmiştir (Sawadogo ve Darmont, 2021). Veri entegrasyonu tamamen otomatikleştirilmiş ve optimize edilmiştir. Kapsamlı üst veri yönetimi uygulanır ve ileri düzey veri kalitesi kontrolleri ile veri temizleme süreçleri mevcuttur. Güçlü veri güvenliği ve erişim kontrolü politikaları uygulanır. İleri analitik ve makine öğrenimi modelleri rutin olarak kullanılır ve veri yönetişimi politikaları tam olarak uygulanır.

Olgunluk tabanlı mimari, organizasyonların mevcut veri yönetimi kapasitelerini değerlendirmelerine ve gelecekteki hedeflerini belirlemelerine yardımcı olur. Bu yaklaşım, veri gölü uygulamalarının aşamalı olarak geliştirilmesine ve organizasyonun ihtiyaçlarına göre ölçeklendirilmesine olanak tanır (Giebler ve diğerleri, 2021). Olgunluk tabanlı mimarinin başarılı bir şekilde uygulanması için, organizasyonların düzenli olarak veri yönetimi pratiklerini değerlendirmeleri ve iyileştirmeleri gerekmektedir. Bu süreç, veri kalitesinin artırılması, veri güvenliğinin güçlendirilmesi ve veri analitiği kapasitesinin geliştirilmesi vb. alanlarda sürekli iyileştirmeyi gerektirir (Sawadogo ve Darmont, 2021). Bu mimari yaklaşım, özellikle büyük ve karmaşık veri ekosistemlerine sahip organizasyonlar için uygundur. Olgunluk tabanlı mimari, veri göllerinin zamanla daha etkili ve verimli kullanılmasını sağlayarak, organizasyonların veri odaklı karar verme süreçlerini güçlendirir ve veri bataklığı riskini azaltır (Giebler ve diğerleri, 2021).

4.4.4.3. Fonksiyonel ve Olgunluk Tabanlı Melez Mimariler

Melez mimariler, fonksiyonel ve olgunluk tabanlı mimarilerin avantajlarını bir araya getirerek veri gölü mimarilerinde daha esnek ve ölçeklenebilir bir çözüm

sunmaktadır. Sawadogo ve Darmont'un (2021) sınıflandırmasında, melez mimariler hem işlevsel optimizasyonu destekleyen modüler yapıyı hem de organizasyonun veri yönetimi olgunluğunu göz önünde bulunduran kademeli gelişim yapısını birleştirmektedir. Bu sayede organizasyonlar veri yönetim süreçlerini hem işlevselliğe hem de olgunluk düzeyine göre uyarlayabilmektedir.

Melez mimarilerde, veri gölü fonksiyonları (veri alımı, işleme, depolama ve analiz gibi) bağımsız modüller olarak yapılandırılır ve her modül organizasyonun mevcut olgunluk düzeyi ile uyumlu hale getirilir. Bu esneklik, organizasyonların başlangıç seviyesinde temel veri toplama ve depolama modüllerine öncelik vermesine olanak tanırken, daha olgun organizasyonların analitik süreçleri optimize etmek amacıyla ileri düzey veri işleme ve makine öğrenimi modüllerini entegre etmesini sağlar (Sawadogo ve Darmont, 2021).

Melez mimariler, organizasyonlara iki ana avantaj sunar. İlk olarak, veri yönetimi süreçleri organizasyonun ihtiyaçları ve veri yönetimi olgunluğu doğrultusunda esnek bir şekilde uyarlanabilir. Bu uyarlanabilirlik, veri gölünün her seviyede işlevsel verimlilik sunmasına olanak tanır ve organizasyonel değişimlere hızlıca adapte olabilen bir yapı oluşturur. İkinci olarak, melez mimariler olgunluk seviyesine göre farklı fonksiyonel modüllerin birbirine entegre edilmesini sağlayarak veri güvenliği ve veri analitiği süreçlerinin bütünleşik bir yapıda yönetilmesine imkân tanır (Hai ve diğerleri, 2021).

Kısaca melez mimariler, veri göllerinin hem işlevselliğini optimize ederek hem de organizasyonun olgunluk düzeyine göre uyarlanmasını sağlayarak daha etkili bir veri yönetimi çözümü sunar. Bu yaklaşım, özellikle büyük ve çeşitli veri setleriyle çalışan organizasyonların veri göllerini yönetmelerini ve optimize etmelerini kolaylaştırır.

4.4.5. Bulut Tabanlı Veri Gölleri (Cloud-Based Data Lakes)

Bulut tabanlı veri gölleri, modern veri yönetimi stratejilerinin önemli bir parçasıdır. Bu çözümler (örneğin, AWS S3, Google Cloud Storage ve Azure Data Lake) verilerin dağıtık bir altyapı üzerinde depolanmasını ve geniş veri analitiği işlemleri için optimize edilmesini sağlar (Fang, 2015). Bu sistemler özellikle ölçeklenebilirlik ve maliyet etkinliği açısından avantajlıdır. Organizasyonlar, ihtiyaçları doğrultusunda depolama kapasitelerini hızlıca ayarlayabilirler. Ayrıca bulut sağlayıcılarının sunduğu güvenlik önlemleri ve bütünleşik analitik araçlar, veri yönetimini kolaylaştırır (Mathis, 2017). Ancak, verilerin buluttan alınması veya farklı bulut sağlayıcılarına taşınması sırasında oluşan veri taşıma maliyetleri ve bazı sektörlerdeki veri lokalizasyonu gereksinimleri, bulut tabanlı çözümlerin kullanımını sınırlayabilir. Bu nedenle, organizasyonlar bulut tabanlı veri gölü stratejilerini belirlerken, bu faktörleri dikkatle değerlendirmelidir (Sawadogo ve Darmont, 2021).

Veri gölü mimarisi seçimi, organizasyonun ihtiyaçlarına, veri tiplerine ve kullanım senaryolarına bağlı olarak yapılmalıdır. Bölge mimarisi, veri güvenliği ve süreç yönetimi açısından güçlü bir yapı sunarken, gölet mimarisi esneklik ve hız kazandırabilir. Her mimari, farklı kullanım durumları ve veri işleme gereksinimlerine göre tasarlanmıştır ve farklı avantajlar sunmaktadır.

4.4.6. Veri Göllerinde Altyapı, Arama ve Analitik Çalışmalar

4.4.6.1. Altyapıyı (Infrastructure) Kullanma

Daha önce açıklandığı gibi, bir şirkete veya kuruma ait tüm verinin herhangi bir düzenleme yapılmadan doğrudan bir veri gölüne gönderilmesi, bir süre sonra veri gölünü "tek yönlü" hale getirir. Yani, veri gölüne sürekli veri akışı sağlanır; ancak bu verilerin analiz edilmesi mümkün olmaz. Bu nedenle, kurumsal veri göllerinin belirli bir sistematik mimari dikkate alınarak oluşturulması ve verilerin bu yapı çerçevesinde yönetilmesi gerekmektedir. Kurumların veri ambarı ve benzeri

depolama ve veri yönetim çözümlerinden veri gölü mimarisine geçiş süreçleri, emek, bütçe ve zaman gerektirir. Ancak, bu maliyetlerin karşılanması sonucunda şirketler, verilerini analiz ederek değer üretebilecekleri büyük bir varlığa sahip olurlar.

Basit bir örnekle büyük bir şirketin yıllık gelirlerini farklı para birimlerinde kaydederek veri gölünde tuttuğunu varsayalım. Bu durumda şirketin yıllık gelirini hesaplaması mümkündür ancak gelirin ilgili tarihteki değeri baz alınarak istenen para birimde hesaplanması vakit alacaktır (Inmon, 2016, s. 140). Ancak şirket entegre bir veri gölünde gelir verilerini tuttuğunda yıllık gelirini hızla elde etmesi mümkün olacaktır (Inmon, 2016, s. 141). Kurumların uygun bir veri gölü mimarisi kullanması verinin yük olmasını engelleyerek istenilen veriye hızlıca ulaşabilmesini sağlayacak ve uzun vadede toplam maliyeti azaltacaktır (Inmon, 2016, s. 149).

Veri gölleri için dokümantasyon kritik öneme sahiptir. Veri göllerindeki verilerin anlaşılarak analiz edilmesi için uygun şekilde belgelenmiş olmaları gerekir (Sawadogo ve Darmont, 2021). Veri gölü ve seçilen veri gölü mimarisi için en az iki tür dokümantasyona ihtiyaç duyulur. İlk dokümantasyon yüksek sistem seviyesinde olup verilerin veri gölü ve alt modüllere nasıl aktarıldığı, modüller arasında nasıl veri transferi yapıldığı gibi konularındaki bilgileri içerir. İkinci seviye dokümantasyon ise veri gölü mimarisi ile ilgili daha detaylı bilgi içerir. Bu bilgiler veriye ait üst veri, veri mimarisinin alt modüllerinde gerçekleşen faaliyetlerle ilgili üst işleme verileri, veri dönüşüm süreçleri, veri akışlarının mimari açıklaması, veri seçim kriterleri ve veri çıkarma kriterleri gibi açıklamaları içermelidir (Giebler ve diğerleri, 2019). Veri gölündeki ve alt modüllerdeki verilere etkili ve verimli bir şekilde ulaşılabilmesi ve bu verilerin nasıl işleneceği hakkında bilgi sahibi olunabilmesi için uygun dokümantasyon yapılmalıdır (Inmon, 2016, s. 179, 180).

4.4.6.2. Arama ve Analiz

Veri analizi ile verinin özetlenmesi ve görselleştirilmesi mümkün olur. Veri analizi için önce ilgili veriye ulaşmak, sonra ise ulaşılan veriyi analiz etmek gerekir. Veriler veri gölüne uygun bir şekilde kaydedilmişse ve dizinlenmişse istenilen veriye ulaşmak oldukça kolay olacaktır (Miloslavskaya ve Tolstoy, 2016). Verilerin analiz edilmesi için farklı araçlar bulunmaktadır. Veri analizi türüne göre basit veya karmaşık olabilir.

Veri aramanın farklı şekilleri bulunmaktadır. Belirli bir veri seti için veya daha büyük bir veri kümesi için arama yapılabilir. Örneğin tek bir kişi için son tıbbi kayıtlar aranabilir veya belli bir gruba ait bütün bireyler için son tıbbi kayıtlara ulaşmak istenebilir (Inmon, 2016, s. 155).

Şifreli veriler veya gizli amaçlarla işletildiği için zayıf olarak işaretlenen verilerde arama yapmak oldukça zordur. Yine aradığımız verinin sınırlarının geniş olması, örneğin ABD'de yaşayan bir kadın için arama yapılması, ABD'de yaşayan bütün kadınlar içinde arama yapılmasını gerektirdiğinden zor ve verimsiz olacaktır (Inmon, 2016, s. 152). Arama kriterlerinin net olmadığı durumlarda makine öğrenimi ve kavram arama kullanılan yöntemlerdendir (Inmon, 2016, s. 154). Arama ve niteleme (search and qualify) teknolojisi, metin verisi gibi optimal olarak organize edilmemiş veriler içinde karmaşık aramalar yapılabilmesine olanak sağlar. Elasticsearch ve Apache Solr gibi araçlar, bu tür arama ve analiz yetenekleri sunar (Sawadogo ve Darmont, 2021).

Veri göllerinde depolanan ham ve işlenmemiş verilerin içinde arama yapmak ve istenilen veriye ulaşmak oldukça zordur. Bu zorluğun en büyük nedeni verinin entegre edilmemiş ve yapılandırılmamış olmasıdır (Sawadogo ve Darmont, 2021). Verinin diğer veriler arasında gizlenmesi veya diğer verilerden ayırt edilememesi veri aramayı zorlaştırmaktadır. Veri arama kriterlerinin belirsiz olması bir diğer problemdir. Bir veriye ulaşıldığında bu verinin istenilen veri olduğundan emin olmak zor olmakta veya veriye ulaştıktan sonra veriyi dönüştürmek gerekebilmektedir (Inmon, 2016, s. 157). Koşullandırılmış (işlenmiş

ve yapılandırılmış) veri göllerindeki veri entegre edilmiş, filtrelenmiş, dönüştürülmüş, organize edilmiş ve düzenlenmiş olduğu için bu göllerde veri aramak ve bulmak kolaydır (Inmon, 2016, s. 160).

Veri analiz etmenin birçok yöntemi bulunmaktadır. Basitçe verilerin sıralanması, verilerin özetlenmesi, verilerin karşılaştırılması, aykırı değerlerin bulunması bu yöntemlerin bir kısmıdır. Verilerin bir diyagram veya resimle temsil edildiği görselleştirme büyük miktarda verinin sonuçlarının hızlıca görülmesini sağladığı için güçlü bir analiz yöntemidir (Inmon, 2016, s. 163). Veri analizi farklı lokasyonlarda yapılabilmektedir. Çevrim içi, gerçek zamanlı verilerin analizi çevrim içi işletim sistemlerinde yapılır (Inmon, 2016, s. 184). Veri göl ve göletlerinde ise daha çok model analizi, ilişki analizi, sebep sonuç analizi ve uzun vadeli trend analizi gibi analizler yapılır (Inmon, 2016, s. 190).

4.4.6.3. Veri Analitiği ve Entegrasyon Araçları

Veri gölü ve veri gölü mimarileri içinde kullanılabilen farklı işlevleri bulunan çeşitli araçlar bulunmaktadır (Inmon, 2016, s. 164). Bu araçlar, görselleştirme, veri analizi entegrasyon ve yönetim gibi çeşitli amaçlara hizmet eder.

Görselleştirme araçları, verileri düzenleyerek ve görüntüleyerek, başka türlü anlaşılması zor olabilecek kalıp ve eğilimlerin hızlıca görülebilmesini sağlar. Tableau, Power BI ve Qlik Sense gibi popüler görselleştirme araçları, kullanıcıların interaktif arayüzler ve raporlar oluşturmasına olanak tanır (Giebler ve diğerleri, 2019). Bu araçlar, verilerin ilişkisel formatta saklanmasını gerektirir (Inmon, 2016, s. 194).

İstatistiksel analiz, büyük veri setlerini analiz etmek ve sonuçları grafiksel olarak göstermek için kullanılan önemli bir yöntemdir (Inmon, 2016, s. 199). Örneğin, bir firmanın geçmiş satış verileri analiz edilebilir ve gelecekteki ürün taleplerini tahmin edebilen istatiksel modeller oluşturulabilir (Sawadogo ve Darmont, 2021). Klasik ETL (Extract, Transform, Load) işleme yöntemi, uygulama tabanlı verileri okur ve

entegre edilmiş kurumsal verilere dönüştürür. Apache Nifi, Talend ve Informatica gibi ETL araçları, veri entegrasyonu ve dönüşümü için yaygın olarak kullanılmaktadır (Sajida ve Ramakrishna, 2015). Bu araçlar, veri gölü ortamının oluşturulması ve desteklenmesine katkı sağlar (Inmon, 2016, s. 201).

4.5. TÜRKİYE'DE KURUMLARDA VERİ MİMARİSİ GELİŞTİRME SÜREÇLERİ VE VERİ GÖLLERİNİN KULLANIMI

4.5.1. Türkiye'deki Mevcut Veri Sistemleri: Türkiye'deki Kurumların Veri Mimarisi Altyapılarının Analizi ve Mevcut Durum

Türkiye'deki veri yönetim uygulamaları, genellikle geleneksel yapılandırılmış veri odaklı sistemlere dayanmaktadır. Ancak, artan veri hacmi ve çeşitliliği, modern veri mimarilerine olan ihtiyacı daha da belirgin hale getirmiştir. Bu bölümde, Türkiye'deki önemli kurumların mevcut veri mimarisi altyapıları incelenmekte, bu sistemlerin güçlü ve zayıf yönleri değerlendirilmekte ve karşılaşılan temel zorluklar ele alınmaktadır.

Bu bağlamda, Türkiye'deki mevcut veri mimarisi altyapılarının sınırlamaları ve modern yaklaşımlara olan ihtiyacın, yalnızca kurumlar düzeyinde değil, akademik çalışmalarda da olduğu söylenebilir. Akademik literatürde veri gölü kavramının yeterince ele alınmamış olması, yenilikçi veri mimarilerinin benimsenmesi ve uygulanması için gerekli olan teorik ve metodolojik birikimin oluşturulmasını engellemektedir. Bu tez kapsamında yapılan arama sonuçlarına göre, DergiPark'ta "Veri Gölü" (ve "Verigölü") terimlerini içeren hiçbir yayına rastlanmazken, Web of Science veri tabanında "Data Lake" (ve "DataLake") terimlerini içeren toplam 615 yayın listelenmiştir. Bu durum, uluslararası akademik çalışmalarda oldukça önem verilen veri gölü konusunun Türkiye’de henüz sistematik bir şekilde ele alınmadığını ve yenilikçi veri mimarileri konusundaki akademik çalışmaların yetersizliğini ortaya koymaktadır.

Bu bağlamda, Türkiye'deki akademik çalışma eksiklikleri, veri gölü ve modern veri mimarileri konusundaki uygulamaların yaygınlaşmasını ve geliştirilmesini sınırlayan önemli bir etken olarak karşımıza çıkmaktadır. Mevcut akademik eksikliklerin, özellikle kamu kurumlarında yenilikçi veri yönetimi stratejilerine geçişi de geciktirdiği görülmektedir.

Bu eksikliğin somut bir yansıması olarak, T.C. Sağlık Bakanlığı'nın veri yönetimi süreçlerinde geleneksel veri ambarı yaklaşımını benimsemiş olması dikkat çekmektedir. Bakanlığın resmî internet sitesinde yer alan "Veri Ambarı ve Büyük Veri Birimi" ("T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü", 2024a) ve "Büyük Veri Uygulamaları ve Veri Yönetimi Koordinatörlüğü" ("T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü", 2024b) sayfalarında bu strateji ayrıntılı olarak açıklanmaktadır. Veri ambarları, yapılandırılmış sağlık verilerinin etkin yönetimini sağlamada önemli bir araç olsa da hızla artan dijitalleşme ile birlikte yapılandırılmamış verilerin de üretilmesi, bu geleneksel yaklaşımların tek başına yeterli olamayacağını göstermektedir. Sağlık verilerinin çeşitliliği ve karmaşıklığı, modern veri yönetimi stratejilerinin yalnızca veri ambarları ile sınırlı kalmaması, büyük veri teknolojileri ve veri gölleri gibi daha esnek ve kapsamlı çözümleri de içermesi gerektiğini göstermektedir (Raghupathi ve Raghupathi, 2014).

Bir başka örnek olarak TÜBİTAK ULAKBİM, 3 petabaytın üzerinde kapasiteye sahip veri ambarı altyapısıyla ulusal ölçekte projelere destek sağlamaktadır. Bu altyapı, Türkiye Deprem Veri Merkezi gibi projelerde veri paylaşımına olanak tanıyarak araştırmacıların ihtiyaçlarını karşılamaktadır ("TÜBİTAK ULAKBİM", t.y.). Bununla birlikte TÜBİTAK, son yıllarda veri yönetiminde daha yenilikçi yaklaşımlar geliştirmeye odaklanmıştır. Bu kapsamda, FAIR (Findable, Accessible, Interoperable, Reusable) veri prensiplerini benimseyerek, Türkiye'de bilimsel veri yönetimi süreçlerinin daha işbirlikçi ve erişilebilir hale gelmesine katkı sağlamaktadır (Türkyılmaz-van der Velden, 2021).

Gelir İdaresi Başkanlığı (GİB) da vergi mükelleflerine ait bilgileri yönetmek amacıyla veri ambarı teknolojilerini kullanmaktadır. GİB'in veri mimarisi, Defter Beyan Sistemi, Hazır Beyan Sistemi, İade Takip Sistemi ve KDVİRA Sistemi gibi çeşitli modülleri içermektedir ("Gelir İdaresi Başkanlığı", t.y.). Bu modüller, geniş kapsamlı verilerin zamanında işlenmesi ve raporlanması için güçlü bir altyapı sunmaktadır ancak veri ambarı tabanlı mimarinin sınırlılıkları ve hızla artan veri hacimleri dikkate alındığında özellikle büyük veri ve yapılandırılmamış veri yönetiminde zorluk yaşanabileceği düşünülmektedir.

Türkiye'de veri gölü uygulamalarına yönelik kamu kurumlarında örnekler sınırlı olmakla birlikte, Türkiye İstatistik Kurumu (TÜİK) tarafından, TÜBİTAK BİLGEM işbirliği ile yürütülen "Büyük Veri İleri Analitik Projesi" bu alanda önemli bir girişimdir. Bu projede, John ve Misra (2017) tarafından geliştirilen Lambda mimarisi temel alınarak bir veri gölü modeli geliştirilmiş ve kullanılmıştır. İnternet sitelerinden günlük fiyat bilgileri ve iş ilanları gibi verilerin toplanması ve işlenmesi süreçlerinde yığın ve akan veri işleme yaklaşımları bir arada kullanılarak gerçek zamanlı analiz sağlanmıştır (B3LAB, 2020; TÜİK, 2020). Proje, Türkiye'de kamu kurumlarının modern veri yönetimi stratejilerine uyum sağlaması açısından dikkat çekmektedir.

Ayvaz ve Salman (2020), Türkiye'deki firmaların büyük veri teknolojilerini kullanma olgunluğunu inceledikleri çalışmada, firmaların bu alanda ilerleme kaydettiğini, ancak olgunluk seviyelerinin henüz yeterince yüksek olmadığını belirtmektedir. Bu durum, modern veri yönetimi çözümlerinin, özellikle veri gölü gibi esnek ve büyük veri odaklı yaklaşımların, daha geniş ölçekte benimsenmesi gerektiğini ortaya koymaktadır. Kurumların bu tür çözümlere yatırım yaparak stratejik kararlar alması ve bu bağlamda gerekli yönlendirmeleri yapması, küresel rekabet ortamında geri kalmamaları açısından önemlidir.

Türkiye'de veri bilimi ve analitiği alanında Sabancı Üniversitesi Veri Analitiği Araştırma ve Uygulama Merkezi (VERİM) önemli çalışmalar yapmaktadır. VERİM, büyük veri analitiği, yapay zekâ ve makine öğrenimi gibi alanlarda öncü projeler

yürütmektedir (“VERİM”, 2024). Veri teknolojileri alanındaki araştırma merkezleri, Türkiye’deki veri yönetimi kapasitelerinin artırılmasına ve modern veri mimarilerinin daha yaygın bir şekilde benimsenmesine katkı sağlamaktadır.

Türkiye’deki kamu kurumları veri yönetiminde çoğunlukla yapılandırılmış veriler için tasarlanmış veri ambarı ve ilişkisel veri tabanı gibi geleneksel mimarileri kullanmaları nedeniyle büyük hacimli ve yapılandırılmamış veri ihtiyaçlarında modern veri mimarilerinin sunduğu esnekliği tam olarak değerlendirememektedir. Bu eksiklik, büyük veri çağında hızlı karar alma ve yenilikçi projeleri hayata geçirme süreçlerinde kurumları önemli ölçüde kısıtlamaktadır. Bu durum, tezde ortaya konulan “Kamu kurumlarının mevcut veri süreçlerinde kullanılan teknolojiler, büyük veri çağının gereksinimlerini karşılamamaktadır” hipotezini destekler niteliktedir. TÜİK Büyük Veri İleri Analitik Projesi gibi örneklerin genişletilmesi ve veri gölü gibi modern veri mimarilerinin benimsenmesi, kurumların dijital dönüşüm süreçlerini hızlandırmalarına ve veri yönetimi kapasitelerini artırmalarına olanak sağlayacaktır.

4.5.2. Veri Göllerinin Kurumlarda Veri Yönetimi Süreçlerine Sağlayabileceği Faydalar ve İlgili Zorluklar

Veri gölleri, kurumlardaki büyük veri yönetimi ve veri analitiği süreçlerini dönüştürebilecek önemli bir çözüm olarak öne çıkmaktadır. Veri gölleri, yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verileri aynı ortamda saklama ve işleme yeteneği sunar. Türkiye'de kamu kurumları, araştırma merkezleri, belediyeler vb. kurumlar veri gölleri sayesinde büyük veri analizi ve veri paylaşımı alanlarında önemli adımlar atabilirler. Örneğin, T.C. Sağlık Bakanlığı'nın e-Nabız sistemine veri gölleri entegre edildiğinde, hasta bilgileri, sosyal medya verileri ve IoT sensörlerinden gelen sağlık verileri birleştirilerek daha kapsamlı analizler yapılabilir. Bu entegrasyon, bazı hastalıkların erken teşhisini sağlayabilir, salgın hastalıkların yayılmasını tahmin etmede yardımcı olabilir, hastane kaynaklarının daha etkin kullanımını mümkün kılabilir. Böylelikle

kişiselleştirilmiş sağlık hizmetlerine geçişi hızlandırabilir, daha etkili kararlar alınmasını sağlayabilir ve böylece sağlık hizmetlerinin kalitesini artırabilir (Gökalp, Gökalp, Çoban ve Eren, 2018a; Raghupathi ve Raghupathi, 2014).

TÜBİTAK ULAKBİM, yüksek kapasiteli veri ambarlarıyla ulusal ölçekte projeler yürütmektedir. Türkiye Deprem Veri Merkezi gibi oluşumların veri paylaşımına olanak sağlayan bu altyapı, araştırmacıların ihtiyaçlarını karşılamaktadır ("TÜBİTAK ULAKBİM", t.y.). Ancak, TÜBİTAK gibi kurumların mevcut veri ambarı altyapılarını veri gölü teknolojisine dönüştürmesi, bilimsel araştırma ve inovasyon süreçlerinde önemli avantajlar sağlayabilir. Deprem araştırmaları örneğinde, veri gölü tabanlı bir altyapı sayesinde sismograf sensör kayıtları (yapılandırılmamış büyük log dosyaları), KKBS (Küresel Konum Belirleme Sistemi - Global Navigation Satellite Systems) verileri ve yer altı jeolojik haritalar gibi farklı formatlardaki veriler tek bir mimaride bir araya getirilebilir. Bu yaklaşım, veriler üzerinde gelişmiş analiz yöntemlerinin uygulanmasını kolaylaştırır. Ayrıca, schema-on-read ilkesinin sağladığı esneklik sayesinde, araştırmacılar katı bir şema tanımlamak zorunda kalmadan farklı veri tipleriyle bilimsel çalışmalarını yürütebilir. Veri gölü altyapısının benimsenmesi, büyük miktarda bilimsel yayının, araştırma verilerinin, proje sonuçlarının ve yapılandırılmamış verilerin tek bir merkezde toplanmasına olanak tanır. Bu dönüşüm, araştırmaların daha hızlı ve kapsamlı analiz edilmesini, disiplinler arası çalışmaların kolaylaşmasını ve bilimsel inovasyon süreçlerinin hızlanmasını sağlayabilir (Türkyılmaz-van der Velden, 2021). Yine bu dönüşüm, Özdemirci'nin (2019a) vurguladığı "Veri-Bilgi ve Belge Analizi Modülü" konseptiyle uyumlu olarak, ulusal ve uluslararası platformlarda devletin ihtiyaç duyduğu veri-bilgi ve belgelere hızlı bir şekilde erişilmesini ve analizini de kolaylaştırabilir.

Bankacılık sektöründe veri gölleri, yapılandırılmış finansal verilerin yanı sıra sosyal medya etkileşimleri, müşteri hizmetleri kayıtları ve mobil uygulama verilerini birleştirerek daha kapsamlı risk analizleri, müşteri segmentasyonları ve dolandırıcılık tespiti yapılmasına olanak tanır. Örneğin, Dong, Liao ve Zhang (2018) finansal sosyal medya verilerinin dolandırıcılık tespiti için nasıl

kullanılabileceğini, bankalara bu verilerle risk değerlendirmelerinde nasıl destek sağlanabileceğini incelemektedir. Belediyeler ve benzeri kurumlar, IoT cihazlarından gelen büyük miktardaki sensör verisini gerçek zamanlı olarak yönetmek için veri göllerini kullanabilirler. Bu yaklaşım, trafik yönetimi, enerji kullanımı ve çevre izleme gibi alanlarda daha etkili kararlar alınmasına olanak sağlayabilir (“T.C. Çevre, Şehircilik”, t.y.). Veri göllerinin bu amaçla kullanılması, akıllı şehir uygulamalarının etkinliğini artırma ve vatandaşlara sunulan hizmetleri iyileştirme potansiyeli taşımaktadır. Veri gölleri, geleneksel veri ambarlarına kıyasla daha düşük depolama maliyetleri ve daha yüksek esneklik sağlar. Bu durum, özellikle bütçe kısıtlamalarının söz konusu olduğu kamu kurumları için önemli bir avantaj sunar (Giebler ve diğerleri, 2021).

Veri gölleri, EBYS kullanan tüm kurumlar için önemli avantajlar sunmaktadır. Türkiye'de kamu kurumlarında yaygın olarak kullanılan EBYS'lerin (Özdemirci, 2019b; Külcü, 2009) veri gölü tabanlı mimarilere entegre edilmesi, uygulama ve metin verileri gibi farklı veri türlerinin daha etkin bir şekilde yönetilmesi ve kullanılmasının önünü açacaktır. EBYS’ler, kamu kurumlarının yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerini üreten temel kaynaklardan biridir (Kutlutürk ve Özdemirci, 2023). Veri göllerinin bu veri yapılarıyla çalışabilme kapasitesi, veri gölü tabanlı mimarileri EBYS'ler için doğal bir altyapı olarak öne çıkarmaktadır. Geleneksel veri ambarları yerine veri gölü tabanlı bir mimariyle entegrasyon, EBYS'lerde üretilen verilerden elde edilebilecek bilgilerin kapsamını ve derinliğini artırma potansiyeline sahiptir.

EBYS’lerin mevcut veri tabanı mimarilerinden veri gölü altyapılarına geçişi, köklü dönüşümleri beraberinde getirmektedir. EBYS’ler genellikle ilişkisel veri tabanları, veri ambarları veya klasik dosya sistemleri üzerine inşa edilmiştir. Ancak veri gölleri, yapılandırılmış ve yapılandırılmamış veri kümelerini merkezi bir platformda birleştirerek büyük boyutlu dokümanların (PDF, görüntü, taranmış belge vb.) doğrudan veri gölünde saklanmasını mümkün kılmaktadır (Khine ve diğerleri, 2021). Bu geçiş sürecinde, EBYS’lerin mevcut veri tabanı mimarileri üzerinde düzenlemeler yapılması, üst veri yönetiminin yeni altyapıya uyarlanması ve veri

gölü entegrasyon süreçlerinin (ETL, API vb.) geliştirilmesi gerekebilir. Ayrıca, kullanıcıların yeni altyapıyı daha etkin bir şekilde kullanabilmesi için eğitim programlarının uygulanması ve değişim yönetimi stratejilerinin hayata geçirilmesi önem taşımaktadır.

Veri gölü teknolojisinin benimsenmesi, avantajların yanı sıra bazı zorlukları da beraberinde getirebilir. Geçiş süreci, entegrasyonun yeniden tasarlanmasını, teknik uyumun sağlanmasını ve kullanıcıların yeni sistemlere uyarlanması için gerekli eğitimlerin verilmesini gerektirir. Veri yönetişimi politikalarının oluşturulması, veri kalitesi ve güvenliği standartlarının belirlenmesi ve kurumlar arası iş birliğinin güçlendirilmesi bu dönüşümün sürdürülebilirliği açısından kritik öneme sahiptir (Uslu, 2023). Bununla birlikte, Türkiye'nin genç ve teknolojiye yatkın nüfusu, bu teknolojinin hızlı bir şekilde benimsenmesi için önemli bir fırsat sunmaktadır.

Sonuç olarak, veri gölleri Türkiye'deki kurumlar için büyük bir potansiyel sunmakta ve dijital dönüşüm süreçlerine hız kazandırabilecek önemli bir araç olarak öne çıkmaktadır. Ancak bu teknolojinin etkin kullanılabilmesi için sadece teknik altyapı değil, aynı zamanda organizasyonel ve kültürel dönüşümlerin de gerçekleştirilmesi gerekmektedir.

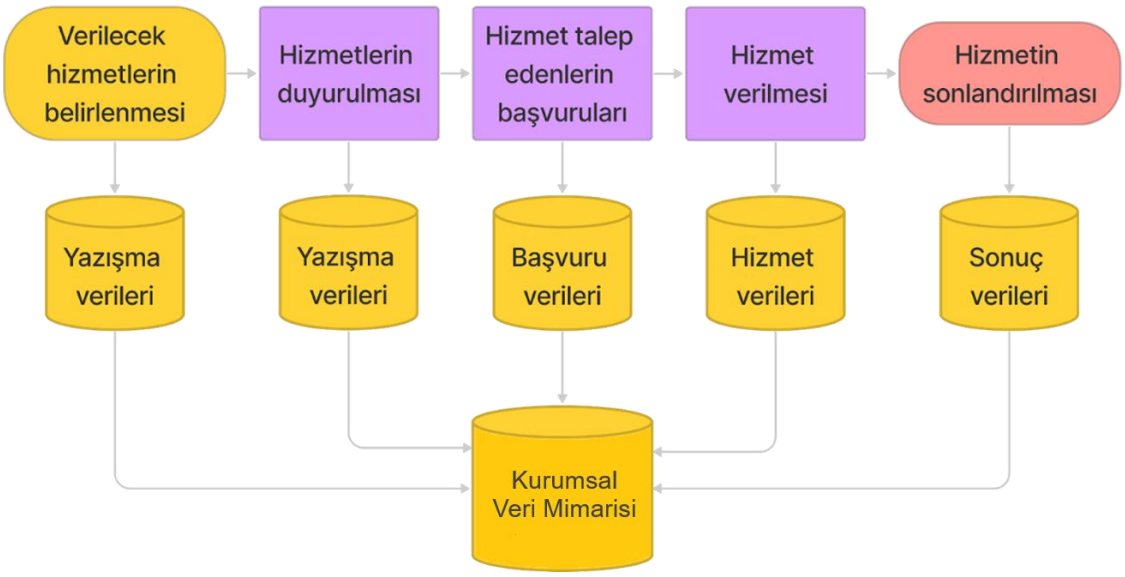
Literatür taraması ve sistem analizi teknikleri ile oluşturulan Veri Mimarisi Geliştirme ana başlıklı bu bölümde, kapsamlı bir şekilde veri gölleri, veri göllerinin avantajları, farklı veri gölü mimarileri ve başarılı veri gölü uygulamaları ele alınmıştır. Bölümde, veri gölleri ile ilgili somut teknik verilere dayanılarak ve literatürden destek alınarak, tezde sunulan "Veri gölü mimarilerinin uygulanması, kamu kurumlarının veri yönetim süreçlerini iyileştirebilir" hipotezi desteklenmiştir. Bu hipotez, modern veri gölü mimarilerinin kamu kurumlarının veri yönetiminde esneklik, verimlilik ve ölçeklenebilirlik gibi alanlarda sağladığı katkılara işaret etmektedir.

5. BÖLÜM
BULGULAR VE DEĞERLENDİRME: VERİ MİMARİSİ MODELİ

5.1. TÜRKİYE'DE KURUMSAL İŞ SÜREÇLERİ VE KURUMSAL VERİLER

Bu bölümde, kurumsal iş süreçleri ve bu süreçlerde elde edilen veriler örneklendirilmiş ve bu süreçlerin kavramsal modelleri oluşturulmuştur. Böylece kamu kurumlarının veri yönetimi ihtiyaçlarına uygun bir altyapı oluşturulması ve etkili bir veri mimarisi geliştirilmesine yönelik adımlar atılması hedeflenmiştir.

Şekil 25'te bir araştırma kurumu örneğinde gözlemlere dayanarak verilecek hizmetlerin belirlenmesinden sonlandırılmasına kadarki süreçleri ve bu süreçlerde oluşan verileri gösterir örnek şema verilmiştir.

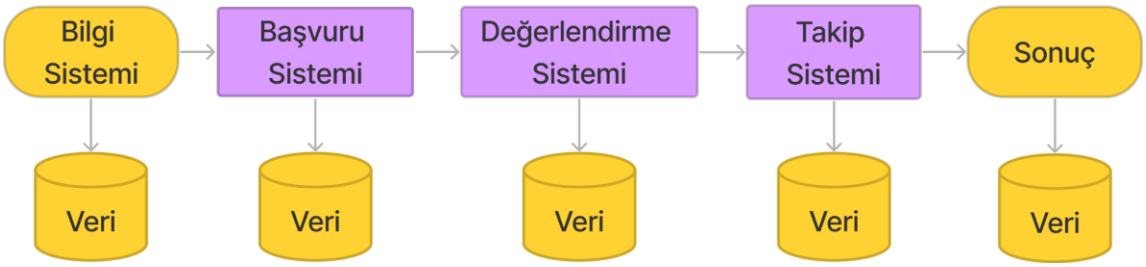


Şekil 25. Kurumsal hizmetlerin belirlenmesinden sonlandırılmasına kadarki süreçleri ve veri oluşumunu gösterir şema

Şekilde gösterildiği üzere, bir kurumun mevzuatı doğrultusunda belirlenen kurum hedeflerinin gerçekleştirilmesi için öncelikle hangi hizmetlerin verileceği belirlenir. Daha sonra verilecek hizmetler çeşitli kanallar (kurumun internet sitesi, ilgili

paydaşlara e-posta gönderimi, sosyal medya vb.) aracılığı ile duyurulur ve hizmet talep edenlerin başvuruları alınarak yapılan değerlendirmeler sonucunda hizmet verilmesi uygun görülen kişilere talep edilen hizmet verilir ve hizmet için belirlenen süre sona erdiğinde hizmet sonlandırılır. Bütün bu aşamalarda yazışma verileri, başvuru verileri, hizmet verileri ve sonuç verileri gibi farklı veriler elde edilir.

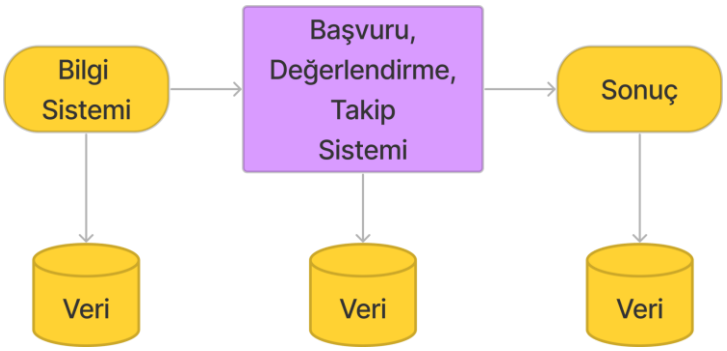
Birçok kamu kurumunda bilgi ve belge süreçleri farklı sistemler üzerinden yürütülmektedir. Aynı kurum içinde dahi farklı birimlerde süreçler farklı yönetilebilmekte, farklı sistemler kullanılabilmektedir. Sunulan hizmet süreçlerinde farklı iş uygulama yazılımları ve veri yönetim sistemleri kullanılabilmektedir. Kurumsal bilgi sistemleri ve veri oluşumunu gösterir örnek bir şema Şekil 26'da verilmiştir.



Şekil 26. Kurumsal bilgi sistemleri ve veri oluşumunu gösterir örnek şema 1

Bu şemada örnek olarak akademisyenlerin başvuruda bulunabileceği burs, proje desteği, altyapı desteği vb. bir hizmet örnek olarak verilmiştir. Genellikle bu hizmetlerde kurumlar (TÜBİTAK, YÖK, Bakanlıklar vb.) hizmete başvuruda bulunacak kişinin öncelikle merkezi bir bilgi sistemine (ARBİS, YOKSİS vb.) kaydolmasını istemektedir. Örnek bir destek başvurusunda, akademisyenin kurumdaki ilgili birimin başvuru sistemi üzerinden başvuruda bulunması gerekir. Daha sonra başvuru verileri farklı bir değerlendirme sistemine aktarılır ve başvuru değerlendirilerek sonucu belirlenir. Destek kapsamına giren başvurulara ait veri yine başka bir takip sistemine aktarılarak hizmet sona erene kadar gerekli süreçler takip edilir. Destek süresi sona erdiğinde hizmet sonuçlandırılarak süreç tamamlanmış olur. Farklı bir hizmette ise başvuruda bulunacak kişi, bilgi

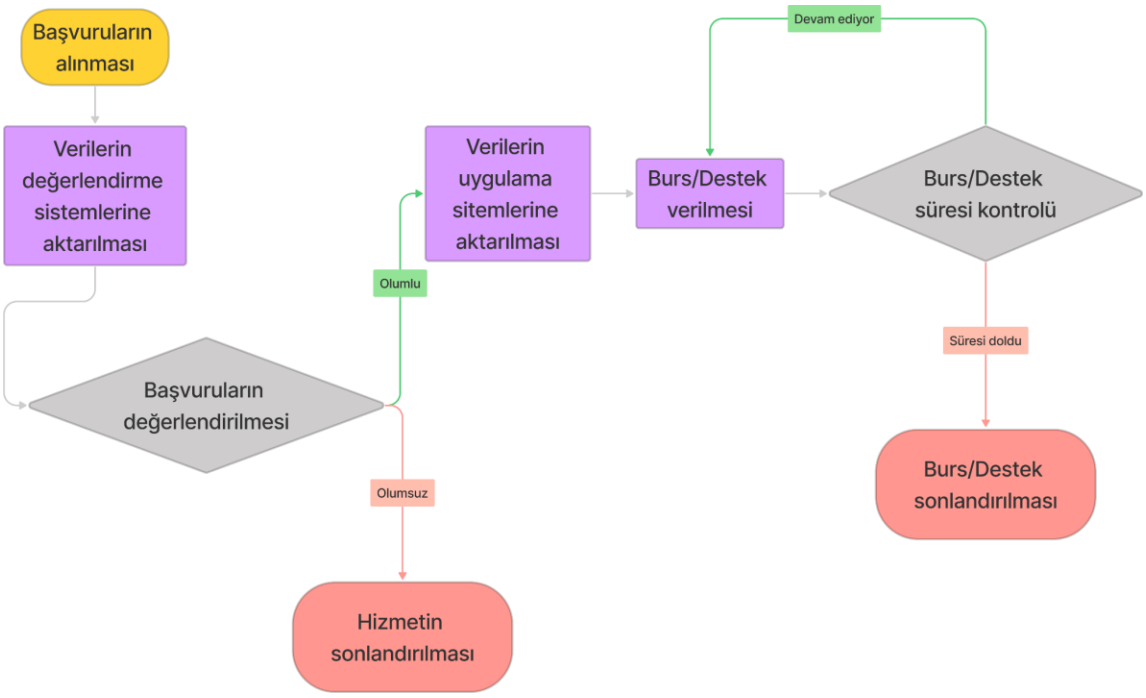
sistemine kaydolduktan sonra, değerlendirme ve takip süreçlerinin tek bir sistem üzerinden yürütülmesi mümkündür. Şekil 27'de ise bu sürece ait şema verilmiştir.



Şekil 27. Kurumsal bilgi sistemleri ve veri oluşumunu gösterir örnek şema 2

Farklı birimlerin farklı bilgi sistemlerini kullanması veya süreçleri farklı şekilde yürütmesi sebebi ile kurum içinde çalışan bir personelin belirli bir kullanıcının farklı hizmetlere olan başvurularını ve süreçlerini takip etmek için farklı arayüzler ve sistemler üzerinden kişinin kurum ile ilişkisini ve sunduğu verileri takip etmesi gerekir. Birimlerin süreçleri farklı biçimlerde yürütmesi hizmete başvuran kişinin farklı arayüze sahip çeşitli sistemlere bilgilerini tekrar tekrar girmesine sebep olarak verimliliği düşürecektir.

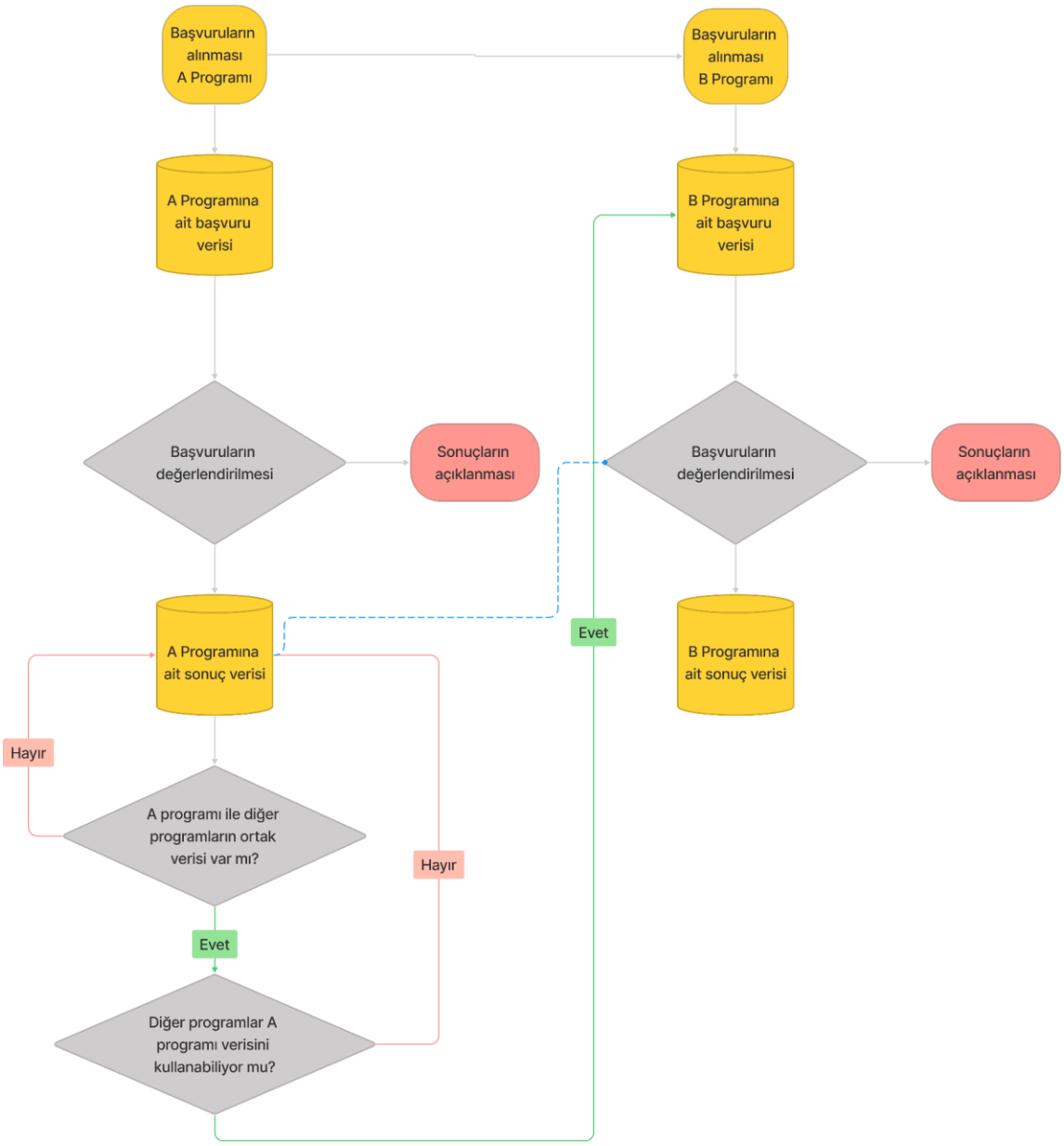
Başvuru sistemi aracılığıyla başvuruların alınmasından verilen hizmetinin sonlandırılmasına kadarki örnek bir süreci gösteren iş akış şeması Şekil 28'de gösterilmektedir.



Şekil 28. Kurumsal bir hizmetin sunulmasına yönelik iş akış şeması

Başvurular alındıktan sonra başvuru verileri değerlendirmek üzere ilgili değerlendirme sistemlerine aktarılarak gerekli değerlendirmeler yapılmakta ve değerlendirme sonucu olumsuz olan başvurular için süreç sonlandırılmaktadır. Değerlendirme sonucu olumlu olan başvurular için ilgili veriler uygulama sistemlerine aktarılarak destek verilme aşaması başlamaktadır. Belirli aralıklarla destek süresi kontrol edilerek süresi devam eden hizmetlerinin verilmesine devam edilmektedir. Destek süresi kontrolü sonucunda süresi dolan hizmet sonlandırılarak süreç tamamlanmış olmaktadır.

Bir kurum tarafından sunulan bir hizmet sonucu oluşan verilere kurumun sunduğu farklı bir hizmet için ihtiyaç duyulabilmektedir. Şekil 29'da iki farklı hizmetin sunulması sırasında bir hizmete ait verinin farklı bir hizmet için kullanımını gösteren örnek şema verilmiştir.

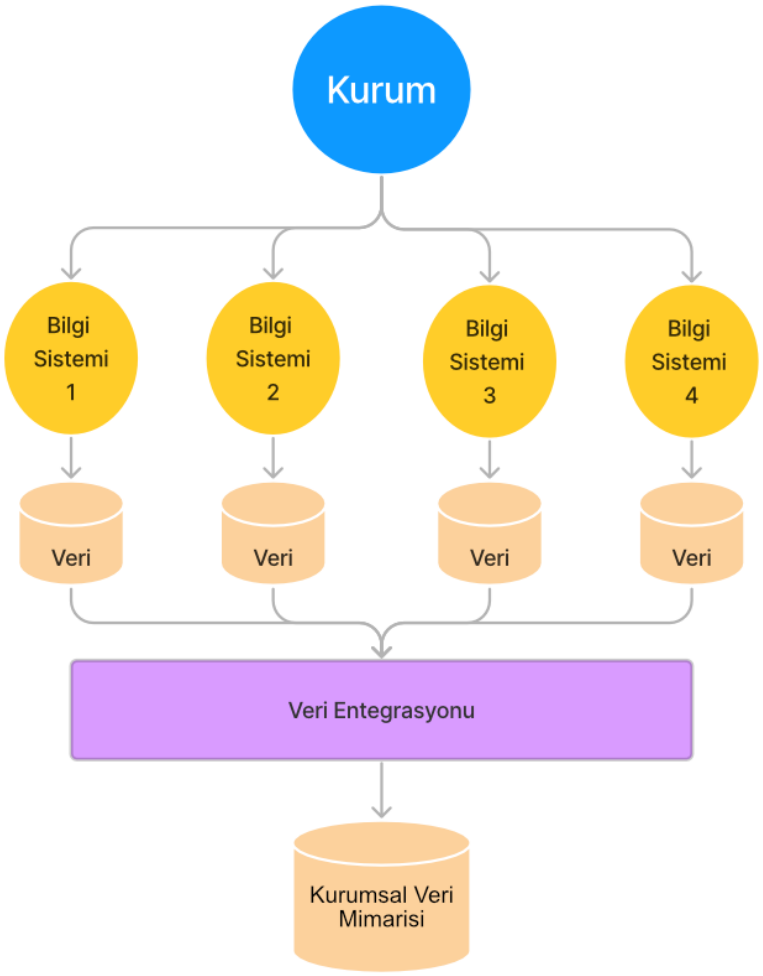


Şekil 29. Farklı hizmetlerin sunumları esnasında oluşan verilerin ortak kullanımını gösterir iş akış şeması

A programına ait hizmetin sunulması sonucunda elde edilen sonuç verisine B programına yapılan başvuruların değerlendirilmesi aşamasında ihtiyaç duyulabilir. B programına yapılan başvuru verilerine A programına ait ilgili sonuç verilerinin sistemsel olarak aktarılması B programının yürütülmesi sürecine doğrudan katkı sağlayacaktır. Örneğin B programına başvuruda bulunabilmek için

gerekli şartlardan birinin A programına yapılan başvurunun olumlu sonuçlanması olduğunu varsayalım. A programının sonuç verisi B programına başvuru esnasında kullanabilirse, B programına sadece A programı başvurusu olumlu sonuçlananların başvurabilmesi sağlanabilir. Bu veri B programına doğrudan aktarılmadığında ise B programına A programı başvurusu olumlu veya olumsuz sonuçlanan herkes başvuru yapabilecek ve B programına yapılan başvuruların değerlendirilmesi aşamasında A programına ait sonuç verisi kullanılarak A programı başvuru olumsuz sonuçlananların B programı başvuruları da olumsuz olarak sonuçlandırılacaktır. Verilerin A programından B programına doğrudan aktarılmaması hem süreçleri uzatacak hem de A programına ait verinin B programı için tekrar kaydedilmesine neden olarak veri tekrarına sebep olacaktır.

Bütün bu süreçlerde elde edilen veriler kurumsal veri mimarisinin (veri gölü, veri ambarı vb.) bileşenleri olmaktadır. Ancak bütün bu veriler farklı bilgi sistemlerinden elde edildiği için mimarinin içindeki modüller aracılığı ile veri ve verinin geçirdiği işlemler üst veri ile etiketlenmeli ve ilgili bağlam bilgileri oluşturulmalı, verinin yapılandırılma ve olgunluk durumuna göre ilgili depolama birimlerine gönderilmeli, verinin ilgili olduğu veri analitiği ve makine öğrenme katmanlarına uygun olacak şekilde veri hazırlanmalı ve yukarıda belirtildiği gibi verimsiz durumlardan ve veri yığını oluşumundan kaçınılmalıdır. Şekil 30’da farklı sistemlerden elde edilen verilerin kurumsal veri mimarisine entegre edilmesini özetleyen model verilmiştir.



Şekil 30. Farklı sistem verilerinin kurumsal veri mimarisine entegrasyonu

Daha önce belirtildiği gibi kurumsal veri mimarisi oluşturulurken farklı bilgi sistemlerinden gelen verilerin entegre edilerek sistemde tutulması gereklidir. Bu bağlamda, kurumsal veri mimarisi oluşturma süreci, yalnızca teknolojik altyapının entegrasyonu ile sınırlı kalmayıp, aynı zamanda kurumsal hedefler ve süreçlerle bilgi teknolojilerinin uyumlu hale getirilmesini gerektirmektedir.

1980'lerde giderek artan kurumsal karmaşıklığı ele almak için kurumsal mimari kavramı ortaya çıkmıştır. Bu süreçte bilgisayar yazılım geliştiricileri ancak kurumların süreçlerini, organizasyon yapısı ve hedeflerini anladıklarında, kuruma uygun bilgi teknolojilerini geliştirebileceklerini anlamışlardır. Kurumsal mimari, farklı kurumsal ekosistemlerin neden olduğu karmaşıklığı çözebilecek bir kavram olarak görülmektedir (Katuu, 2018, s. 2702). Modern kurumlar, kurumsal

görevlerini yerine getirecek teknolojileri geliştirmek ve iş uygulamaları oluşturmak için büyük yatırımlar yapmaktadır (Katuu, 2018, s. 2703). Kurumsal mimarinin yaygınlaşmasıyla bilgi teknolojilerinin hedeflerini işletmenin hedefleriyle uyumlu hale getirmesi amaçlamaktadır, ancak çok az kurum kurumsal mimariyi benimsemiştir ve bunların çok daha azı kapsamlı bir şekilde kullanmaktadır (Hazen, Kung, Cegielski ve Jones-Farmer, 2014).

5.2. TÜRKİYE'DE KURUMSAL VERİ SİSTEMLERİNE UYGUN VERİ GÖLÜ MİMARİLERİNİN GELİŞTİRİLMESİ

Veri gölleri, büyük veri yönetimi açısından stratejik bir çözüm sunmakla birlikte, veri göllerinin etkin kullanımı için doğru mimari yaklaşımlar, etkili üst veri yönetimi ve veri güvenliği politikalarının benimsenmesi gerekmektedir. Türkiye'deki kurumların mevcut veri yönetim sistemleri ve gelecekteki ihtiyaçları göz önünde bulundurularak ve uluslararası uygulamalardan yararlanılarak bu çalışmada kapsamlı bir uygulama çerçevesi geliştirilmiştir. Model, veri göllerinin Türkiye'deki kurumlara özgü gereksinimler doğrultusunda etkin bir şekilde uygulanmasını hedeflemektedir.

5.2.1. Fonksiyonel ve Olgunluk Tabanlı Melez Bir Mimari

Bu çalışmada kurumlar için fonksiyonel ve olgunluk tabanlı veri gölü mimarilerinin kombinasyonu temeline dayanan ayrıca gölet mimarisindeki gibi verilerin kategorilere ayrıldığı, kurumların mevcut veri yönetimi kapasitelerini ve gelecekteki hedeflerini dikkate alan kapsamlı bir melez mimari sunulmaktadır. Bu hibrit yaklaşım, veri göllerinin hem esnek hem de sürdürülebilir bir yapıya kavuşmasına yardımcı olurken, aynı zamanda kurumların veri yönetim süreçlerini aşamalı olarak geliştirmelerine olanak tanıyacaktır. Esneklik, farklı veri türlerini aynı platformda yönetebilme kapasitesini içermektedir. Böylece kurumların veri süreçlerinin değişen ihtiyaçlarına uyum sağlamasına katkıda bulunacaktır.

Fonksiyonel mimari, veri gölünün farklı işlevlerini (veri alımı, işleme, analiz gibi) ayrı ayrı ele alarak her bir aşamayı optimize etmeyi amaçlarken, olgunluk tabanlı mimari ise organizasyonun veri yönetimi kapasitesinin zaman içinde gelişimini göz önünde bulundurur. Sawadogo ve Darmont'un (2021) fonksiyonel ve olgunluk tabanlı mimari sınıflandırmasına göre melez kategoride yer alan Inmon'un gölet mimarisinde, ham veriler tür ve kaynaklarına göre çeşitli veri göletlerine ayrılmaktadır. Bu çalışmada önerilen melez mimaride ise, verilerin olgunlaşma seviyelerine göre üst katmanlarda yer alan veriler farklı veri göletlerinde toplanmaktadır. Bu bağlamda, farklı mimarilerde yer alan çeşitli yaklaşımların bileşenlerini bir araya getiren bir melez mimari model ile veri göllerinin daha etkin ve verimli bir biçimde kullanılması hedeflenmektedir. Ayrıca bu yaklaşım veri bataklığı riskini azaltacak ve veri yönetim sürecinin kalitesini artıracaktır. Bu süreçte verilerin ham halden olgunlaşmış bir veri kaynağına dönüşmesi kademeli bir sürece dayanmaktadır ve bu olgunlaşma süreci boyunca kurumlar veri yönetim stratejilerini adım adım geliştirme şansına sahiptir (Sawadogo ve Darmont, 2020).

Örneğin bir kurum için veri alımı ve temel analiz fonksiyonlarına odaklanan bir başlangıç seviyesi veri gölü tasarlanabilir. Zaman içinde, gölet mimarisinin önerdiği şekilde veriler kategorilerine göre ayrılarak veri gölünün veri işleme kapasitesi geliştirmek ve göle ileri analitik yetenekler ekleyerek mimariyi daha olgun bir seviyeye taşımak mümkün olacaktır. Bu aşamalı yaklaşım kurumların veri gölü teknolojisini kademeli olarak benimsemesine ve mevcut sistemlerle entegrasyonunu kolaylaştırmasına yardımcı olacaktır. Olgunlaşmış veri setleri, giderek daha karmaşık analizler için kullanılabilir hale gelecektir.

Fonksiyonel ve olgunluk tabanlı mimarilerin gölet mimarisi ile birleştirilmesi, Türkiye'deki veri göllerinin etkinliğini artırarak veri yönetimi süreçlerini daha sürdürülebilir ve verimli hale getirebilecektir. Bu mimari ile büyük verinin stratejik olarak kullanılmasını ve iş değerine dönüştürülmesini sağlayarak kurumların veri odaklı karar verme süreçlerini güçlendirmek amaçlanmaktadır.

5.2.2. Üst Veri Yönetimi

Üst veri yönetimi, veri göllerinin etkin bir şekilde kullanılabilmesi için kritik bir rol oynamaktadır. Bu çerçevede kurumlar, veri gölü teknolojilerine geçiş sürecinde büyük veri kümelerini yönetmek ve verilerin organizasyonunu şeffaf hale getirmek için etkili üst veri yönetimi stratejileri benimsemelidir.

Veri göllerinin etkin yönetimi için merkezi bir üst veri yönetim sistemi kurulumu kritik öneme sahiptir. Etkin bir üst veri yönetim sistemi, veri göllerindeki verilerin kaynağı, yapısı ve kullanım amacı hakkında kapsamlı bilgi sağlayarak verilerin kolayca bulunmasını ve erişilmesini mümkün kılar. Üst veri kataloglama, bu sistemin temel bir bileşenidir. Bu kataloglar, verilerin yapısı, kaynağı ve işlenme süreçleri hakkında detaylı bilgi sunarak veri göllerinin işlevselliğini ve verimliliğini artırır. Veri göllerinin dinamik yapısı dikkate alındığında, otomatik üst veri toplama ve güncelleme mekanizmalarının geliştirilmesi kaçınılmaz bir gereklilik olarak ortaya çıkmaktadır (Miloslavskaya ve Tolstoy, 2016). Bu mekanizmalar, yapay zekâ ve makine öğrenimi algoritmaları ile desteklenerek veri kaynaklarını düzenli olarak tarar ve üst veriyi sürekli olarak günceller. Örneğin, doğal dil işleme (NLP) teknikleri, yapılandırılmamış verilerin üst verilerini analiz ederek kategorilere ayırabilir. Ayrıca, bilgi grafikleri ve grafik tabanlı modeller, veri göllerindeki ilişkileri otomatik olarak haritalayarak üst veri entegrasyonunu kolaylaştırabilir. Kullanıcı deneyimini geliştirmek için üst veri sistemlerinin kullanıcı dostu hale getirilmesi önemlidir. Etkili arama ve filtreleme özellikleri sunan kullanıcı dostu arayüzler, veri keşfini kolaylaştırır ve analitik süreçleri hızlandırır (Halevy ve diğerleri, 2016).

Bir diğer önemli husus ise veri güvenliği ve uyumluluk konusunun üst veri yönetimiyle uyumlu hale getirilmesidir. Türkiye'deki veri güvenliği gereksinimleri doğrultusunda, üst veri yönetim sistemlerinin 6698 sayılı Kişisel Verilerin Korunması Kanunu'na (KVKK) uygun olarak hassas ve gizli verilerin korunmasını sağlaması kritik önemdedir. Uygulama alanına özgü veri türleri ve ilişkileri dikkate alınarak uygulamaya özgü üst veri standartları geliştirilmelidir. Son olarak, kurum içinde üst veri yönetimi konusunda eğitim ve farkındalık yaratılmalıdır. İlgili

personele düzenli eğitimler verilmeli ve üst veri yönetiminin önemi kurum genelinde vurgulanmalıdır. Böylece veri yönetimi süreçlerinin daha etkili hale gelmesine katkı sağlanacaktır.

5.2.3. Veri Yönetişimi ve Güvenliği

Veri gölleri, farklı veri türlerini, yapılandırılmamış veriler de dahil olmak üzere, büyük miktarlarda depolamak ve analiz süreçlerinde kullanmak amacıyla tasarlanmıştır. Ancak bu büyük ve karmaşık veri kümelerinin etkin yönetimi ve güvenliğinin sağlanması, veri gölleri ile çalışmanın en kritik unsurlarından biridir. Türkiye'deki kurumlar için veri göllerinin etkin ve güvenli bir şekilde yönetilmesi, veri yönetişimi ve veri güvenliği stratejilerinin bir arada düşünülmesini gerektirmektedir. Bu kapsamda, veri kalitesi, erişim kontrolü, şifreleme ve güvenlik politikalarının uyumlu bir şekilde uygulanması büyük önem taşımaktadır.

Veri yönetişimi, bir organizasyonun veri varlıklarının etkin bir şekilde yönetilmesi, kullanılması ve korunması için gerekli olan politikaların, prosedürlerin, süreçlerin ve sorumlulukların bütünüdür. Bu kapsamda, veri göllerinin başarısını artırmak için geniş bir yelpazede faaliyetler yürütülür. Bu faaliyetler arasında veri kalitesinin iyileştirilmesi, erişim kontrolünün sağlanması, uyumluluk yönetiminin geliştirilmesi, veri stratejisinin belirlenmesi, veri sahipliği ve sorumluluklarının tanımlanması, veri yaşam döngüsü yönetimi ve veri standartlarının oluşturulması yer alır. İlk olarak, veri kalitesini sağlamak için Türkiye'deki kurumların veri profilleme teknikleri kullanarak veri yapısını ve kalitesini anlamaları gerekir. Veri profilleme, veri setlerinin içeriğini, yapısını ve kalitesini analiz etmek için kullanılan bir süreçtir. Bu süreç, veri setlerindeki örüntüleri, eksik veya hatalı verileri, aykırı değerleri ve veri dağılımlarını belirlemeyi içerir. Veri profilleme sayesinde, kurumlar veri setlerinin genel özelliklerini anlayabilir ve potansiyel veri kalitesi sorunlarını erken aşamada tespit edebilirler (Sawadogo ve Darmont, 2021). Veri doğruluğunu artırmak amacıyla otomatik doğrulama kuralları ve veri temizleme süreçleri uygulanmalı, hatalı veriler düzeltilmeli ve veriler standartlaştırılmalıdır

(Terrizzano ve diğerleri, 2015). Bunlara ek olarak, veri erişiminde rol tabanlı erişim kontrolü ve veri maskeleme teknikleri kullanılarak, verilerin yetkisiz erişimlere karşı korunması sağlanabilir. Rol tabanlı erişim kontrolü, kullanıcılara organizasyon içindeki görevlerine veya pozisyonlarına göre belirli veri erişim yetkileri atanmasını içerir. Bu yaklaşım, kullanıcıların yalnızca görevleri için gerekli olan verilere erişmesini sağlayarak veri güvenliğini artırır (Halevy ve diğerleri, 2016; Suriarachchi ve Plale, 2016).

Veri güvenliği, veri göllerinde saklanan bilgilerin korunmasında bir diğer kritik bileşendir. Türkiye'de özellikle kişisel verilerin korunmasına yönelik düzenlemeler (KVKK) ve siber güvenlik tehditleri göz önüne alındığında, veri şifreleme ve anonimleştirme gibi stratejilerin kullanılması önemlidir. Kişisel verilerin güvenliği için AES ve RSA gibi güçlü şifreleme algoritmalarına başvurulmalı ve sağlık, finans gibi sektörlerde yasal uyumluluğun sağlanması için anonimleştirme teknikleri kullanılmalıdır (Karaarslan ve Akbaş, 2017). Ayrıca iki faktörlü kimlik doğrulama ve erişim loglarının izlenmesi gibi güvenlik önlemleri, gizli verilere yetkisiz erişimi önlemek açısından kritik önem taşır (Sawadogo ve Darmont, 2020).

Hem veri yönetişimi hem de veri güvenliği kapsamında Türkiye'deki kurumlar, ulusal ve uluslararası düzenlemelere uyum sağlayacak şekilde kapsamlı politikalar ve standartlar geliştirmelidir. Veri sınıflandırma stratejisi ile veriler hassasiyetlerine göre kategorize edilmeli ve her sınıf için uygun güvenlik önlemleri belirlenmelidir (Inmon, 2016). Bunun yanı sıra, verilerin yaşam döngüsü yönetimi politikaları ile, oluşturulmasından imha edilmesine kadar olan süreçlerin bütünsel bir şekilde yönetilmesi gerekir. ISO 27001 gibi uluslararası bilgi güvenliği standartlarına uyum sağlanarak, veri güvenliği politikaları daha sağlam bir çerçeveye oturtulabilir (Miloslavskaya ve Tolstoy, 2016). Bu stratejilerin uygulanması, kurumların veri göllerini daha güvenli ve etkin bir şekilde yönetmelerine olanak tanıyacaktır.

5.3. ÖNERİLEN VERİ GÖLÜ MİMARİSİ

Bu bölümde Türkiye'deki kurumların büyük veri yönetim süreçlerini etkin ve sürdürülebilir bir şekilde yönetmesi amacıyla geliştirilen veri gölü mimarisi sunulmuştur. Mimarinin temel dayanağı, Sawadogo ve Darmont (2021) tarafından ortaya konulan fonksiyonel ve olgunluk tabanlı veri gölü mimari sınıflandırmasıdır. Sawadogo ve Darmont (2021), hem fonksiyonel hem de olgunluk tabanlı ilkeleri bir araya getiren hibrit mimarilerin kurumların veri yönetim ihtiyaçlarına daha iyi yanıt verdiğini ve veri göllerinin avantajlarının kullanılmasında daha etkin olduklarını vurgulamışlardır. Bu çerçevede Türkiye'deki kamu kurumları referans alınarak fonksiyonel mimari, olgunluk tabanlı mimari ve veri göleti yaklaşımlarının özelliklerini bir araya getiren modüler hibrit bir mimari önerilmiştir. Bu çerçevede, mimaride yer alan tüm unsurlar, kurumların veri gölü altyapılarına aşamalı olarak geçiş yapmasını destekleyecek bir şekilde modüler ve esnek bir yapıda planlanmıştır. Mimarinin geliştirilmesinde, özellikle kademeli olgunlaşma süreci ve farklı olgunluk seviyelerinde veri göletlerinin oluşturulmasının, Türkiye'deki kamu kurumları için büyük bir kolaylık sağlayacağı öngörülmüştür. Bu sayede, kurumlar veri gölü altyapısına daha kolay geçiş yapabilir ve veri yönetim süreçlerini geliştirdikçe daha yüksek olgunluk seviyesindeki katmanları devreye alabilirler.

Mimarinin modüler yapısı, kurumların ihtiyaçlarına göre belirli modülleri devre dışı bırakmalarına veya ihtiyaç duyulduğunda yeniden entegre etmelerine olanak tanır. Örneğin, başlangıç aşamasında makine öğrenimi süreçlerine ihtiyaç duymayan bir kurum, bu modülü devre dışı bırakabilir ve ilerleyen dönemlerde olgun veri göletleri ile birlikte bu süreci sisteme dahil edebilir. Bu esneklik, kurumların kaynaklarını daha verimli kullanmalarını ve veri yönetim altyapılarını gelişen ihtiyaçlarına göre ölçeklendirmelerini sağlar.

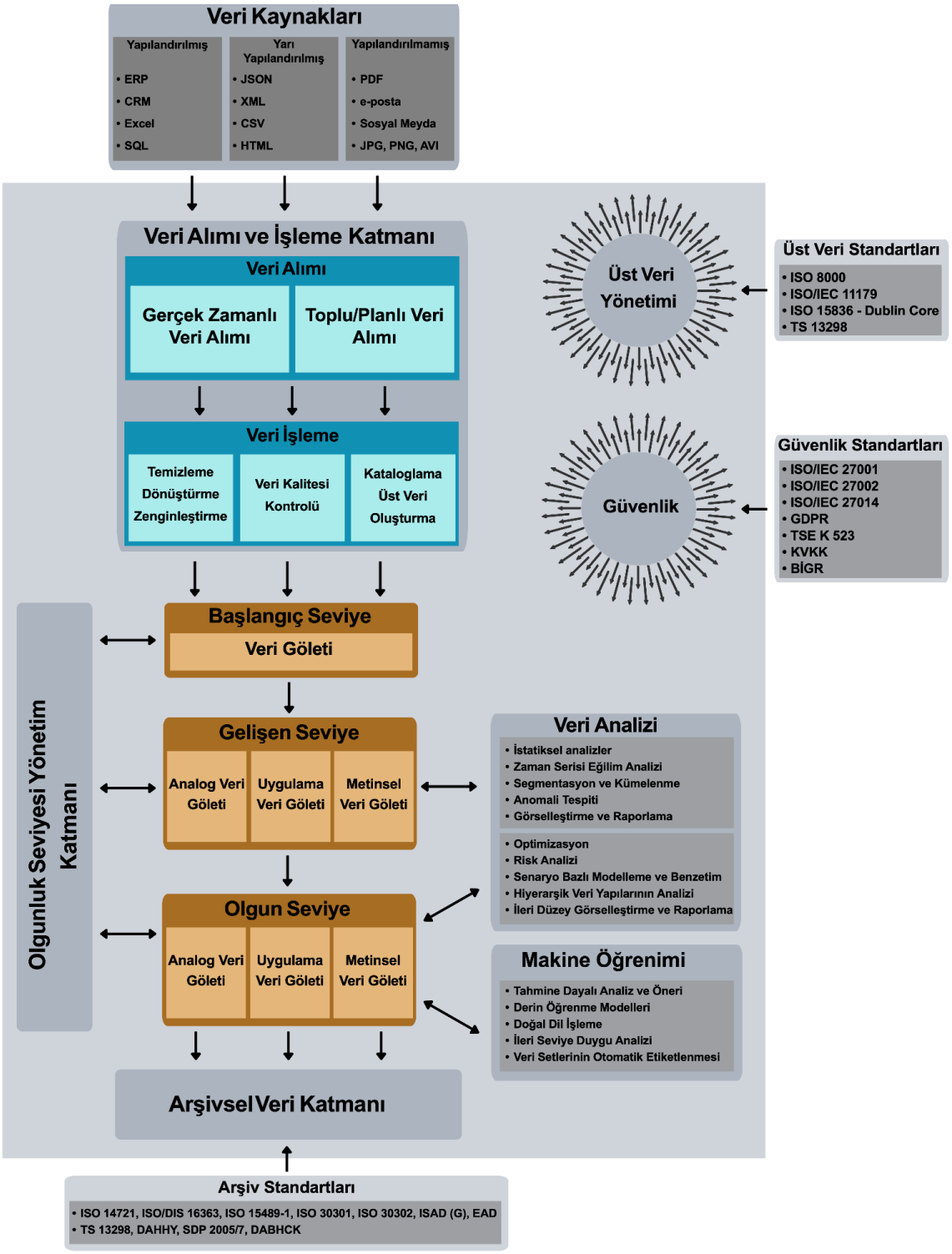
Önerilen veri gölü mimarisinin temel bileşenlerinden biri olan veri göletleri, verilerin türlerine ve olgunluk seviyelerine göre organize edilmesini sağlamaktadır. Önerilen modelde, veri göleti bölgeleri üç olgunluk seviyesinde (başlangıç, gelişen ve olgun) yapılandırılmıştır. Gelişen ve olgun veri göleti

bölgelerinde ise analog, uygulama ve metin veri göletleri bulunmakta, bu sayede veri yönetimi ve organizasyonu daha düzenli hale getirilmektedir. Ayrıca, bu yapı sayesinde kurumlar, yalnızca ihtiyaç duydukları veri göletlerini kullanarak kaynaklarını optimize edebilir ve operasyonel verimliliklerini artırabilirler.

Mimarinin bu şekilde yapılandırılması, Türkiye'deki kamu kurumlarının büyük veri yönetim süreçlerini etkin bir şekilde desteklemek ve sürdürülebilir kılmak amacı taşımaktadır. Bu yapı, dijital dönüşüm hedefleriyle uyumlu olarak, kurumların veri odaklı karar verme süreçlerini güçlendirme potansiyeline sahiptir.

5.3.1. Önerilen Veri Gölü Mimarisi: Ana Bileşenler ve Katmanların Entegrasyonu

Önerilen veri gölü mimarisi beş ana katmandan oluşmaktadır. Bu katmanlar, verilerin ham halinden işlenmiş bilgiye dönüştürülmesi sürecinde birbirleriyle uyumlu bir şekilde çalışmaktadır. Şekil 31'de önerilen veri gölü mimarisi modelini görsel olarak açıklayan bir blok şema verilmiştir. Bu şema, veri alımından işlenmiş bilgiye ulaşana kadar geçen sürecin tüm aşamalarını, katmanlar arasındaki etkileşimleri ve verilerin yönetimini kapsamaktadır. Veri alımından başlayarak, verilerin işlenme ve analiz süreçlerine kadar nasıl yönetileceğini adım adım göstermektedir. Veri alımı ile başlayan süreç, üst veri ve güvenlik yönetimi ile desteklenerek verilerin güvenli bir şekilde işlenmesi ve olgunlaşmasını, daha sonra ise arşivlenmesini kapsamaktadır. Olgunlaşan veriler ileri düzey analizler ve makine öğrenimi modelleri ile işlenerek iş zekâsı sistemlerine dâhil edilmektedir. Model, verinin yaşam döngüsünün tüm aşamalarını – veri oluşturma, işleme, saklama, analiz ve imha süreçlerini – dikkate alarak, bu süreçlerin kesintisiz bir şekilde yönetilmesini sağlamayı amaçlamaktadır.



Şekil 31. Önerilen veri gölü mimarisi modeli

Veri Alımı ve İşleme Katmanı: Bu katman, Bölüm 4.4.4.1'de tanımlanan fonksiyonel mimari ilkelerine dayanır. Farklı veri kaynaklarından gelen veriler bu katman aracılığıyla veri gölüne alınır. Kurumlar için bu süreç, özellikle büyük veri hacimlerini doğru bir şekilde depolamak ve işlemek açısından önemlidir. Veri alımı sırasında, üst veri bilgileri otomatik olarak oluşturulup veri gölüne eklenir (Sawadogo ve Darmont, 2021). Bu katman, üst veri yönetimi katmanı ile aktif etkileşim içinde olmalıdır. Veri alımı sırasında otomatik olarak üst veri bilgileri oluşturulur ve her veri kümesinin kaynağı, yapısı ve işlenme durumu tanımlanır. Bu işlem veri gölündeki verilerin organize edilmesini ve daha sonra erişilebilir hale gelmesini sağlar.

Olgunluk Seviyesi Yönetim Katmanı: Bölüm 4.4.4.2'de tanımlandığı üzere, veri gölü zamanla olgunlaşarak veri yönetimi süreçlerini geliştirir. Bu katman, ham verilerin işlenmiş ve anlamlandırılmış verilere dönüşmesini sağlayan süreçleri yönetir. Verilerin kalitesi artırıldıkça analiz süreçlerinde kullanılabilirlikleri artar (Miloslavskaya ve Tolstoy, 2016). Örneğin bu katmanda ham veri işlenerek iş zekâsı sistemlerine aktarılmak üzere olgunlaştırılır. Olgunluk Seviyesi Yönetim katmanı, veri gölünün zaman içinde daha gelişmiş analizler yapabilmesini sağlar ve kurumların veri odaklı karar verme süreçlerini güçlendirir (Sawadogo ve Darmont, 2021). Olgunluk Seviyesi Yönetimi Katmanı 3 aşamalıdır. İlk seviye "Başlangıç Seviye", ikinci seviye "Gelişen Seviye", üçüncü seviye ise "Olgun Seviye" olarak adlandırılmaktadır. Gelişen Seviye basamağında verinin kategorilere ayrılması mümkündür. Önerilen modelde bu aşamada verinin türüne göre 3 ayrı gölet oluşturulması önerilmektedir. Böylece veri türüne göre "2.4.1. Veri İşleme ve Metodolojiler" başlığında açıklanan araçlar kullanılarak veriyi analize hazır hale getirmek mümkün olacaktır. Gelişen seviyede gerçekleştirilen veri türüne göre ayrı gölet oluşturulması kavramı, Olgun seviye katmanına da taşınacaktır. Bu durumda, her veri türü kapsamında mimaride bir adet gelişen veri göleti ve bir adet olgun veri göleti yer alacaktır.

Üst Veri Yönetim Katmanı: Bölüm 5.2.2'de ele alındığı üzere, bu katman veri gölündeki verilerin kaynağını, yapısını ve işlenme süreçlerini organize eder. Üst

veri yönetim katmanı, verilerin izlenebilirliğini sağlayarak veri yönetişimi ve güvenlik süreçlerine dahil olur. Özellikle ulusal ve uluslararası düzenleyici gereksinimler doğrultusunda, üst veri yönetimi verilerin doğru bir şekilde izlenebilmesini ve organize edilmesini sağlar (Halevy ve diğerleri, 2016). Bu katman veri gölünün veri bataklığına dönüşmesini önlemek için en önemli adımları içerir ve veri keşfi süreçlerini hızlandırır (Suriarachchi ve Plale, 2016). Üst veri katmanının ürettiği üst veriler aynı zamanda hangi verilerin hangi kullanıcılar tarafından erişilebileceğini belirleyerek güvenlik katmanı ile etkileşim içinde hareket eder (Halevy ve diğerleri, 2016).

Veri Yönetişimi ve Güvenlik Katmanı: Bölüm 5.2.3'te açıklanan veri yönetişimi ve güvenlik ilkeleri bu katmanda uygulanır. Veri yönetişimi politikaları, verilerin kalitesini, güvenliğini ve kullanılabilirliğini denetler. Veri güvenliği katmanı ise şifreleme, anonimleştirme ve erişim kontrolü gibi güvenlik önlemleri sağlar. Kurumlar için veri güvenliği, yasal düzenlemelere uyum sağlamak açısından kritik bir bileşendir (Sawadogo ve Darmont, 2021). Bu katman, veri gölünün güvenilirliğini artırırken, aynı zamanda yasal uyumluluğu da garanti eder. Üst veri yönetimi katmanı ile bu katman arasında güçlü bir etkileşim olmalıdır. Üst veri, veri güvenliği için gerekli şifreleme ve anonimleştirme bilgilerini sağlar. Hassas ve gizli veriler, üst veri katmanında tanımlandıktan sonra bu katmanda işlenir ve korunur.

Arşivsel Veri Katmanı: Önerilen veri gölü mimarisinin son katmanı arşivsel veri katmanıdır. Olgun seviyede depolanan verilerden aktif olarak analiz edilmesine ihtiyaç duyulmayan ve kullanım ihtimali azalmış veriler arşivsel veri katmanına yönlendirilir. Bu verilerden gelecekte analiz edilebilecek veya kullanılma ihtimali olanlar ile yasal gereklilikler nedeniyle saklanacak olanlar kurumsal veri arşivinde depolanır. Arşivlenme sürecine alınmayan veriler imha edilerek veri gölünden çıkarılır.

Bu önerilen yapı ile Türkiye'deki kurumlar, büyük veri yönetiminde etkinliklerini artırabilirler. Katmanlar arası entegrasyon sayesinde veriler daha düzenli bir

şekilde yönetilirken, güvenlik ve uyumluluk da sağlanmış olur. Bu yapı kurumların değişen ihtiyaçlarına göre esnek bir şekilde uyarlanabilir ve ölçeklenebilir bir çözüm sunar.

5.3.2. Önerilen Mimari ile İlgili Standartlar

5.3.2.1. Önerilen Mimari ile İlgili Ulusal Standartlar

Aşağıda 3. Bölümde açıklanan üst veri yönetimi, güvenlik, gizlilik, arşivleme ve belgeleme süreçleriyle ilişkili ulusal standartlar önerilen mimariyle ilişkilerine göre gruplandırılmış olarak sunulmuştur:

Üst Veri ve Veri Yönetimi Standartları:

- Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (9/6/2020 Tarih ve 2646 Sayılı)
- TS 13298 Elektronik Belge Yönetimi Standardı
- E-Yazışma Teknik Rehberi (E-YTR)

Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (2020) ve TS 13298 Elektronik Belge Yönetimi Standardı, önerilen veri mimarisinin Üst Veri Yönetim Katmanı ile doğrudan ilişkilidir. Bu standartlar verilerin kaynağı, yapısı ve işlenme süreçlerinin organize edilmesine rehberlik ederek veri gölündeki üst veri yönetimi stratejilerinin temelini oluşturmaktadır.

Özellikle TS 13298 üst verilerle veri sınıflandırmasını ve organizasyonunu desteklerken, Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (2020) resmî belgelerde kullanılan üst veri standartlarının belirlenmesine ve bu verilerin doğru tanımlanarak etkin bir şekilde yönetilmesine rehberlik etmektedir. Bu standartlar, Üst Veri Yönetim Katmanının veri güvenliği ve yasal uyumluluk süreçleriyle entegre bir şekilde çalışmasını mümkün kılmaktadır.

E-Yazışma Teknik Rehberi, kamu kurumları arasında gerçekleştirilen elektronik yazışmalarda üst verilerin güvenli, standartlara uygun ve uyumlu bir şekilde işlenmesi için bir yol haritası sunmaktadır. Bu rehber, üst veri yönetiminde kullanılan belge formatlarının ve güvenlik protokollerinin belirlenmesini sağlayarak veri güvenliği ve yasal uyumluluk süreçlerine katkı sunmaktadır.

Bilgi Güvenliği ve Gizlilik Standartları:

- TSE K 523: Bilgi Varlıklarının Gizlilik Derecelerine Göre Sınıflandırılması Kriteri
- Kişisel Verilerin Korunması Kanunu (KVKK) (No: 6698)
- Bilgi ve İletişim Güvenliği Rehberi (BİGR)
- Gizlilik Dereceli Belgelerde Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (GDB-Y)
- E-Yazışma Teknik Rehberi (E-YTR)

Veri Yönetişimi ve Güvenlik Katmanı, önerilen mimaride verilerin sınıflandırılmasını, korunmasını ve yasal düzenlemelere uyumunun sağlanmasını mümkün kılmaktadır. Bu katman, TSE K 523 ve GDB-Y rehberliğinde gizlilik derecelerine göre verilerin sınıflandırılmasını ve korunmasını sağlamaktadır. KVKK, kişisel verilerin güvenliği için şifreleme ve anonimleştirme gibi teknik önlemleri zorunlu kılmakta, Bilgi ve İletişim Güvenliği Rehberi (BİGR) ise erişim kayıtlarının izlenmesi ve iki faktörlü kimlik doğrulama gibi güvenlik uygulamaları için kapsamlı rehberlik sunmaktadır. Bu standartlar, Veri Yönetişimi ve Güvenlik Katmanının, üst veri yönetimiyle entegre bir şekilde çalışarak verilerin güvenliğini ve yasal uyumluluğunu garanti altına almasını sağlamaktadır. E-Yazışma Teknik Rehberi ise elektronik yazışmalarda veri bütünlüğü, erişim kontrolü ve güvenlik mekanizmalarının uygulanması için standartlar sağlayarak bu süreçleri desteklemektedir.

Arşiv ve Belge Yönetimi Standartları:

- Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik (9/6/2020 Tarih ve 2646 Sayılı)
- TS 13298 Elektronik Belge Yönetimi Standardı
- Devlet Arşiv Hizmetleri Hakkında Yönetmelik (DAHHY)
- Standart Dosya Planı ile İlgili 2005/7 Sayılı Başbakanlık Genelgesi (SDP 2005/7)
- Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi (Kararname Numarası: 11) (DABHCK)

Arşivsel Veri Katmanı, sunulan veri gölü mimarisinin temel bileşenlerinden biri olarak, yasal gereklilikler ve gelecekteki ihtiyaçlar doğrultusunda verilerin düzenli ve ulusal standartlarla uyumlu şekilde yönetilmesini sağlamaktadır. TS 13298 Elektronik Belge Yönetimi Standardı, belgelerin üst verilerle düzenlenmesini ve saklama sürelerinin belirlenmesini kolaylaştırırken, Resmi Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik, bu belgelerin resmi süreçlere uygunluğunu temin etmektedir. Standart Dosya Planı (SDP 2005/7) arşiv düzenini ve erişilebilirliğini desteklerken, Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi (DABHCK), belge saklama ve imha süreçlerinin ulusal politikalara uygun olarak yürütülmesini sağlamaktadır. Tüm bu standart ve düzenlemeler, Arşivsel Veri Katmanının uyumlu, güvenli ve işlevsel bir yapı sunmasına olanak tanımaktadır.

5.3.2.1. Önerilen Mimari ile İlgili Uluslararası Standartlar

Aşağıda 3. Bölüm'de açıklanan üst veri yönetimi, güvenlik, gizlilik, arşivleme ve kayıt süreçleriyle ilişkili uluslararası standartlar önerilen mimariyle ilişkilerine göre gruplandırılmış olarak sunulmuştur:

Üst Veri ve Veri Yönetimi Standartları:

- ISO 8000: Veri Kalitesi
- ISO/IEC 11179: Üst Veri Kayıtları
- ISO 15836: Dublin Core Üst Veri Eleman Seti

ISO 8000, ISO/IEC 11179 ve ISO 15836 gibi uluslararası standartlar, önerilen mimarinin Üst Veri Yönetim Katmanı ile güçlü bir bağlantı içerisindedir. Bu standartlar verilerin düzenlenmesi, tanımlanması ve yönetilmesine yönelik küresel kabul görmüş yöntemler sunmaktadır.

ISO 8000, veri kalitesini sağlamak için doğruluk, bütünlük ve kullanılabilirlik gibi kriterler belirlemektedir. Bu kriterler, mimaride üst veri yönetim sistemleri aracılığıyla veri bütünlüğünün korunmasına ve kullanıcıların ihtiyaçlarına uygun şekilde veri erişiminin sağlanmasına olanak tanımaktadır.

ISO/IEC 11179, üst veri kayıtları için standart bir yapı sunarak, mimaride verilerin tanımlanmasını ve organize edilmesini desteklemektedir. Bu yapı, sistemler arasında birlikte çalışabilirliği artırırken, verilerin hızlı bir şekilde keşfedilmesini ve etkin bir şekilde yönetilmesini mümkün kılmaktadır.

ISO 15836 (Dublin Core), dijital kaynakların açıklanması ve sınıflandırılması için kullanılan üst veri elemanlarını tanımlamaktadır. Bu standart, veri kaynaklarının standartlaştırılmış bir biçimde tanımlanmasını sağlayarak mimaride erişilebilirlik ve yeniden kullanılabilirliği artırmaktadır.

Bu standartlar, önerilen mimarinin uyumlu, kaliteli ve sürdürülebilir bir yapı sunmasını sağlamaktadır. Böylece, mimarideki üst veri yönetim katmanı, uluslararası kabul görmüş yöntemlerle entegre bir şekilde çalışarak, veri yönetiminin genel etkinliğini arttırmaktadır.

Bilgi Güvenliği ve Gizlilik Standartları:

- ISO/IEC 27001: Bilgi Güvenliği Yönetim Sistemleri - Gereksinimler
- ISO/IEC 27002: Bilgi Güvenliği Kontrolleri

- ISO/IEC 27014: Bilgi Güvenliği Yönetimi
- Avrupa Birliği GDPR (2018)
- ABD HIPAA: Sağlık Sigortası Taşınabilirliği ve Hesap Verebilirlik Yasası

ISO/IEC 27001, ISO/IEC 27002, ISO/IEC 27014 ve GDPR gibi uluslararası standartlar, önerilen mimarinin Veri Yönetişimi ve Güvenlik Katmanı için güçlü bir temel sağlamaktadır. Bu standartlar bilgi güvenliği ve gizliliğin sağlanmasında hem teknik hem de yönetsel süreçleri entegre eden kapsamlı bir çerçeve sunmaktadır.

ISO/IEC 27001 bilgi güvenliği yönetim sistemlerine yönelik gereksinimleri belirlerken ISO/IEC 27002 bu gereksinimlerin uygulanmasına yönelik ayrıntılı kontroller sağlamaktadır. ISO/IEC 27014 bilgi güvenliği hedeflerini kurumsal stratejilerle uyumlu hale getiren bir yönetişim çerçevesi sunarak mimarinin güvenlik süreçlerini daha sağlam bir temele oturtmaktadır. Bu standartlar önerilen mimaride veri şifreleme, anonimleştirme ve erişim kontrolü gibi güvenlik önlemlerinin uygulanmasına doğrudan rehberlik etmektedir.

GDPR, kişisel verilerin korunmasını küresel bir öncelik haline getirerek Türkiye'deki veri yönetim süreçlerini de güçlü bir şekilde etkilemektedir. GDPR'nin getirdiği erişim kontrolü, veri sınıflandırması ve ihlal yönetimi gibi uygulamalar, önerilen mimarinin güvenlik katmanında doğrudan karşılık bulmaktadır.

HIPAA ise sağlık sektörü odaklı bir düzenleme olarak sağlık ve benzeri sektörlerde hassas veri yönetimi için yararlanılabilir bir rehberdir.

ISO standartları ve GDPR, önerilen mimarinin güvenlik yapısını şekillendiren ana çerçeveyi oluşturmakta ve teknik ile yasal gerekliliklerin birlikte ele alınmasını sağlamaktadır. Böylece, mimarinin güvenli, uyumlu ve sürdürülebilir bir yapı sunması mümkün olmaktadır.

Bilgi güvenliği ve gizlilik standartları, bilgi varlıklarının korunması, gizliliğin sağlanması ve yasal düzenlemelere uyum için kapsamlı rehberlik sağlamaktadır.

Hem teknik hem de yönetişim düzeyinde kontrol mekanizmaları sunmakta ve bilgi güvenliğini kurumsal bir öncelik haline getirmektedir.

Bilgi Teknolojileri Yönetimi ve Yapay Zekâ Standartları:

- COBIT: Bilgi ve İlgili Teknolojiler İçin Kontrol Hedefleri
- ITIL: Bilgi Teknolojisi Altyapı Kütüphanesi
- ISO/IEC JTC 1/SC 42: Yapay Zekâ
- ISO/IEC 42001: Yapay Zekâ - Yönetim Sistemi

COBIT, ITIL, ISO/IEC JTC 1/SC 42 ve ISO/IEC 42001 gibi uluslararası standart ve çerçeveler önerilen melez veri mimarisinin farklı katmanlarına rehberlik ederek stratejik yönetişim, operasyonel verimlilik ve yapay zekâ süreçlerinin güvenilir bir şekilde uygulanmasını sağlamaktadır.

Olgunluk Seviyesi Yönetim Katmanı açısından, COBIT ve ITIL'in BT yönetimi için sunduğu çerçeve, veri gölünün olgunlaşma sürecine önemli katkılar sunmaktadır. COBIT, BT süreçlerini kurumsal hedeflerle uyumlu hale getirerek, veri yönetim süreçlerinin stratejik bir perspektiften ele alınmasını mümkün kılmaktadır. ITIL ise operasyonel süreçlerin etkin yönetimi için rehberlik ederek ham verilerin işlenip anlamlı hale gelmesini desteklemektedir. Bu çerçeveler, veri gölünün başlangıç seviyesinden olgun bir yapıya dönüşmesine olanak tanırken organizasyonların veri odaklı karar alma süreçlerini daha etkin hale getirmektedir.

Veri Alımı ve İşleme Katmanı, ITIL'in önerdiği hizmet yönetimi süreçleri ile optimize edilmektedir. Farklı veri kaynaklarından alınan verilerin işlenmesi sırasında üst veri entegrasyonuna yönelik süreçlerin düzenlenmesi, ITIL'in operasyonel verimliliği artırıcı rehberliği ile uyumlu bir şekilde gerçekleştirilmelidir.

Veri Analizi ve Makine Öğrenimi Katmanları, ISO/IEC JTC 1/SC 42 ve ISO/IEC 42001 standartlarının rehberliğinde, yapay zekâ süreçlerinin teknik ve etik gerekliliklerle uyumlu hale getirilmesine olanak tanımaktadır. ISO/IEC JTC 1/SC 42, yapay zekâ uygulamalarının teknik standartlarını belirleyerek, ileri analitik tekniklerin güvenilir bir şekilde kullanılmasını sağlamaktadır. Öte yandan ISO/IEC

42001, yapay zekâ uygulamalarının kurumsal stratejilere entegre edilmesi için bir yönetim çerçevesi sunmaktadır.

COBIT ve ITIL, BT yönetimi için stratejik ve operasyonel rehberlik sağlarken ISO/IEC yapay zekâ standartları teknik ve etik bir çerçeve sunmaktadır. Bu iki yaklaşım, önerilen veri mimarisinin olgunluk seviyelerine ulaşmasını ve ileri analitik süreçlerle uyumlu bir şekilde çalışmasını desteklemektedir. Bu bütünleşik yaklaşım, kurumların veri yönetimi kapasitelerini artırırken yapay zekâ ve veri analitiği süreçlerini daha güvenilir, verimli ve stratejik bir hale getirmeyi amaçlamaktadır.

Arşiv ve Belge Yönetimi Standartları:

- ISO 14721: OAIS - Açık Arşivsel Bilgi Sistemi
- ISO/DIS 16363: Güvenilir Dijital Depoların Denetimi ve Sertifikasyonu
- ISO 15489-1: Belge Yönetimi
- ISO 30301: Belge Yönetim Sistemleri - Gereksinimler
- ISO 30302: Belge Yönetim Sistemleri - Uygulama Rehberi
- ISAD (G): Genel Uluslararası Standart Arşivsel Tanım
- EAD: Kodlanmış Arşivsel Tanımlama

ISO 14721, Arşivsel Veri Katmanında depolanan verilerin uzun vadeli saklama ve erişim süreçleri için temel bir çerçeve sunarken ISO/DIS 16363, dijital depoların güvenilirliğini garanti altına alacak kriterleri sağlamaktadır. ISO 15489-1, belgelerin yaşam döngüsüne odaklanarak arşivde düzenli ve yapılandırılmış bir veri yönetimi sağlamaktadır. Buna ek olarak, ISO 30301 ve ISO 30302, belge yönetim sistemlerinin etkili şekilde kurulması ve işletilmesine yönelik ayrıntılı rehberlik sunmaktadır. Dijital belgelerin standart hale getirilmesi için ISAD (G) ve EAD, arşiv katmanında verilerin erişilebilirliği ve düzenlenmesi süreçlerinde kullanılmaktadır. Tüm bu standartlar, Arşivsel Veri Katmanında verilerin düzenli ve sürdürülebilir şekilde yönetilmesini mümkün kılmaktadır.

5.3.3. Önerilen Mimarinin Avantajları

Sunulan veri mimarisi modelinin Türkiye'deki kurumlar için sunacağı avantajlar aşağıda açıklanmıştır:

Esneklik ve Ölçeklenebilirlik: Farklı veri kaynaklarından gelen verilerin entegre edilmesi ve işlenmesi, kurumların ihtiyaçlarına göre esnek ve ölçeklenebilir bir yapı sağlamaktadır. Bu, büyük veri setlerini yönetmek zorunda olan kurumlar için önemlidir. Veri gölü mimarilerinin esnekliği, özellikle hızla değişen veri ekosistemleri için kritiktir (Giebler ve diğerleri, 2021).

Güvenlik ve Uyumluluk: Güvenlik katmanı özellikle hassas verilerin korunmasını sağlamaktadır. Türkiye'de veri güvenliği ve uyumluluk gereksinimleri, özellikle 7 Nisan 2016 tarihli ve 29677 sayılı Resmî Gazete'de yayımlanan 6698 sayılı Kişisel Verilerin Korunması Kanunu'nun (KVKK) yürürlüğe girmesi ile birlikte, kuruluşların veri yönetimi stratejilerinde önemli yer tutmaya başlamıştır ("Kişisel Verileri Koruma Kurumu", 2023; Resmî Gazete, 2016).

Kademeli Olgunlaşma Süreci: Veri yönetimi süreçleri kademeli olarak gelişmektedir. Kurumlar, başlangıç aşamasında ham veri ile çalışmaya başlarken süreç ilerledikçe olgunlaşmış verileri iş zekâsı ve makine öğrenimi süreçlerine dahil etmektedir. Olgunlaşmış veri, ham verinin temizlenmiş, anlamlandırılmış ve iş süreçlerinde kullanılabilir hale getirilmiş halidir. Bu süreç, veri temizleme, dönüşüm, entegrasyon ve zenginleştirme gibi adımları içerir. Örneğin, bir sağlık kuruluşu başlangıçta yalnızca ham hasta kayıtlarını saklarken, süreç ilerledikçe bu veriler temizlenerek eksik bilgiler tamamlanır, farklı kaynaklardan gelen tıbbi geçmişler entegre edilir ve hastalık gruplarına göre sınıflandırmalar eklenir. Bu olgunlaştırılmış veri seti, daha ileri düzey analizlerde kullanılabilir hale gelir. Sawadogo ve Darmont (2021), veri gölü olgunluk modellerinin, organizasyonların veri yönetimi kapasitelerini aşamalı olarak geliştirmelerine olanak tanıdığını vurgulamıştır.

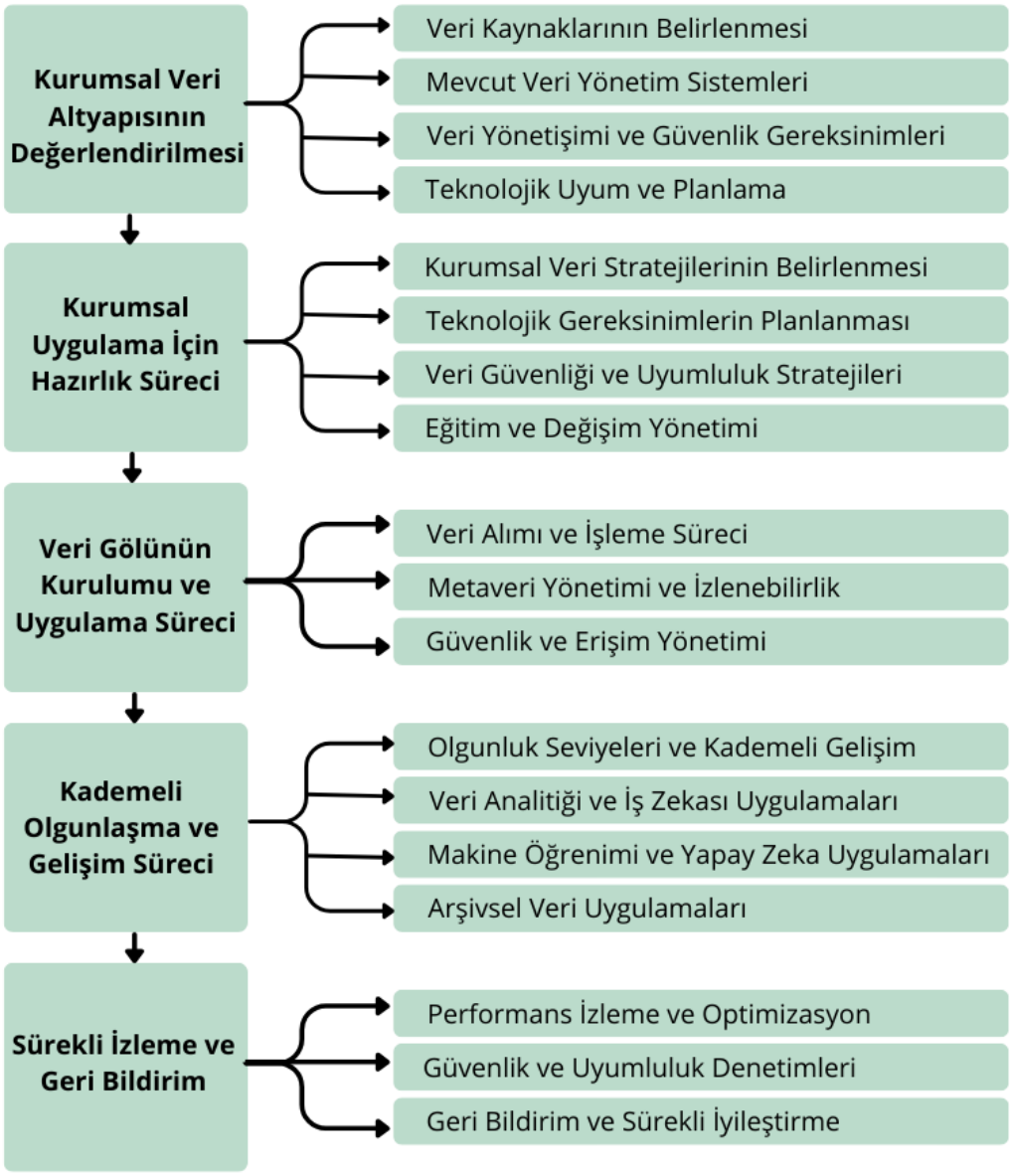
Üst Veri ile İzlenebilirlik: Üst veri yönetim katmanı sayesinde veri gölündeki verilerin izlenebilirliği ve düzenlenmesi sağlanmaktadır. Bu süreç, özellikle Türkiye'deki düzenleyici kurumlar için verilerin şeffaf ve izlenebilir bir yapıda olmasını sağlayacaktır. Halevy ve diğerleri (2016), etkili üst veri yönetiminin veri keşfi ve analizi süreçlerini önemli ölçüde iyileştirdiğini göstermiştir.

Önerilen veri gölü mimarisi, Türkiye'deki kurumların büyük veri yönetimi süreçlerini desteklemek ve sürdürülebilir kılmak için kapsamlı bir yapı sunmaktadır. Bu mimari, fonksiyonel ve olgunluk tabanlı veri gölü sınıflandırmasını temel alırken üst veri yönetimi, veri yönetişimi ve güvenlik katmanlarını da içeren bütünleşik bir sistem oluşturarak esnek ve güvenli bir veri yönetimi modeli sunmaktadır. Bu model, verilerin kademeli olarak olgunlaştırılmasını, güvenlik ve yönetişim ilkeleri doğrultusunda kullanılabilir hale getirilmesini sağlamayı amaçlamaktadır. Önerilen mimari, Türkiye'deki kurumların veri odaklı karar verme süreçlerini güçlendirme potansiyeline sahiptir. Ayvaz ve Salman (2020), Türkiye'deki firmaların büyük veri analitiği kullanım olgunluğunu inceledikleri çalışmada, firmaların bu alanda ilerleme kaydettiğini, ancak bütünleşik veri yönetimi çözümlerine olan ihtiyacın devam ettiğini vurgulamıştır. Önerilen model, bu ihtiyaca cevap vererek kurumların veri varlıklarından maksimum değer elde etmelerine olanak tanıyacaktır. Ayrıca bu mimari model, Türkiye'nin dijital dönüşüm hedefleriyle de uyumludur. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi (DDO), Türkiye'nin dijital dönüşüm girişimlerinin başarısının etkili veri yönetimi stratejilerine bağlı olduğunu vurgulayarak Ulusal Veri Stratejisi'nin hazırlık çalışmalarını başlatmıştır (“Cumhurbaşkanlığı Dijital Dönüşüm Ofisi”, 2024). Önerilen veri gölü mimarisi, bu bağlamda Türkiye'deki kurumların dijital olgunluklarını artırmalarına ve veri odaklı inovasyonu teşvik etmelerine yardımcı olabilecek potansiyele sahiptir.

5.4. VERİ GÖLLERİNİN KURUMSAL UYGULAMA SÜRECİ İÇİN MODEL ÖNERİSİ

Veri göllerinin kurumsal yapılara entegrasyonu, büyük veri yönetimi süreçlerini optimize etmek ve kurum ihtiyaçlarına uygun veri stratejileri geliştirmek açısından kritik bir süreçtir. Kurumların veri gölü entegrasyon sürecinde dikkate alması gereken önemli unsurlar bulunmaktadır. Bu bölümde, Türkiye'deki kamu kurumlarının önerilen veri gölü mimarisi modelini başarılı bir şekilde uygulayabilmeleri için atması gereken stratejik adımları ve kritik unsurları ortaya koyan bir model sunulmuştur. Bu uygulama modeli, mevcut kurumsal veri yönetim altyapısının değerlendirilmesi, veri gölünün entegrasyonu için hazırlık yapılması, veri gölünün kurulması, kademeli olgunlaşma süreci ve sürekli izleme gibi unsurları kapsamaktadır.

Şekil 32'de önerilen veri gölü uygulama sürecini açıklayan model yer almaktadır. Bu şema veri kaynaklarından başlayarak işlenmiş bilgilere ulaşılana kadar olan süreçleri göstermektedir. Bu model içindeki her bir aşama kurumların veri yönetimi ve arşivleme süreçlerini etkili bir şekilde yapılandırması ve elde ettikleri veriden en yüksek seviyede yararlanması için kritik öneme sahiptir. Tablo 4'te önerilen model oluşturulurken temel alınan ve model uygulanırken dikkate alınması gereken ulusal ve uluslararası standartlar her bir katman ve alt katman için listelenmiştir. Bütün bu adımların izlenmesi ile, kurumun stratejik hedeflerini ve yasal gereklilikleri karşılayan bir veri mimarisi ile verilerin etkin bir şekilde yönetilmesi, güvenli bir şekilde saklanması ve uzun vadeli olarak korunmasına katkı sağlamak hedeflenmektedir.



Şekil 32. Önerilen veri gölü mimarisinin kurumsal uygulama modeli

Tablo 4. Önerilen modelin katmanlara göre ilgili olduğu mevzuat ve standartlar[1]

Model Katmanı	Model Alt Katmanı	İlgili Mevzuat ve Standartlar
Kurumsal Veri Altyapısının Değerlendirilmesi	Veri Kaynaklarının Belirlenmesi	ISO/IEC 11179, ISO 15836, ISO 8000, TS ISO/IEC 27001, TSE K 523*, KVKK*, TS 13298*
Kurumsal Veri Altyapısının Değerlendirilmesi	Mevcut Veri Yönetim Sistemleri	ISO/IEC TR 10032, ISO/IEC 19583, ISO/IEC 38500, COBIT, ISO/IEC 20000, ITIL, TS 13298*, SDP 2005/7*, DABHCK*, DABHCK*
Kurumsal Veri Altyapısının Değerlendirilmesi	Veri Yönetimi ve Güvenlik Gereksinimleri	TS ISO/IEC 27001, TS ISO/IEC 27002, ISO/IEC 27014, ISO/IEC 29100, ISO/IEC 27701, Avrupa Birliği GDPR, ABD FISMA, ABD HIPAA, COBIT, TSE K 523*, KVKK*, BİGR*, GDB-Y*, E-YTR*
Kurumsal Veri Altyapısının Değerlendirilmesi	Teknolojik Uyum ve Planlama	ISO/IEC 19763, ISO/IEC 19510, ITIL, ISO/IEC 20000, ISO/IEC 42010, ISO/IEC 38500
Kurumsal Uygulama için Hazırlık Süreci	Kurumsal Veri Stratejilerinin Belirlenmesi	ISO/IEC 38500, ISO 30300, ISO 30301, ISO 30302, ISO 14721, COBIT, DABHCK*
Kurumsal Uygulama için Hazırlık Süreci	Teknolojik Gereksinimlerin Planlanması	ISO/IEC 19510, ISO/IEC TR 10032, ITIL, ISO/IEC/IEEE 42010
Kurumsal Uygulama için Hazırlık Süreci	Veri Güvenliği ve Uyumluluk Stratejileri	TS ISO/IEC 27001, ISO/IEC 27002, ISO/IEC 27701, Avrupa Birliği GDPR, ABD HIPAA, BİGR*, GDB-Y*, E-YTR*
Kurumsal Uygulama için Hazırlık Süreci	Eğitim ve Değişim Yönetimi	ITIL, ISO/IEC 38500, PROSCI Change Management
Veri Gölünün Kurulumu ve Uygulama Süreci	Veri Alımı ve İşleme Süreci	ISO/IEC 19583, ISO 8000, ISO/IEC 11179, ISO/IEC 29100, GDPR, SDP 2005/7*, VPKY*
Veri Gölünün Kurulumu ve Uygulama Süreci	Üst Veri Yönetimi ve İzlenebilirlik	PREMIS, ISO 15836, ISO 19115, ISO 23081-1, Dublin Core, VPKY*, E-YTR*
Veri Gölünün Kurulumu ve Uygulama Süreci	Güvenlik ve Erişim Yönetimi	ISO/IEC 27001, ISO/IEC 27002, ISO/IEC 27701, Avrupa Birliği GDPR, ABD FISMA, ABD HIPAA
Kademeli Olgunlaşma ve Gelişim Süreci	Olgunluk Seviyeleri ve Kademeli Gelişim	COBIT, ISO/IEC 19763, ITIL
Kademeli Olgunlaşma ve Gelişim Süreci	Veri Analitiği ve İş Zekâsı Uygulamaları	ISO/IEC 19583, ISO/IEC 19510, Colibra, DMBOK, DGI Standards
Kademeli Olgunlaşma ve Gelişim Süreci	Makine Öğrenimi ve Yapay Zekâ Uygulamaları	ISO/IEC 38500, ISO/IEC JTC 1/SC 42, ISO/IEC 42001

[1] Tablo içerisinde kullanılan kısaltmalar, “3.2. İLGİLİ MEVZUAT VE UYGULAMALAR” bölümünde ilgili başlıklarda parantez içinde verilmiştir.

Kademeli Olgunlaşma ve Gelişim Süreci	Veri Yönetimi ve Sürekli İyileştirme	ISO/IEC TR 10032, ITIL, COBIT, DMBOK, Deming Döngüsü, DAHHY*, DAGM DAY*, DABHCK*, SDP 2005/7*, VPKY*, DABHCK*
Sürekli İzleme ve Geri Bildirim	Performans İzleme ve Optimizasyon	ISO/IEC 38500, COBIT, ITIL, ISO/IEC 33000
Sürekli İzleme ve Geri Bildirim	Güvenlik ve Uyumluluk Denetimleri	ISO/IEC 27014, TS ISO/IEC 27001, ISO/IEC 27005, AB GDPR, ABD HIPAA, ABD FISMA, VPKY*, BİGR*, GDB-Y*, E-YTR*
Sürekli İzleme ve Geri Bildirim	Geri Bildirim ve Sürekli İyileştirme	ITIL, ISO/IEC 19510, ISO 8000, Deming Döngüsü

**Ulusal*

5.4.1. Kurumsal Veri Altyapısının Değerlendirilmesi

Veri göllerinin başarılı bir şekilde uygulanabilmesi için ilk adım, mevcut kurumsal veri altyapısının detaylı bir değerlendirmesini yapmaktır. Mevcut veri yönetim sistemlerinin esneklik, ölçeklenebilirlik, güvenlik ve entegrasyon açısından değerlendirilmesi gereklidir. Bu değerlendirme süreci, hangi veri kaynaklarının mevcut olduğu, verilerin hangi formatta olduğu, mevcut veri yönetim sistemlerinin kapasitesi, veri güvenliği ve veri gölünün bu altyapıya nasıl adapte edileceği gibi önemli unsurları kapsar (Giebler ve diğerleri, 2021).

5.4.1.1. Veri Kaynaklarının Belirlenmesi

Veri gölünün kuruma entegrasyonu sürecinde, kurumların sahip olduğu veri kaynaklarının detaylı bir analizinin yapılması gerekmektedir. Bu aşamada öncelikle sunulan ve sunulması planlanan hizmetler için ihtiyaç duyulan verilerin belirlenmesi gerekmektedir. Bu verilerin nasıl elde edileceği, hangi bilgi sistemlerinin kullanılacağı ve hizmet süreçlerinde hangi tür verilerin toplanacağı gibi konular detaylı bir şekilde ele alınmalıdır. Kurumsal veriler; yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verileri içerir. Türkiye'de kamu kurumları büyük ölçüde yapılandırılmış verilere (örneğin veri tabanlarındaki finansal bilgiler) sahipken, sosyal medya, IoT cihazları ve sensörlerden gelen yapılandırılmamış

veriler de giderek yaygınlaşmaktadır. Veri gölleri, bu çok çeşitli veri kaynaklarını tek bir depolama sistemi içinde toplayarak, kurumların tüm veri varlıklarını etkin bir şekilde yönetmesine olanak sağlamaktadır. Veri kaynaklarının belirlenmesi, veri gölünün doğru şekilde yapılandırılabilmesi ve farklı veri türlerinin entegre edilebilmesi açısından kritik bir adımdır. Kurumların, veri gölünü oluştururken hangi veri kaynaklarının hangi iş süreçleri için kullanılacağını belirlenmesi gerekmektedir (Khan, Uddin ve Gupta, 2014).

Veri kalitesi ve güvenliğinin değerlendirilmesi, sürecin önemli bir parçasıdır (Sawadogo ve Darmont, 2021). Veri kaynaklarının belirlenmesi sürecinde ulusal ve uluslararası yasal düzenlemelerin ve standartların (ISO TSE K 523, TS ISO/IEC 27001 v.b.) dikkate alınması kritik öneme sahiptir. Kişisel verilerin korunması, veri güvenliği ve kurumsal mevzuatlara uyum konuları, veri gölü entegrasyonunun her aşamasında göz önünde bulundurulmalıdır.

5.4.1.2. Mevcut Veri Yönetim Sistemlerinin Değerlendirilmesi

Bir veri gölünün mevcut kurumsal yapıya entegrasyonunun başarılı olabilmesi için, mevcut veri yönetim sistemlerinin kapasitesinin ve veri gölü ile entegrasyon potansiyelinin dikkatle değerlendirilmesi gerekmektedir. Kurumların mevcut sistemlerinin veri gölü ile uyumlu olmasını sağlayacak altyapıya sahip olup olmadığı göz önünde bulundurmalıdır (Giebler ve diğerleri, 2021). Bu değerlendirme birkaç önemli boyuta odaklanmalıdır. Veri depolama kapasitesi değerlendirme sürecinde ilk ele alınması gereken alanlardan biridir. Veri gölleri genellikle büyük miktarda veriyi depolamak için kullanıldığından, mevcut depolama altyapısının kapasitesi gözden geçirilmelidir. Kurumların mevcut veri ambarlarının kurulu olduğu sistemler (donanım ve yazılım altyapıları) veri göllerinin ve veri gölleri çerçevesinde gerçekleştirilmesi planlanan büyük veri uygulamalarının kapasite ve veri transferi hız gereksinimlerini karşılayamıyorsa bu eksiklikleri gidermek için bulut veya büyük veri platformu tabanlı veri gölü çözümlerine geçiş değerlendirilmelidir (Fang, 2015). Bununla birlikte, bulut dışı

çözümler için donanımsal güncellemeler ve mevcut altyapının aşamalı olarak geliştirilmesi de bir alternatif olarak ele alınabilir. Böylece, tamamen yeniden yapılanma yerine mevcut altyapının kademeli bir şekilde yenilenmesi sağlanabilir. Böylece kurumların daha fazla veri depolayabilme kapasitesine sahip olması sağlanarak, veri gölü tabanlı büyük veri analizlerine olanak tanınmış olur.

Bir diğer kritik unsur, veri işleme yetkinliğidir. Mevcut veri yönetim sistemlerinin veri işleme yetkinlikleri kapsamlı bir şekilde incelenmelidir. Özellikle kurumların yapılandırılmamış veya yarı yapılandırılmış verileri işleyebilme kapasitesi değerlendirilmelidir. Örneğin bir kurumun mevcut sistemleri, yapılandırılmış verilerin işlenmesinde güçlü olabilir. Ancak yapılandırılmamış veri işleme kapasitesini artırmak için büyük veri çözümleri olan Hadoop ve Spark gibi platformlar gerekebilir (Miloslavskaya ve Tolstoy, 2016). Bu platformlar, veri göllerinin esnek veri işleme gereksinimlerine uyum sağlamasını kolaylaştırmaktadır.

Mevcut sistemlerin entegrasyonu da göz ardı edilmemesi gereken bir faktördür. Veri gölünün, kurum içindeki iş süreçlerini yönetmek ve veri akışını sağlamak için kullanılan ve veri oluşturma ve işleme süreçlerinde kullanılan ERP (Enterprise Resource Planning - Kurumsal Kaynak Planlaması), CRM (Customer Relationship Management - Müşteri İlişkileri Yönetimi) veya SCM (Supply Chain Management - Tedarik Zinciri Yönetimi) gibi iş uygulama yazılımları ile uyumlu olması gerekmektedir. Bu entegrasyon, kurumların veri gölü ile mevcut iş süreçleri arasında kesintisiz bir bilgi akışı sağlayarak verimliliği artıracaktır. Bu nedenle, entegrasyon planları, mevcut sistemlerin veri gölü ile nasıl bütünleştirileceği konusunda ayrıntılı bir değerlendirme içermelidir (Sawadogo ve Darmont, 2020). Bu unsurların birlikte ele alınması, kurumların mevcut veri yönetim sistemlerini değerlendirmesine ve veri gölü entegrasyon sürecini daha verimli ve başarılı bir şekilde planlamasına olanak tanıyacaktır.

5.4.1.3. Veri Yönetişimi ve Güvenlik Gereksinimlerinin Değerlendirilmesi

Kurumların mevcut veri yönetim politikaları ve güvenlik gereksinimlerinin veri gölü mimarisiyle nasıl uyum sağlayacağı kritik bir değerlendirme alanıdır. Özellikle süreçlerin kurum mevzuatına ve yasal düzenlemelere uyum sağlaması büyük önem taşımaktadır. Kurumlar ilgili kanun, Cumhurbaşkanlığı Kararnamesi, Cumhurbaşkanı Kararı, yönetmelik vb. ilkelere uygun olarak hazırladıkları mevzuat hükümleri doğrultusunda veri süreçlerini yönetmelidir. Bu kapsamda veri gölüne geçiş sürecinde veri güvenliği, kişisel verilerin korunması, erişim hakları gibi yasal düzenlemeler dikkate alınarak süreç yürütülmelidir. Ayrıca veri yönetişimi süreçlerinin veri gölü mimarisine dahil edilebilmesi için mevcut güvenlik protokollerinin yeterli olup olmadığı titizlikle incelenmelidir. Kurumlar veri yönetişimi politikalarını düzenli olarak izlemeli ve veri kalitesini artırmak için güncellemeler yapmalıdır. Türkiye'nin Ulusal Veri Stratejisi hazırlıkları kapsamında, veri yönetişimi çerçevesine yönelik politikalar, mevzuat, düzenlemeler, kurumlar, mekanizmalar, süreçler, insan kaynakları, teknoloji ve altyapı gibi konularda öneriler şekillendirilmektedir (“T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi”, 2024).

Mevcut veri yönetim sistemlerinin veri gölüne entegrasyonu sırasında erişim kontrolleri ve veri şifreleme süreçleri büyük önem taşımaktadır. Bu aşamada, özellikle hassas ve gizli verilerin şifrelenmesi ve yetkilendirilmiş kullanıcıların belirlenmesi gerekmektedir (Miloslavskaya ve Tolstoy, 2016). Bu strateji, veri gölündeki hassas ve gizli verilere erişimi sınırlandırırken aynı zamanda veri analistlerinin ve veri bilimcilerin gerekli verilere ulaşabilmesini sağlayacaktır (Sawadogo ve Darmont, 2021).

Erişim kontrolü ve veri güvenliğinin yanı sıra, veri sınıflandırma ve etiketleme de önemli bir güvenlik stratejisidir. Sınıflama ve etiketleme ile veri göllerinde saklanan verilerin hassasiyet derecesine göre düzenlenmesi sağlanmaktadır. Türkiye'deki kurumlar, KVKK kapsamında kişisel verilerin doğru bir şekilde tanımlanmasını ve sınıflandırılmasını sağlamalıdır. Bu sınıflandırma süreci, hangi

verilerin şifrelenmesi gerektiğini, hangi verilerin anonimleştirilmesi gerektiğini ve hangi verilerin açık bir şekilde saklanabileceğini belirlemede kritik rol oynamaktadır (Giebler ve diğerleri, 2021). Hassas ve gizli verilerin korunmasında veri maskeleme ve anonimleştirme uygulamaları veri gölü stratejisinin önemli bir parçası olarak ele alınmalıdır. Özellikle test ve geliştirme ortamlarında kullanılan verilerin anonimleştirilmesi, veri gizliliğini koruma açısından büyük önem taşımaktadır. Bu teknikler yazılım testi, kullanıcı eğitimi gibi gerçek verinin kullanılmasının gerekli olmadığı durumlarda güvenli ve etkili bir çözüm sunmaktadır (Tunks, 2024). Veri gölündeki verilerin kalitesini ve doğruluğunu korumak için veri yönetişimi ve güvenlik politikaları belirlenmeli ve sıkı bir şekilde uygulanmalıdır.

5.4.1.4. Teknolojik Uyum ve Planlama

Veri gölünün mevcut altyapıya uyumu kadar, bu altyapının gelecekteki veri yönetimi gereksinimlerine ne kadar yanıt vereceği de dikkatle değerlendirilmelidir. Bu çerçevede kurumların mevcut iş süreçleri ve sunduğu hizmetlerin yanı sıra uzun vadeli stratejik planlarını da dikkate alarak veri kapsam ve yönetim planlaması yapması gerekmektedir. Veri göllerinin kurumların stratejik hedefleri doğrultusunda yalnızca bugünkü ihtiyaçlara değil, aynı zamanda gelecekteki işlevsellik beklentilerine de yanıt verebilecek bir yapıda olması önemlidir. Bu bağlamda kurumlar veri hacminin ve veri türlerinin zamanla artacağını göz önünde bulundurarak, ölçeklenebilir çözümler geliştirmelidir. Kurumlar sundukları ve gelecekte sunmayı planladıkları hizmetleri kapsamlı bir şekilde gözden geçirerek teknik alt yapı gereksinimlerini planlamalı ve gerekiyorsa alt yapı değişiklik ve yenileme çalışmalarını öncelikli olarak başlatmalıdır. Bulut tabanlı veri depolama ve işleme sistemleri gelecekteki veri yönetimi ihtiyaçlarını karşılamada büyük bir potansiyele sahiptir (Fang, 2015).

Bulut tabanlı çözümler veri gölleri için ölçeklenebilirlik sağlayarak, veri hacmi arttıkça depolama kapasitesinin kolayca genişletilmesini mümkün kılar. Bulut

sistemlerinin sunduğu esneklik, kurumlar için maliyet etkin bir çözüm sunarken, aynı zamanda depolama ihtiyaçlarına hızlı yanıt verebilme olanağı sağlar. Veri hacmi arttıkça, bulut tabanlı depolama sistemleri otomatik olarak genişleyerek kapasiteyi artırır ve böylece kurumların maliyetlerini optimize eder (Mathis, 2017).

Büyük veri platformları da veri gölleri için önemli bir çözüm sunar. Hadoop ve Spark gibi platformlar, dağıtık veri işleme yetenekleri sayesinde yüksek hacimli verilerle başa çıkabilme kapasitesine sahiptir. Bu platformlar, veri hacmi artsa da işlem performansını ve veri işleme hızını korur (Miloslavskaya ve Tolstoy, 2016).

Kurumların gelecekteki veri yönetimi ihtiyaçlarına yanıt verebilmesi için ölçeklenebilir bulut çözümleri ve büyük veri platformları gibi yenilikçi teknolojilere yatırım yapmaları gerekmektedir. Ayrıca, özellikle kamu kurumları gibi büyük ve karmaşık yapılarda, bu teknolojilere geçişin aşamalı bir planlama ile yapılması, geçiş sürecindeki kesintilerin önlenmesine ve mevcut sistemlerin yeni yapıyla uyumlu hale getirilmesine yardımcı olacaktır. Bu çözümler, veri göllerinin esnekliğini ve verimliliğini artırarak, kurumların değişen veri yönetimi ihtiyaçlarına dinamik bir şekilde uyum sağlamalarını kolaylaştıracaktır.

5.4.2. Kurumsal Uygulama İçin Hazırlık Süreci

Veri gölü kurulumu, yalnızca teknik bir altyapının sağlanması ile değil, aynı zamanda kurumun mevcut iş süreçlerinin, veri stratejileri ve politikalarının ve teknolojik gereksinimlerinin veri gölü mimarisine uygun hale getirilmesiyle başlar. Bu nedenle, hazırlık süreci kritik öneme sahiptir ve veri gölünün kurumsal stratejilerle uyumlu olması için dikkatli bir planlama gerektirmektedir (Giebler ve diğerleri, 2021). Bu aşamada çalışanların veri gölü sistemine uyum sağlayabilmesi için eğitim süreçleri ve değişim yönetimi stratejilerinin uygulanması önemlidir.

5.4.2.1. Kurumsal Veri Stratejilerinin Belirlenmesi

Veri gölünün başarılı bir şekilde uygulanabilmesi için kurumların mevcut veri strateji ve politikalarının gözden geçirilmesi ve veri gölü ile uyumlu hale getirilmesi kritik bir adımdır. Bu süreçte, kurumların büyük veri yönetimindeki stratejik hedeflerini veri gölü ile uyumlu hale getirmeleri gerekmektedir. Veri gölü, karar destek sistemleri, analitik süreçler ve veri yönetişimi gibi stratejik hedeflerin merkezinde yer alarak verilerin işlenmesi, analiz edilmesi ve iş süreçlerine entegrasyonu için ideal bir platform sunmaktadır (Sawadogo ve Darmont, 2021).

Veri gölünün kurumun stratejik hedeflerine hizmet edebilmesi için hangi iş süreçlerine katkı sağlayacağı ve hangi stratejik hedeflere hizmet edeceği net bir şekilde tanımlanmalıdır. Stratejik hedeflerin desteklenebilmesi için veri gölünün hangi veri kaynaklarını içereceği, hangi analiz araçlarının kullanılacağı ve hangi karar destek sistemlerine adapte edileceği detaylı bir şekilde planlanmalıdır (Alharthi, Krotov ve Bowman, 2017).

Bir diğer önemli adım, veri yönetimi ihtiyaçlarının belirlenmesidir. Kurumların iş süreçlerine, işlevsel gereksinimlerine ve karar destek sistemlerine katkı sağlayacak verilerin tanımlanması, veri gölünün işlevselliği ve etkinliği açısından önemlidir. Bu aşamada, veri kaynaklarının türü, verilerin formatı ve analiz süreçlerine uygunluğu incelenmelidir. Veri gölü stratejileri, kurumların iş süreçlerine ve stratejik hedeflerine uyumlu olmalıdır. Bu ise, doğru veri kaynaklarının seçilmesi, analiz araçlarının optimize edilmesi ve karar destek sistemlerinin entegrasyonu ile mümkündür.

5.4.2.2. Teknolojik Gereksinimlerin Planlanması

Veri gölünün kurumsal yapıya entegrasyonu için gerekli olan teknolojik altyapı dikkatle planlanmalıdır ve gelecekte yaşanacak teknolojik gelişmelere uyum sağlayacak bir yapı oluşturulmalıdır. Bu süreç, bölüm 5.4.1.'de anlatıldığı üzere ilk olarak mevcut teknoloji altyapısının kapsamlı bir şekilde gözden geçirilmesini

ve gerektiğinde altyapı değişiklik ve yenileme çalışmalarının hayata geçirilmesini içermektedir. Veri depolama kapasitesi, işleme gücü, ağ altyapısı ve güvenlik önlemleri, bu hazırlık sürecinde belirlenmesi gereken temel unsurlar arasında yer almaktadır (Giebler ve diğerleri, 2021). Ayrıca, veri gollerinin entegrasyonu ile birlikte gerçekleştirilmesi planlanan büyük veri analizi, veri madenciliği, makine öğrenmesi gibi işlemler, teknik altyapı gereksinimlerini önemli ölçüde etkiler. Bu tarz analiz süreçlerinin etkin bir şekilde yürütülebilmesi için planlama aşamasında teknik altyapının bu analizleri destekleyecek şekilde yapılandırılması gerekir. Teknolojik altyapının doğru bir şekilde planlanması, veri gölünün etkin ve verimli bir şekilde kullanılmasını sağlayacaktır.

Bulut tabanlı çözümler, veri gölü uygulamaları için esneklik ve ölçeklenebilirlik sağlayan önemli bir çözüm yöntemidir. Bulut bilişim, veri hacmi arttıkça otomatik olarak ölçeklenebilme yeteneğine sahip olduğundan, maliyet avantajı sağlamaktadır (“Snowflake”, t.y.). Büyük veri platformları da veri gölünün yüksek hacimli verilerle başa çıkabilmesi için etkin bir altyapı alternatifidir. Hadoop, Spark gibi platformların entegrasyonu, veri işleme süreçlerini hızlandırarak veri gölüne gelen yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verilerin etkili bir şekilde yönetilmesini sağlamaktadır. HDFS gibi dağıtık depolama çözümleri, veri gölünün depolama kapasitesini artırırken aynı zamanda işleme gücünü de yükseltmektedir (Miloslavskaya ve Tolstoy, 2016). Bu platformların kullanımı, büyük veri uygulamalarında giderek yaygınlaşmakta ve veri gölü mimarilerinde kritik bir rol oynamaktadır (Terrizzano ve diğerleri, 2015). Ancak bu platformların uygulanması sırasında, mevcut altyapı ile entegrasyonun dikkatle planlanması ve büyük ölçekli geçiş süreçlerinin aşamalı bir şekilde hayata geçirilmesi önemlidir. Böylece kurumların geçiş sürecinde operasyonel kesintiler yaşamaması sağlanmış olur.

Veri entegrasyon araçları, veri gölünün mevcut sistemlerle entegrasyonu için kilit bir öneme sahiptir. Bu araçlar, farklı veri kaynaklarından gelen verilerin tek bir veri gölü platformunda toplanmasını sağlamaktadır. Talend, Informatica ve Apache Nifi gibi veri entegrasyon araçları, veri taşıma, veri kalitesi yönetimi ve gerçek

zamanlı veri akışı gibi temel fonksiyonlarıyla, farklı veri türlerinin entegrasyonunu kolaylaştırmaktadır ve veri gölünde düzenli bir yapı oluşturmaya yardımcı olacaktır (Sawadogo ve Darmont, 2021).

Veri gölünün kurumsal yapıya entegrasyonu için teknolojik altyapının kapsamlı bir şekilde planlanması ve en yeni teknolojilerle desteklenmesi, veri gölünün performansını ve esnekliğini artıracaktır. Bu süreçte hem teknik altyapının yenilenmesi hem de kurum içindeki operasyonel süreçlerin veri gölü yapısına uyarlanması, başarının temel unsurları olarak öne çıkmaktadır.

5.4.2.3. Veri Güvenliği ve Uyumluluk Stratejilerinin Planlanması

Veri gölü kurulum sürecinde en kritik aşamalardan biri, veri güvenliği ve uyumluluk stratejilerinin belirlenmesidir. Özellikle yukarıda açıklandığı üzere Türkiye'de KVKK gibi yasal düzenlemelere uyum sağlamak zorunludur. Bu nedenle, veri gölünün güvenliğinin sağlanması ve verilerin yasal düzenlemelere uygun şekilde işlenmesi büyük bir öneme sahiptir.

Veri gölünün güvenliği, verilerin yetkisiz erişimlerden korunması açısından birincil öneme sahiptir. Bu bağlamda, şifreleme, veri maskeleme ve anonimleştirme gibi teknikler, hassas ve gizli verilerin güvenliğini sağlamada kritik bir rol oynamaktadır. Özellikle hassas verilerin işlendiği kurumlarda, verilerin korunması için güçlü güvenlik önlemlerinin alınması zorunludur. Örneğin Bankacılık Düzenleme ve Denetleme Kurumu (BDDK) tarafından yayımlanan yönetmelikler gereği, Türkiye'de finansal kuruluşların veri güvenliği için çok katmanlı bir yaklaşım benimsemesi gerekmektedir. Bu yönetmelikler çerçevesinde kurumların veri şifreleme, erişim kontrolü ve sürekli izleme gibi teknikleri bir arada kullanmaları gerekmektedir (“T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi”, 2020). Sawadogo ve Darmont (2021), veri gölleri için güvenlik stratejilerinin verilerin yaşam döngüsü boyunca uygulanması gerektiğini vurgulamaktadır. Bu

durum, veri alımından depolamaya, işlemeden analize kadar tüm aşamalarda güvenlik önlemlerinin uyarlanmasını gerektirmektedir.

Veri gölünün kurulumu sırasında gölün KVKK ve uluslararası veri koruma düzenlemelerine (örneğin GDPR) uyum sağlanması, süreçlerin başarıyla yürütülebilmesi için gereklidir. Veri gölündeki verilerin işlenmesi ve depolanması, yasal düzenlemelere uygun olarak yapılandırılmalıdır. Bu süreçte, veri erişimi ve kullanımıyla ilgili politikaların tanımlanması ve sürekli denetlenmesi gerekmektedir. Kurumlar, veri gölünün yasal uyumluluğunu sürekli izlemeli ve güncel regülasyonlara uygunluğunu sağlamalıdır. Giebler ve diğerleri (2021), veri gölleri için uyumluluk stratejilerinin dinamik ve sürekli güncellenen bir yapıda olması gerektiğini belirtmiştir. Bu, değişen yasal gereksinimlere hızlı adaptasyon sağlamak için kritik bir unsurdur. Veri gölü mimarilerin kurulumunda ve kurumların veri sistemlerine entegrasyonunda, güvenlik ve uyumluluk stratejileri, verilerin korunması ve yasal gerekliliklerin karşılanması için dikkatle ele alınmalıdır.

5.4.2.4. Eğitim ve Değişim Yönetimi

Veri gölünün başarılı bir şekilde uygulanabilmesi için yalnızca teknik altyapının değil, aynı zamanda insan kaynakları ve organizasyonel süreçlerin de uygun şekilde hazırlanması gerekmektedir. Kurum çalışanlarının veri gölü sistemine uyum sağlayabilmesi için eğitim süreçleri ve değişim yönetimi stratejileri uygulanmalıdır (“DATAVERSITY”, t.y.).

Veri gölüne geçiş sürecinde, veri analitiği uzmanlarının, veri bilimcilerinin ve diğer ilgili çalışanların yeni sistemler hakkında bilgi sahibi olması büyük önem taşımaktadır. Hadoop, Spark, Talend gibi büyük veri platformları ve entegrasyon araçları vb. hakkında eğitim verilmesi, kurum içinde veri gölünün etkin bir şekilde kullanılabilmesi için kritik bir adımdır. Kurumlar, veri yönetimi süreçlerine katkı sağlayacak çalışanlara, büyük veri analitiği, veri işleme araçları ve veri güvenliği konularında kapsamlı eğitim programları sunmalıdır (Alharthi ve diğerleri, 2017).

Kurumların veri yönetişimi ve veri gölü uygulamalarında karşılaştıkları en büyük zorluklardan biri, nitelikli personel eksikliğidir (Nargesian, Zhu, Miller, Pu ve Arocena, 2019). Bu bağlamda, eğitim programlarının Türkiye'deki kurumlar için önemi daha da artmaktadır. Birleşmiş Milletler Kalkınma Programı (United Nations Development Programme - UNDP) tarafından hazırlanan bir raporda, Türkiye için veri yönetişimi çerçevesinde kapasite geliştirme ve eğitim programlarının önemi vurgulanmıştır (“UNDP”, 2024). Veri gölünün uygulanması, kurumların iş süreçlerinde değişikliklere yol açabilir. Bu nedenle, değişim yönetimi stratejilerinin devreye alınması gereklidir. Değişim yönetimi stratejisi, organizasyonlarda yeni teknolojilerin veya süreçlerin benimsenmesi sırasında ortaya çıkabilecek direnci azaltmak, çalışanları yeni sisteme adapte etmek ve geçiş sürecini sorunsuz bir şekilde yönetmek için kullanılan sistematik bir yaklaşımdır. Özellikle kamu kurumlarında, geleneksel sistemlerden veri gölü mimarisine geçiş, iş süreçlerinde önemli değişimlere yol açabileceğinden, değişim yönetiminin erken aşamalarda planlanması kritik bir önem taşır. Çalışanların veri gölü ile çalışmaya uyum sağlaması ve yeni veri işleme süreçlerine uyum sağlayabilmesi için, kurumlar değişim yönetimi süreçlerini yakından takip etmeli ve çalışanlara gerekli destekleri sunmalıdır (“Collectiv”, t.y.).

Etkili bir değişim yönetimi stratejisi, üst yönetim desteği, iletişim planı, kullanıcı katılımı ve sürekli geri bildirim gibi unsurları içermelidir. Kurumlar için bir diğer önemli nokta ise veri kültürünün oluşturulmasıdır. Değişim yönetimi stratejileri, aynı zamanda veri odaklı düşünme ve karar verme kültürünün geliştirilmesine de odaklanmalıdır. Eğitim süreçleri ve değişim yönetimi stratejileri birlikte ele alındığında, kurumlar veri gölü sistemlerini daha etkin bir şekilde kullanabilir ve veri odaklı bir organizasyon kültürü oluşturabilirler.

5.4.3. Veri Gölünün Kurulumu ve Uygulama Süreci

Veri gölünün kurulum süreci, farklı veri kaynaklarından veri alımı, bu verilerin işlenmesi, üst veri yönetimi ve veri güvenliği gibi kritik adımları içermektedir.

Kurumların ihtiyaçlarına ve mevcut altyapılarına bağlı olarak bu süreçler özelleştirilmelidir. Ancak genel olarak veri gölü, büyük miktarda veriyi işleyebilecek ve çeşitli analiz süreçlerine uyum sağlayabilecek şekilde yapılandırılmalıdır (Sawadogo ve Darmont, 2021).

5.4.3.1. Veri Alımı ve İşleme Süreci

Veri alımı ve işleme süreçleri 4.4.4.1 ve 5.3.1 bölümlerinde ele alınmıştır. Bu bölümde, bu süreçlerin veri gölü yapısındaki uygulamalarına odaklanılmıştır.

Veri alımı, veri gölüne dış kaynaklardan veri çekme sürecidir. Verilerin kalitesi, doğruluğu ve işlenebilirliği bu aşamada sağlanmalıdır (Terrizzano ve diğerleri, 2015). Veri alımı, farklı yöntemlerle yapılabilmektedir. Gerçek zamanlı veri alımında, gerçek zamanlı sensörlerden ve sosyal medya gibi kaynaklardan anlık olarak veri gölüne aktarım yapılmaktadır. Toplu veri alımı ise belirli aralıklarla büyük miktarda verinin göle aktarımıdır (“AWS”, t.y.). Hastanelerdeki hasta verilerinin günlük veya haftalık transferi buna örnek verilebilir.

Veri alımı sonrasında verilerin işlenmesi gereklidir. Bu süreç, veri temizleme, dönüşüm ve organize etme adımlarını içermektedir. İlk olarak, veri temizleme aşamasında eksik, hatalı veya tekrarlı veriler düzeltilir. Ardından, farklı formatlarda gelen veriler, analiz edilebilir formata dönüştürülür. Veri işleme aşamasında, Türkiye'deki kurumlar için özellikle önemli olan bazı unsurlar olabilir. Örneğin, Türkçe metinlerin doğru şekilde işlenmesi ve analiz edilmesi için özel dil işleme algoritmalarının geliştirilmesi gerekebilir. Son olarak, verilerin organize edilmesi aşamasında, veriler kategorilere ayrılarak anlamlı hale getirilmektedir. Yerel dinamikleri ve düzenlemeleri göz önünde bulunduran bir veri alımı ve işleme stratejisi, veri gölü uygulamalarının başarısını artıracaktır.

5.4.3.2. Üst Veri Yönetimi ve İzlenebilirlik

Veri gölünde etkili bir üst veri yönetimi, verilerin izlenebilirliğini ve güvenliğini sağlamak açısından kritik bir rol oynamaktadır. 4.2.2 ve 5.3.1 bölümlerinde üst veri yönetiminin veri göllerindeki önemine detaylı olarak değinilmiştir. Kısaca özetlemek gerekirse, üst veri, veri gölündeki verilerin kaynağını, yapısını, veri tanımını ve erişim süreçlerini yöneterek verilerin düzenli bir şekilde işlenmesini sağlamaktadır (Suriarachchi ve Plale, 2016).

Bu bağlamda, veri göllerinde etkili bir izlenebilirlik sağlayan üst veri katalogları, verilerin hangi kaynaklardan alındığını ve hangi işlemlere tabi tutulduğunu izlemek için kullanılmaktadır. Üst veri yönetimi sayesinde, kurumlar verilerinin hangi süreçlerde kullanıldığını kontrol edebilmekte ve veri güvenliği stratejilerini optimize edebilmektedir (Halevy ve diğerleri, 2016). Ayrıca 5.2.3 bölümünde anlatıldığı üzere, bu süreçler veri güvenliğini artırırken aynı zamanda Türkiye'de geçerli olan KVKK gibi düzenleyici gereksinimlere uyum sağlamak için de önemlidir. Doğru yapılandırılmış üst veri yönetimi ve sürekli izlenebilirlik stratejileri, veri keşfi süreçlerini hızlandırır ve veri kalitesini güvence altına alarak büyük veri yönetiminin etkinliğini artırır (Miloslavskaya ve Tolstoy, 2016). Türkiye'deki kurumlar için üst veri yönetimi, yalnızca veri güvenliği açısından değil, aynı zamanda yasal düzenlemeler ve veri yönetimi olgunluk süreçleri açısından da vazgeçilmez bir unsurdur.

5.4.3.3. Güvenlik ve Erişim Yönetimi

Veri gölünde güvenliğin sağlanması, veri ihlallerinin önlenmesi ve yasal düzenlemelere uyum açısından önemlidir. 5.2.3 ve 5.3.1 bölümlerinde detaylı olarak ele alındığı üzere, Türkiye'deki kurumlar veri gölü kurulum sürecinde, ulusal mevzuatlara ve uluslararası veri güvenliği standartlarına uyum sağlamak zorundadır. Bu uygulamalar, özellikle bulut tabanlı çözümlerle kullanılan veri

göllerinin entegrasyonunda dikkatlice ele alınmalıdır (Sawadogo ve Darmont, 2021).

Veri gölünde erişim kontrolü ve yetkilendirme mekanizmaları ise kimin hangi verilere erişebileceğini düzenleyerek kullanıcı rollerine göre erişim yetkilerini tanımlamaktadır. Kurumlar, sıkı erişim kontrol politikaları uygulayarak kullanıcı verilerinin korunmasını güvence altına almak zorundadır (Miloslavskaya ve Tolstoy, 2016). Çok faktörlü kimlik doğrulama sistemleri, güvenliği artırmanın bir başka etkili yoludur. Bu sistemler, sadece şifreyle değil, ek doğrulama katmanları ile kimlik doğrulama gerektirmektedir. Ek doğrulama özellikle hassas verilere yetkisiz erişimi önlemek için kritik bir adımdır (Halevy ve diğerleri, 2016). Veri gölünün güvenliği, yalnızca kurulum sırasında değil, sürekli olarak izlenmeli ve denetlenmelidir. Düzenli güvenlik denetimleri, veri ihlallerinin erken tespiti ve önlenmesi açısından önemlidir. Bu bağlamda veri erişim raporlamaları ve denetimleri, veri gölü güvenliğinin sürekliliğini sağlayacaktır (Miloslavskaya ve Tolstoy, 2016).

Kurumlar için güvenlik ve erişim yönetimi, sadece yasal bir zorunluluk değil, aynı zamanda veri odaklı stratejilerin başarılı bir şekilde uygulanması için kritik bir gerekliliktir. Etkili güvenlik uygulamaları, potansiyel veri ihlalleri ve güvenlik risklerini en aza indirecektir.

5.4.4. Kademeli Olgunlaşma ve Gelişim Süreci

Önerilen veri gölü mimarisi, başlangıçta ham verilerin toplandığı geniş veri deposu olarak işlev görse de zamanla olgunlaşarak daha yapılandırılmış ve işlenebilir hale gelmektedir. Bu süreç, veri gölünün kademeli olarak geliştirilmesi ve olgunlaşmasıyla işleyişini sürekli olarak iyileştiren bir yapı olmasını sağlamaktadır. Önerilen modelde kurumlar, veri gölünü kurduktan sonra verilerin daha etkin kullanılması ve daha ileri analiz yöntemlerinin uygulanması için kademeli olgunlaşma süreci adımlarını izlemelidir.

5.4.4.1. Olgunluk Seviyeleri ve Kademeli Gelişim

Veri gölünün kurumsal sistemlerde olgunlaşması, veri yönetimi süreçlerinin kademeli olarak gelişmesini sağlamaktadır. Bu olgunlaşma süreci, verilerin başlangıçta ham halde depolandığı, ardından işlenip anlamlı hale getirildiği adımları kapsamaktadır.

Başlangıç seviyesinde, veri gölüne alınan ham veriler düşük seviye işleme tabi tutulmaktadır. Bu aşamada, verilerin kalitesi düşük olabilir ve veri yönetimi süreçleri henüz gelişmemiştir. Türkiye'deki birçok kurum, veri gölü kurulumunun ilk aşamalarında bu seviyede olacaktır. Bu dönemde veri gölü daha çok veri depolama amacıyla kullanılmaktadır ve veri işleme süreçleri oldukça sınırlıdır (Sawadogo ve Darmont, 2021). Bir sonraki aşama olan gelişen seviyede, veri gölündeki verilerin kalitesi artmaya başlamaktadır. Veriler temizlenmekte, organize edilmekte ve farklı kategorilere ayrılmaktadır. Önerilen modelde verilerin türüne göre göletler oluşturulması önerilmektedir. “2.1.2. Türüne göre Veri” başlığında anlatılan veri türleri doğrultusunda gelişen seviyede verilerin “Gelişen Analog Veri Göleti”, “Gelişen Uygulama Veri Göleti” ve “Gelişen Metinsel Veri Göleti” olarak 3 ayrı kategoriye ayrılması önerilmektedir. Böylece verileri türlerine göre veri analitiği süreçlerine hazırlamak mümkün olacaktır. Bu aşamada veriler daha düzenli hale getirilmeli ve veri analitiği süreçlerine uyarlanması için gerekli adımlar tamamlanmalıdır. Gelişen Seviyede ayrıca veri yönetişimi süreçleri devreye girmektedir. Olgunlaşmış seviyede, veri gölü tam anlamıyla organize edilmiş yüksek kaliteli verilerden oluşmaktadır ve ileri analizler için hazır hale gelmiştir. Önerilen mimari kapsamında, olgunlaşmış seviye içinde de gelişen seviyede olduğu gibi verilerin türüne göre alt göletler bulunmaya devam edecektir. Bu kapsamda, Olgun seviye katmanında “Olgun Analog Veri Göleti”, “Olgun Uygulama Veri Göleti” ve “Olgun Metinsel Veri Göleti” olmak üzere üç ayrı olgun veri göleti yer alacaktır. Bu aşamada, veriler iş zekâsı araçları veya makine öğrenimi modelleri tarafından kullanılabilmektedir. Veri gölünün olgunluk seviyeleri, kurumların veri yönetimi kapasitelerinin geliştirilmesi ve veri analizlerinin daha ileri seviyelere taşınmasında önemli bir rol oynamaktadır.

5.4.4.2. Veri Analitiği ve İş Zekâsı Uygulamaları

Veri gölünün kademeli olarak olgunlaşması, gelişmiş veri analitiği ve iş zekâsı süreçlerinin devreye sokulmasını sağlamaktadır. Bu süreçte, veri gölü içerisinde toplanan veriler anlamlı hale getirilebilmekte ve iş süreçlerinde karar alma mekanizmalarına dâhil edilebilmektedir. Veri gölünün olgunlaşması, öncelikle ham verilerin temizlenmesi, meta veri yönetimi ve veri entegrasyonu gibi temel aşamalarla başlar. Veri analitiği ve iş zekâsı süreçleri ise bu olgunlaşmayı takip eden ileri aşamalarda devreye girerek verilerin anlamlandırılmasını destekler. Etkin bir veri gölü mimarisi ile sistematik olarak toplanmış kurumsal verilerin analiz edilerek bilgiye dönüştürülmesi, kurumlarım hem mevcut iş süreçlerine katkı sağlama, hem de ileride sunmayı planladıkları hizmetlerin planlanmasına öncülük etme potansiyeline sahiptir. Türkiye'de son yıllarda kurumsal verinin veri analiz ve veri madenciligi gibi yöntemler ile incelenmesi ve değerlendirilmesi büyük önem kazanmıştır ve Türkiye'nin Ulusal Veri Stratejisi hazırlıkları kapsamında, ileri analitik ve yapay zekâ veri yönetişimi ihtiyaçlarına yönelik kapsamlı çalışmalar yürütülmektedir ("T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi", 2024).

Olgunlaşmış veri gölleri, iş zekâsı uygulamaları için güçlü bir temel sağlamaktadır. Bu uygulamalar, veri gölünde depolanan verilerin iş süreçlerine uyarlanmasını ve karar destek sistemlerine katkı sağlamasını mümkün kılmaktadır. İş zekâsı araçları, kurumsal performansı artırmak ve stratejik kararlar için önemli bilgiler sunmaktadır (Sivarajah, Kamal, Irani ve Weerakkody, 2017). Veri gölünün olgunlaşması ile gelişmiş veri analitiğine dayalı modeller de devreye girmektedir. Ayrıca veri göllerinin olgunlaşmasıyla gerçek zamanlı analitik kapasitesi de artmaktadır. Gerçek zamanlı analitik, anlık kullanıcı davranışlarını analiz ederek kişiselleştirilmiş öneriler sunabilir ya da dolandırıcılık tespiti gibi güvenlik uygulamalarında etkili bir şekilde kullanılabilir (Giebler ve diğerleri, 2021). Veri gölünün kademeli gelişimi hem iş zekâsı hem de gelişmiş veri analitiği uygulamaları açısından kurumlara önemli stratejik avantajlar sağlamaktadır.

Özellikle büyük ölçekli veriler için, veri madenciliği ve büyük veri analiz teknikleri kullanılarak kurumun performansını iyileştirecek, operasyonel verimliliğini artıracak, müşteri memnuniyetini iyileştirecek ya da maliyetleri düşürecek stratejiler belirlenmesi veya iş süreçlerinde aksaklığa sebep olan sıkıntıların tespit edilmesi mümkündür. Kurumsal verilerin bilgiye dönüştürülmesi, kurum hedeflerinin daha somut ve veri temelli stratejilerle şekillendirilmesini sağlayacaktır.

5.4.4.3. Makine Öğrenimi ve Yapay Zekâ Uygulamaları

Önerilen modelde veri olgun seviyeye ulaştığında makine öğrenimi modelleri ile veri gölündeki büyük veri setlerini işleyerek anlamlı sonuçlar üretmek mümkün olmaktadır. Makine öğrenimi ve yapay zekâ teknikleri kullanılarak veri işleme süreçleri otomatikleştirilebilmekte ve daha karmaşık analizler yapılabilmektedir. Türkiye'deki büyük kurumlar, müşteri deneyimlerini geliştirmek, finansal riskleri azaltmak ve operasyonel verimliliği artırmak amacıyla bu teknolojileri kullanmaya başlamıştır. Özellikle perakende sektöründe, bulut bilişim altyapısıyla bütünleştirilmiş büyük veri analitiği ve yapay zekâ uygulamaları, müşteri deneyimini iyileştirmek, operasyonel verimliliği artırmak ve rekabet avantajı elde etmek için kullanılmaktadır (“IDC Türkiye”, 2024).

Makine öğrenimi uygulamaları, veri gölüne entegre edilerek büyük veri setlerinin otomatik analizini sağlamaktadır. Bu süreç, veri gölüne eklenen yeni verilerin otomatik olarak sınıflandırılması, analiz edilmesi ve gelecekteki trendlerin tahmin edilmesi amacıyla kullanılmaktadır. Örneğin, bir üniversitenin veri gölünde öğrenci bilgileri, ders kayıtları, ders videoları, çevrim içi öğrenme platformu etkileşimleri, taranmış sınav kağıtları ve kariyer merkezi verileri gibi çeşitli veri türleri toplanabilir. Makine öğrenimi algoritmaları, bu verileri analiz ederek öğrenci başarısını etkileyen faktörleri belirleyebilir, ders içeriklerini kişiselleştirebilir ve mezuniyet sonrası kariyer yollarını tahmin edebilir. Bu yaklaşım, eğitim kalitesini

arttırırken, aynı zamanda öğrenci destek hizmetlerinin optimize edilmesini sağlayacaktır (Baker ve Inventado, 2014).

Yapay zekâ tabanlı çözümler, veri göllerindeki veriler üzerinde daha ileri analizler yapılmasına ve otomatik süreçlerin geliştirilmesine olanak tanımaktadır. Yapay zekâ, veri gölündeki verilerin daha etkin kullanılmasını sağlayarak kurumların stratejik karar alma süreçlerine doğrudan katkı sağlamaktadır (Sawadogo ve Darmont, 2021). Giebler ve diğerleri (2021), yapay zekâ uygulamalarının, özellikle büyük ölçekli ve karmaşık veri setlerinde, geleneksel analitik yöntemlere göre daha etkili sonuçlar verdiğini belirtmiştir.

Derin öğrenme ve doğal dil işleme gibi ileri yapay zekâ teknikleri, veri göllerinin olgunlaşmasıyla birlikte daha yaygın olarak kullanılmaktadır. Bu teknikler, özellikle yapılandırılmamış verilerin analizinde kritik bir rol oynamaktadır. Derin öğrenme modelleri, veri göllerinde yer alan büyük miktardaki veriyi anlamlandırmak için kullanılmaktadır (LeCun ve diğerleri, 2015). Doğal dil işleme teknikleri, veri göllerindeki metinsel verileri analiz ederek müşteri geri bildirimleri, sosyal medya gönderileri ve e-postalar gibi yapılandırılmamış veri kaynaklarından değerli içgörüler çıkarmaya yardımcı olmaktadır (Halevy ve diğerleri, 2022). Veri göllerinin makine öğrenimi ve yapay zekâ ile entegre edilmesi, kurumların büyük veri setlerinden daha anlamlı ve stratejik sonuçlar elde etmesine olanak tanımaktadır.

Makine öğrenimi platformları, veri gölünün potansiyelinden tam anlamıyla yararlanmak için özümsenmelidir. Bu teknolojiler operasyonel verimlilik, risk yönetimi ve müşteri odaklı stratejilerin geliştirilmesi için kullanılabilir. TensorFlow, PyTorch veya Apache Spark MLlib gibi açık kaynaklı çözümler, veri gölündeki büyük veri setleri üzerinde karmaşık analizler ve tahmine dayalı modeller geliştirmek için kullanılabilecek yazılımlardır (Abadi ve diğerleri, 2020).

5.4.4.4. Arşivsel Veri Uygulamaları

Olgun seviyede depolanan verilerden aktif olarak ihtiyaç duyulmayan veya uzun vadeli saklanması gereken veriler arşivsel veri katmanında uygun bir şekilde formatlanır, arşivlenir ve böylelikle gelecekte erişim için verilerin korunması sağlanır. Kurumlar tarafından toplanan veya süreçler yürütülürken oluşan verilerin hangilerinin daha sonra kullanılma olasılığı olduğu belirlenmelidir. Arşivlenecek veriler kurumun stratejik hedefleri, yasal düzenlemeler ve operasyonel ihtiyaçlar dikkate alınarak belirlenmelidir.

Kurumsal veri arşivine aktarılan veriler için kimlere erişim yetkisi verileceği, ayrıca arşivdeki hassas ve gizli veriler için uygulanacak prosedürler güvenlik ve erişim yönetimi katmanındaki kurallar ve kurum mevzuatı ile uyumlu olmalıdır. Kurumların işlevselliğini ve hukuki yükümlülüklerini yerine getirmede veri arşivi büyük önem taşımaktadır.

5.4.5. Sürekli İzleme ve Geri Bildirim

Veri gölünün başarıyla kurulup kullanılmasının ardından, sistemin sürdürülebilirliği ve etkinliği için sürekli izleme ve geri bildirim mekanizmaları devreye girmelidir. Veri gölünün performansını artırmak ve verilerin işlenebilirliğini iyileştirmek için sürekli izleme ve iyileştirme süreçleri gereklidir. Bu süreçler kapsamında, veri gölündeki veriler düzenli olarak izlenmekte ve veri kalitesini artırmak için gerekli iyileştirme adımları atılmaktadır. Sürekli izleme, veri gölünde yer alan verilerin kalitesinin korunmasını, sistemin güvenliğinin sağlanmasını ve performansın optimize edilmesini sağlamaktadır. Geri bildirim süreçleri ise sistemin zamanla iyileştirilmesine ve yeni ihtiyaçlara uyum sağlamasına yardımcı olacaktır. Kurumlar, veri göllerini sürekli olarak izleyerek göllerin işlevselliği artırabilir ve stratejik karar alma süreçlerine doğrudan katkı sağlayabilirler.

5.4.5.1. Performans İzleme ve Optimizasyon

Veri gölü, sürekli olarak veri alıp işleyerek analiz süreçlerine katkıda bulunmaktadır. Bu nedenle sistemin performansının sürekli olarak izlenmesi gereklidir. Veri gölünün performansını izlemek, verilerin işlenme hızını, depolama kapasitesini ve sistemin yanıt verme süresini değerlendirmek açısından büyük önem taşımaktadır. Bu süreç, veri gölünün operasyonel verimliliğini artırmak ve olası darboğazları tespit etmek için kritik rol oynamaktadır (Sawadogo ve Darmont, 2021).

Veri gölünün etkin bir şekilde izlenebilmesi için temel performans göstergeleri (key performance indicator - KPI) belirlenmelidir. Bu göstergeler, veri işlemenin hızı, depolama kapasitesi, sistem yanıt süresi ve veri güvenliği gibi temel unsurları kapsamaktadır. Kurumlar bu göstergeleri düzenli olarak izleyerek veri gölünün performansını optimize edebilirler. Veri gölüne alınan verilerin işlenme hızı, performans açısından sürekli izlenmelidir. Özellikle gerçek zamanlı verilerin hızlı bir şekilde işlenmesi gereklidir; toplu veri işlemleri de benzer şekilde optimize edilmelidir (Miloslavskaya ve Tolstoy, 2016). Ayrıca veri gölünün depolama kapasitesinin izlenmesi, sistemin veri hacmiyle birlikte ölçeklenmesini ve artan veri miktarının sorunsuz bir şekilde saklanabilmesini sağlamaktadır (Sawadogo ve Darmont, 2021). Yanıt süresi ise kullanıcıların veri gölüyle etkileşimi sırasında izlenmesi gereken bir diğer önemli faktördür. Düşük yanıt süreleri, kullanıcı deneyimini iyileştirmekte ve sistemin daha verimli kullanılmasını sağlamaktadır (Giebler ve diğerleri, 2021).

Veri gölünün performansı izlendikten sonra, sistemin optimizasyonu için iyileştirme adımları atılmalıdır. Kurumlar, veri gölünde meydana gelen darboğazları belirleyerek depolama çözümlerini genişletebilir veya veri işleme süreçlerini optimize edebilirler. Veri gölünün performansının sürekli izlenmesi ve optimize edilmesi, veri yönetiminin etkinliğini artırır ve kurumların veri odaklı karar alma süreçlerini hızlandırır.

5.4.5.2. Güvenlik ve Uyumluluk Denetimleri

Veri gölü güvenliği, sürekli izlenmesi ve denetlenmesi gereken kritik bir unsurdur. Kurumlar, veri güvenliğini sağlamak amacıyla düzenli güvenlik denetimleri yapmalı ve sistemin yasal uyumluluk gereksinimlerini karşılayıp karşılamadığını kontrol etmelidir. Kişisel Verileri Koruma Kurumu'nun rehberine göre, veri sorumluları, kendi kurumlarında kanun hükümlerinin uygulanmasını sağlamak amacıyla gerekli denetimleri yapmakla yükümlüdür (“KVKK”, t.y.). Veri gölünün güvenliği düzenli olarak izlenmelidir. Veri erişimlerinin ve kullanıcı aktivitelerinin kaydedilmesi, güvenlik ihlallerinin önceden tespit edilmesine yardımcı olacaktır. Bunun yanı sıra, veri gölündeki şifreleme ve anonimleştirme süreçlerinin sürekli olarak kontrol edilmesi gereklidir.

Güvenlik izleme sürecinde anormallik tespiti, veri gölündeki normal dışı aktivitelerin belirlenmesidir. Yetkisiz erişimler, veri ihlalleri ve veri sızıntılarının tespiti için anormallik tespit sistemleri kullanılmalıdır. Chavan ve Yalagi (2023), veri göllerinde makine öğrenimi temelli anormallik tespit yöntemlerinin, özellikle güvenlik tehditlerini erken aşamada belirlemedeki başarısına dikkat çekmişlerdir. Bu yaklaşımlar, veri ihlallerini ve yetkisiz erişimleri tespit ederek veri göllerinde güvenlik seviyesini önemli ölçüde artırma potansiyeline sahiptir. Veri gölünün uyumluluk denetimleri, ulusal ve uluslararası yasal düzenlemelere uygunluğun sağlanması açısından büyük önem taşımaktadır. Bu denetimler, veri gölünde saklanan kişisel verilerin yasalara uygun işlenip işlenmediğini kontrol etmekte ve potansiyel güvenlik risklerini tespit etmeye yardımcı olmaktadır (“Teradata”, t.y.). Kurumlar, bu denetimlerin yanı sıra iç denetim mekanizmaları geliştirerek veri güvenliğini artırabilirler. Bu mekanizmalar, sürekli izleme, detaylı kayıtların tutulması ve düzenli güvenlik değerlendirmeleri gibi uygulamaları içerebilmektedir (“Cloudian”, t.y.).

Veri gölünün güvenliği ve uyumluluğu, düzenli izleme ve denetim süreçleri ile sağlanmalıdır. Bu süreçlerin etkin bir şekilde uygulanması, veri ihlallerinin

önlenmesine yardımcı olurken yasal düzenlemelere uyum sağlamayı da güvence altına alacaktır.

5.4.5.3. Geri Bildirim ve Sürekli İyileştirme

Veri gölünün etkin bir şekilde yönetilebilmesi için kullanıcı geri bildirimlerinin dikkate alınması ve bu geri bildirimler doğrultusunda sistem iyileştirmelerinin yapılması önemlidir. Kullanıcılar, veri gölünü kullanırken karşılaştıkları sorunları raporlayabilir veya performans iyileştirmeleri konusunda önerilerde bulunabilirler. Bu geri bildirimler, veri gölünün zamanla gelişmesine ve kurumsal ihtiyaçlara daha iyi uyum sağlamasına katkıda bulunacaktır (Sawadogo ve Darmont, 2021). Veri gölünün performansını ve kullanılabilirliğini artırmak amacıyla düzenli olarak kullanıcı geri bildirimleri toplanmalıdır. Kurumlar; veri bilimcilerinden, iş analistlerinden veya diğer kullanıcılardan aldıkları geri bildirimlerle veri gölünün performansını ve erişim süreçlerini iyileştirmelidir. Bu geri bildirimler, veri gölündeki performans sorunlarının veya verimsiz süreçlerin belirlenmesine ve çözümlenmesine katkı sağlayacaktır (Giebler ve diğerleri, 2019).

Geri bildirimlere dayanarak veri gölünün performansı ve güvenliği sürekli iyileştirme süreçleri ile geliştirilmelidir. İyileştirme süreçleri, verilerin daha hızlı işlenmesini, daha güvenli hale getirilmesini ve kullanıcıların verilere daha kolay erişebilmesini sağlayacaktır. Kurumlar, veri gölünün sürdürülebilirliğini sağlamak için bu sürekli iyileştirme mekanizmalarını kullanmalıdır. Kullanıcı geri bildirimleri ve sürekli iyileştirme süreçleri, veri gölünün performansını artırarak sistemin verimli ve güvenli bir şekilde çalışmasına katkıda bulunacaktır.

5.4.6. Modelin Faydaları

Önerilen model, veri göllerinin kurumsal yapılara entegrasyonunu kolaylaştıran ve büyük veri yönetim süreçlerini optimize eden bir yapı sunmaktadır. Esneklik,

modelin öne çıkan avantajıdır ve farklı veri türlerini ve kaynaklarını kurumsal ihtiyaçlara göre uyarlayarak veri gölünün iş süreçlerine entegrasyonunu sağlar. Modelin bir diğer avantajı, sürekli gelişim imkânı sunmasıdır. Geri bildirim ve izleme süreçleriyle veri gölü zamanla daha verimli hale gelmektedir. Kullanıcı geri bildirimlerine dayalı iyileştirmeler, veri işleme ve analiz süreçlerini sürekli olarak geliştirmektedir.

Güvenlik ve uyumluluk stratejileri, modelin kritik bir parçasıdır. KVKK ve ISO 27001 gibi yasal düzenlemelere uyum sağlanarak veri güvenliği artırılmaktadır. Düzenli denetimler ve güvenlik önlemleri hem verilerin korunmasını hem de yasal gereksinimlerle uyumlu olmayı sağlamaktadır. Modelin en ileri seviyesinde, makine öğrenimi ve yapay zekâ teknikleriyle daha derin analizler yapılabilmektedir. Böylece kurumların stratejik karar alma süreçlerini güçlendiren tahmine dayalı analizlerle veri odaklı kararlar almalarına olanak tanınmaktadır.

Kısaca bu model, esnekliği, sürekli gelişimi, güvenliği ve ileri analiz yetenekleriyle Türkiye'deki kurumlara kapsamlı bir büyük veri çözümü sunmayı amaçlamaktadır.

SONUÇ VE ÖNERİLER

Bu çalışmada sunulan fonksiyonel ve olgunluk tabanlı veri gölü mimarisi ile veri yönetişimi stratejileri, Türkiye'deki kurumların büyük veri yönetim süreçlerini iyileştirmeye yönelik önemli fırsatlar sunmaktadır. Kurumların veri yönetimi ihtiyaçları göz önünde bulundurulduğunda önerilen model, veri toplama, analiz etme ve güvenli bir şekilde yürütme süreçlerinde önemli kolaylıklar sağlamaktadır. Özellikle büyük miktarda veri üreten ve bu verileri analiz etmeye ihtiyaç duyan sektörler için veri gölü mimarilerinin kullanımı, verinin değerini ortaya çıkarmak açısından kritik rol oynamaktadır (Sawadogo ve Darmont, 2021). Türkiye'deki TÜİK, Sağlık Bakanlığı ve TÜBİTAK gibi büyük veri toplayan kurumlar için bu mimari, veri işleme süreçlerinin daha etkin bir şekilde yönetilmesine olanak tanıyacaktır. TÜİK ve TÜBİTAK BİLGEM'in pilot bir proje kapsamında veri gölü tabanlı bir mimariyi uygulamaya koyması, bu kurumların veri göllerinin yenilikçi özelliklerini ve gelecekte sunabileceği potansiyeli stratejik bir çözüm olarak gördüğünü ortaya koymaktadır (B3LAB, 2020; TÜİK, 2020).

Fonksiyonel ve olgunluk tabanlı veri gölü mimarisi, Türkiye'deki kurumların mevcut veri mimarilerini kademeli olarak geliştirmelerini sağlayacaktır. Bu yaklaşım, kurumların veri toplama aşamasından başlayarak verileri temizleme, organize etme ve olgunlaştırma süreçlerini sistematik bir yaklaşımla ele almasına olanak tanımaktadır (Giebler ve diğerleri, 2019). Türkiye'deki birçok kurum büyük veri yönetim süreçlerinde henüz ham veri kullanımı ve veri işleme kapasitesini tam anlamıyla olgunlaştıramamıştır. Kurumlar henüz büyük verinin tüm potansiyelinden tam anlamıyla yararlanamasa da bu çalışmada önerilen mimari ile veri yönetimi kapasitesini adım adım geliştirebilir. Ayrıca üniversiteler ve diğer araştırma kurumları, veri göllerini kullanarak büyük veri kümelerini organize edebilir ve disiplinler arası çalışmalarını daha etkin hale getirebilirler. Önerilen olgunluk tabanlı yaklaşım, verilerin ham halde depolanmasından, işlenmiş ve analiz edilebilir veri setlerine dönüşmesine kadar olan süreci adım adım

geliştirmektedir. Bu sayede kurumlar, veri yönetişimi ve güvenlik gibi kritik konularda daha yüksek bir olgunluk seviyesine ulaşabilecektir.

Bu kapsamda yapılan değerlendirmeler, veri gölü mimarilerinin Türkiye'deki kamu kurumlarının mevcut veri yönetim altyapısını önemli ölçüde iyileştirebileceğini göstermektedir.

Mevcut sistemlerin büyük ölçüde veri ambarı tabanlı olması, veri akışlarının optimize edilmesini sınırlandırmakta ve veri gölü gibi yenilikçi teknolojilerin sunduğu avantajlardan tam anlamıyla faydalanılmasını engellemektedir. Veri gölleri, makine öğrenmesi tabanlı analizler gibi ileri düzey uygulamaları destekleyen esnek yapılarıyla bu kısıtların aşılmasına olanak tanımaktadır.

Önerilen model ile bu eksikliklerin giderilmesi adına uluslararası standartlarla uyumlu bir veri gölü mimarisi modeli ve bu modelin kurumsal yapılara entegrasyonuna yönelik stratejiler geliştirilmiştir.

Çalışmada sunulan üst veri yönetimi ve veri yönetişimi stratejileri, veri göllerinin veri yığınlarına dönüşmesini engellemek için kritik öneme sahiptir (Suriarachchi ve Plale, 2016). Üst veri yönetimi, verilerin izlenebilirliği ve erişim kontrolü açısından yasal düzenlemelere uyum sağlamaya yardımcı olmaktadır. Üst veri yönetimi sayesinde veriler daha düzenli bir şekilde sınıflandırılarak analiz süreçlerinde kullanılabilecektir.

Türkiye'de sağlık sektöründe dijitalleşme süreci hızlanmış olsa da mevcut sistemlerde veri gölü kullanılmaması, sağlık verilerinin etkin yönetimini sınırlamaktadır. Veri gölü mimarisine geçiş, yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerin bir arada depolanmasını ve işlenmesini mümkün kılarak, makine öğrenmesi tabanlı yenilikçi çözümler üretilmesine olanak sağlayarak sağlık verilerinin de daha etkili kullanılmasını sağlayacaktır. Mevcut veri ambarı tabanlı mimariler için üst veri yönetimi genellikle yeterli olmakla birlikte, veri göllerine geçişle birlikte işlenen veri kapasitesi ve türünün artması nedeniyle daha kapsamlı bir üst veri yönetimi stratejisine ihtiyaç duyulacaktır. Bu

yaklaşım, büyük çaplı veri setlerinin izlenebilirliğini ve erişim kontrolünü sağlayarak veri güvenliğini daha üst düzeyde destekleyecektir.

Veri gölleri, veri toplama, temizleme, organize etme ve analiz etme gibi süreçleri sistematik bir yapıya oturtarak, kamu kurumlarının veri yönetim süreçlerini optimize etmek için güçlü bir araç olarak öne çıkmaktadır. Ayrıca, üst veri yönetimi ve izlenebilirlik süreçlerinin iyileştirilmesi, verilerin güvenliğini artırmak ve yasal düzenlemelere uyumu sağlamak açısından kritik öneme sahiptir.

Önerilen modelin temellendirilmesi, kamu kurumlarının mevcut veri politikalarının değerlendirilmesi, literatürle karşılaştırmalı analizler yapılması ve ulusal ve uluslararası standartların göz önünde bulundurulması ile sağlanmıştır.

Veri güvenliği ve uyumluluk, kurumlar için büyük önem taşımaktadır. Önerilen mimari ve kapsamındaki güvenlik stratejileri, veri göllerinde saklanan verilerin güvenliğini sağlamak için gerekli olan şifreleme, erişim kontrolü ve anonimleştirme gibi teknikleri içermektedir. Veri göllerinin güvenlik süreçlerine uyarlanması, kurumların büyük miktardaki müşteri verisini korumasına yardımcı olacaktır. Özellikle KVKK gerekliliklerine uyum sağlanması, hassas verilerin korunması açısından zorunludur. Veri göllerine entegre edilen veri güvenliği stratejileri hem yasal düzenlemelere uygunluğu sağlamakta hem de kurumların veri ihlalleri ve siber saldırılara karşı güvenlik önlemlerini güçlendirmektedir (Madera ve Laurent, 2016).

Veri göllerinin sürdürülebilirliği için sürekli izleme ve geri bildirim mekanizmaları oldukça önemlidir. Bu mekanizmalar, veri gölünün performansını optimize etmek, güvenlik açıklarını tespit etmek ve kullanıcıların deneyimlerini iyileştirmek için kullanılabilir. Kurumlar, özellikle büyük veri analitiği ve iş zekâsı uygulamalarında bu tür izleme süreçlerini devreye alarak veri göllerini sürekli olarak geliştirebilirler. Büyük veri yöneten kurumlar, geri bildirim mekanizmalarını kullanarak verilerin kalitesini ve doğruluğunu artırabilirler. Performans izleme araçları sayesinde veri göllerinin yanıt süresi, veri işleme kapasitesi ve erişim süreleri sürekli

iyileştirilebilir. Bu süreç, kurumların daha verimli bir şekilde veri analitiği yapmasını sağlayacaktır.

Bu çalışmada, Türkiye'deki kurumlar için, fonksiyonel ve olgunluk tabanlı veri gölü mimarisi ile üst veri yönetimi ve veri yönetişimi stratejilerinin büyük veri yönetiminde etkin bir şekilde kullanılabilmesi için bir model önerisi sunulmuştur. Veri göllerinin başarılı bir şekilde uygulanabilmesi için veri toplama, işleme ve analiz süreçlerinin sürekli izlenmesi ve geri bildirimlerle geliştirilmesi gerektiği açıklanmıştır. Ayrıca veri göllerinin kurumsal iş süreçlerine entegrasyonunda yasal düzenlemelere uyum sağlanmasının ve veri güvenliğinin önemi vurgulanmıştır.

Yapılan değerlendirmeler, veri göllerinin kademeli olarak geliştirilmesi ve olgunlaşması ile kamu kurumlarının büyük veri analitiği kapasitelerinin artırılabileceğini ve bu süreçlerin sürdürülebilirlik temelinde desteklenmesi gerektiğini göstermiştir.

Yapılan araştırma sonucunda, veri göllerinin özellikle büyük veri yönetimi gereksinimi olan kurumlar için önemli fırsatlar taşıdığı ortaya çıkmıştır. Ancak, bu fırsatların hayata geçirilebilmesi için güvenlik, yönetişim ve uyumluluk gibi zorlukların dikkatle ele alınması gerekmektedir.

Uzun vadeli bir ulusal veri yönetimi sisteminin başarıyla hayata geçirilmesi için önerilen adımlar aşağıda özetlenmiştir:

- Tüm kurumları kapsayacak şekilde, ulusal bir veri arşivi vizyonu ve stratejik yol haritası belirlenmelidir.
- Veri yönetimi ve paylaşımı konularında ulusal düzeyde politika ve düzenlemeler oluşturulmalıdır.
- Veri formatları, üst veri yapıları ve sınıflandırma yöntemleri için ulusal standartlar geliştirilmelidir.
- Ulusal düzeyde güvenli, ölçeklenebilir ve sürdürülebilir bir veri altyapısı tasarlanmalıdır.

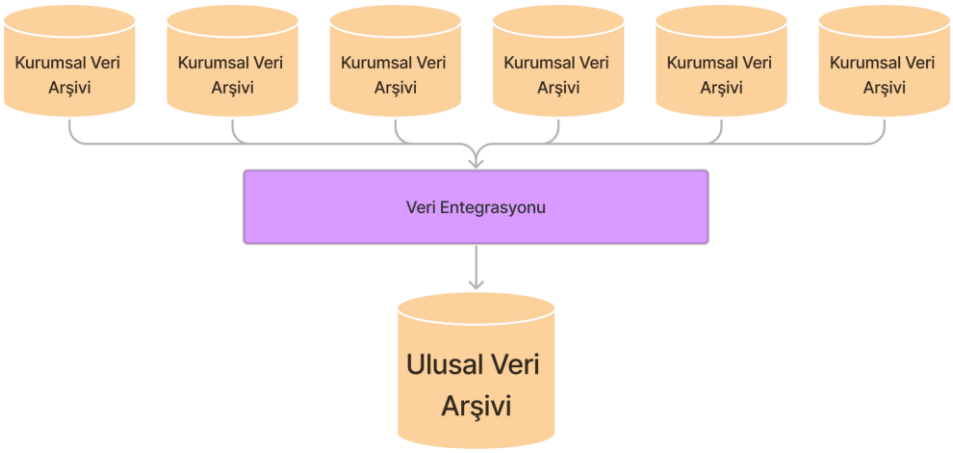
- Yüksek kapasiteli veri saklama ve işleme sistemleri (bulut tabanlı veya hibrit çözümler) kullanılmalıdır.
- Ulusal veri arşivine katkıda bulunacak tüm kurumlar için yasal uyumluluk standartları net bir şekilde tanımlanmalıdır.
- Kişisel verilerin korunması için KVKK ve diğer yasal düzenlemelere uygun güvenlik önlemleri alınmalıdır.
- Farklı kurumlar tarafından kullanılan veri yönetimi araçları arasında entegrasyon sağlanmalıdır.
- Mevcut veri gölü ve veri ambarı tabanlı altyapıların bir araya getirilmesi için teknik bir çözüm geliştirilmelidir.
- Sektörlere yönelik özel entegrasyon protokolleri geliştirilmelidir.
- Verilere erişim düzeyleri belirlenmeli; hangi kurumun veya kullanıcının hangi verilere hangi koşullarda erişebileceği düzenlenmelidir.
- Şifreleme, veri maskeleme ve erişim kontrol mekanizmaları yaygınlaştırılmalıdır.
- Veri arşivleri üzerinde makine öğrenmesi tabanlı uygulamaları destekleyecek araçlar ve altyapılar sağlanmalıdır.
- Büyük veri analitiği için veri işleme süreçleri optimize edilmelidir.
- Veri gölleri ve ulusal veri arşivinin performansını izlemek için kurumsal ve ulusal düzeyde sürekli geri bildirim mekanizmaları oluşturulmalıdır.
- Erişim süreleri, veri kalitesi ve kullanıcı memnuniyeti gibi metrikler düzenli olarak izlenmelidir.
- Kamuya açık veri setleri oluşturularak vatandaşların bu verilere erişimi sağlanmalı ve demokratik süreçlere katkı artırılmalıdır.
- Veri paylaşımı ve kullanım politikaları kamuoyuna açık ve anlaşılır bir şekilde sunulmalıdır.

Kurum düzeyinde geliştirilen bu modelin kurumların önemli bir kısmında hayata geçirilmesi sonrasında ulusal düzeyde bir arşivsel veri mimarisi modeli geliştirilmesi önerilmektedir. Bölüm 6.1'de bu öneri detaylarıyla açıklanmıştır.

ULUSAL ARŞİVSEL VERİ MİMARİSİ MODEL ÖNERİSİ

Bu çalışmadan sonra, her bir kurumun kendi süreçlerinde topladığı ve ürettiği arşivsel verilerin ulusal düzeyde sistematik olarak bir araya getirilmesi ile Ulusal Arşivsel Veri Mimarisi oluşturulması için bir çalışma yapılması önerilmektedir. Kurumların kullandığı bilgi ve belge sistemlerinin birbirinden farklı olması, kurumlar arası toplanan ve işlenen verilerin farklılıklar gösterilmesi, kurumların kullandığı veri yönetim, işleme ve analiz araçlarında farklılıklar olması sebepleri ile kurumlardan gelen arşivsel verilerin Ulusal Arşive aktarılması için sistematik veri entegrasyonunun ve veri yönetişimi stratejilerinin geliştirilmesi gerekecektir. Ancak bu çalışmada kullanılan veri arşivi modelinin kamu kurumlarında kullanılması durumunda her kurumun Arşivsel Veri Katmanı'na ulaşan veri belirli bir yapı ve standartta olduğu için daha basit bir entegrasyonun ardından Ulusal Arşiv'e aktarılabilecektir.

Şekil 33'te kurumsal veri arşivlerinin entegre edilerek Ulusal Veri Arşivine aktarılmasını gösteren model verilmiştir.



Şekil 33. Kurumsal veri arşivlerinin entegre edilerek Ulusal Veri Arşivi'ne aktarılması

Ulusal veri arşivi, bir ülkenin farklı kurumlarında toplanan verilerin merkezi bir yapı altında birleştirilerek uzun vadeli saklanması ve uygun verilerin erişime açılması

amacıyla oluşturulmuş kapsamlı bir veri saklama ve yönetim sistemi olarak tanımlanabilir. Ulusal veri arşivi, veri paylaşımını artırmak, araştırma ve analizleri kolaylaştırmak, karar alma süreçlerine katkıda bulunmak ve veri yönetiminde şeffaflık sağlamak vb. farklı amaçlara hizmet edebilir. Böylece aynı verilerin farklı kurumlar tarafından birden fazla kez toplanması gerekliliği ortadan kalkar, verimlilik sağlanır. Kurumlar, birbirlerinin verilerden yararlanabilir, ortak çalışmaları daha rahat gerçekleştirebilir. Örneğin Tarım ve Orman Bakanlığı ile Çevre, Şehircilik ve İklim Değişikliği Bakanlığı gibi kurumların verileri bir araya getirilerek, tarım ve çevre politikalarının oluşturulmasında etkili olabilir. Bu ortak çalışma su kaynakları ve tarımsal üretim ile ilgili kararların daha doğru verilmesini sağlayabilir. Farklı sektörlerden gelen verilerin analiz edilmesi, politika yapıcıların farklı bakış açılarına sahip olmasına ve daha bütünsel çözümler geliştirmesine yardımcı olur. Örneğin sağlık politikaları bölgesel sağlık ihtiyaçlarına göre optimize edilebilir. Kamuya açık veri setleri, vatandaşların devlet hizmetlerini değerlendirmesine, bilimsel çalışmalara katkıda bulunmasına ve sivil toplum kuruluşlarının çalışmalarını desteklemesine olanak tanır. Ayrıca kurumların işleyişi hakkında daha fazla şeffaflık sağlar ve demokratik süreçleri güçlendirir. Ulusal bir veri arşivi, yasal düzenlemelere ve veri yönetimi politikalarına uygunluğu kolaylaştırır. Merkezi bir yapıda yönetilen veriler, yasal düzenlemelere uygun olarak saklanır ve erişim izinleri kontrollü bir şekilde yönetilir. Ayrıca veri arşivi üzerinden yapılan tüm işlemler kayıt altına alınır böylece denetim süreçleri kolaylaşır.

Ulusal bir veri arşivi oluşturulurken, çeşitli sektörlerden gelen verilerin etkili bir şekilde toplanması, saklanması ve yönetilmesi için farklı bileşenler dikkate alınmalıdır. Bunların en önemlisi standartlaşma ve veri entegrasyonudur. Ulusal düzeyde bir veri arşivinin başarılı bir şekilde hayata geçirilebilmesi için, farklı kurumların verilerini tek bir yapıda birleştirirken veri standartlarının oluşturulması kritik önem taşır. Veri entegrasyonu, çeşitli kaynaklardan elde edilen verilerin standart bir yapıda toplanması, saklanması ve analiz edilebilir hale getirilmesidir. Ulusal veri arşivi kapsamında, bu sürecin güvenli, sürdürülebilir ve ölçeklenebilir bir şekilde yapılandırılması önem taşımaktadır. Farklı kurumlar genellikle

birbirinden farklı veri formatları, sınıflandırma yöntemleri ve arşivleme teknikleri kullanır. Bu durum, veri entegrasyonu sırasında uyumsuzluk sorunlarına yol açabilir. Örneğin bir eğitim kurumu öğrenci bilgilerini belirli bir formatta saklarken, sağlık kurumu hasta verilerini farklı bir yapıda saklayabilir. Kurumların tüm veri mimarilerini ulusal olarak tamamen uyumlu hale getirmek teorik olarak bu uyumsuzluk sorununu çözme potansiyeline sahip olsa da bu ölçekte bir entegrasyonun maliyeti astronomik boyutlara ulaşabilir ve uygulamada büyük zorluklarla karşılaşılabilir. Bu çalışmada sunulan veri gölü mimarisi modeli ve önerilen uygulama modeli benimsendiğinde, her kurumun ürettiği arşiv verisi belirli bir olgunluğa ve standarda erişmiş olacaktır. Ancak, ulusal bir veri arşivi oluşturmak için yine de bir entegrasyon süreci yürütülmelidir. Önerilen mevcut mimari sayesinde, bu entegrasyon çalışması verilerin olgun seviyeden arşiv katmanına geçişi sırasında ve arşiv katmanının içinde yürütülerek verimli bir entegrasyon sağlanabilir. Standartlaşma süreci, veri modelleri, dosya formatları, üst veriler ve veri sınıflandırma sistemleri gibi birçok unsuru içerir. Bu unsurların ulusal düzeyde tanımlanması, verilerin kurumlar arası uyumlu bir şekilde kullanılmasına olanak tanır. Örneğin, Türkiye'de farklı kamu kurumları arasında veri paylaşımını kolaylaştırmak amacıyla kullanılan e-Devlet entegrasyonu, bu tür standartlaşma ve veri entegrasyonu süreçlerine iyi bir örnektir.

Ulusal bir veri arşivi, büyük miktarda veriyi etkili bir şekilde saklayacak ve yönetecek güçlü bir teknik altyapıya ihtiyaç duyar. Bu altyapı, verilerin güvenli bir şekilde saklanmasını, yedeklenmesini ve gelecekte erişim ihtiyaçlarına yanıt verecek şekilde düzenlenmesini sağlamalıdır. Ayrıca verilerin hangi kurumlar tarafından ne şekilde kullanılacağına dair düzenlemeler de bu sistemler aracılığıyla uygulanabilir. Örneğin Avrupa’da bulunan European Data Archive (EDA) (Strollo ve diğerleri, 2021) gibi uluslararası veri arşivleri, bulut tabanlı sistemler kullanarak farklı ülkelerden gelen verileri merkezi bir yapıda bir araya getirmektedir. Bu yapı, sadece verilerin güvenli saklanmasını değil, aynı zamanda verilerin geniş kullanıcı kitlelerine erişilebilir olmasını da sağlayacaktır. Bilimsel araştırmalarda büyük veri setlerine erişim, özellikle veri analitiği, makine öğrenimi ve yapay zekâ gibi alanlarda büyük önem taşır. Ulusal veri arşivleri,

araştırmacıların güvenilir veri kaynaklarına erişimini sağlayarak bilimsel çalışmaların kalitesini ve kapsamını artırır.

Farklı kurumlardan gelen verilerin ulusal düzeyde bir arşivde birleştirilmesi, güvenlik ve gizlilik sorunlarını beraberinde getirir. Bu nedenle, ulusal veri arşivi oluşturulurken güçlü güvenlik önlemleri alınmalıdır. Kişisel verilerin korunması, yetkisiz erişimlerin önlenmesi, veri bütünlüğünün korunması gibi güvenlik unsurları, veri arşivinin sürdürülebilirliği için oldukça önemlidir. Bu nedenle verileri korumak için şifreleme, veri maskeleme, erişim kontrolü ve yedekleme vb. yöntemler kullanılmalıdır. Ayrıca veri koruma politikalarının ulusal düzeyde düzenlenmesi ve denetlenmesi gerekmektedir. Bu güvenlik tedbirleri, kullanıcıların veriye erişim izinleriyle ilgili net kurallar koyar ve böylece veri ihlali durumlarının önüne geçmek hedeflenir. Erişim politikaları, verilerin hangi kurumlar, araştırmacılar veya vatandaşlar tarafından hangi koşullarda kullanılabileceğini belirler. Farklı veri setleri için farklı erişim seviyeleri tanımlanabilir. Örneğin ABD'deki Ulusal Arşivler ve Belgeler İdaresi (National Archives and Records Administration - NARA), farklı kamu verilerine bireylerin erişim sağlamasına izin verirken, hassas veya gizli veriler için sadece yetkili kurumlarla erişimi sınırlandırır (*National Archives and Records Administration*, 2007). Bu politikalar, aynı zamanda veri paylaşımını da düzenler. Farklı kurumlar arası veri paylaşımı süreçlerinde, verilerin gizliliğini ve güvenliğini koruyacak prosedürler uygulanmalıdır.

Ulusal bir veri arşivi, ülkedeki veri yönetimi süreçlerini optimize ederek, uzun vadeli bilgi yönetimi stratejilerine katkı sağlayacaktır. Farklı kurumlardan gelen verilerin birleştirilmesiyle oluşturulacak bu yapı hem veri paylaşımını artıracak hem de verilerin güvenli ve düzenli bir şekilde saklanmasını sağlayacaktır. Farklı kurumların veri arşivlerini bir araya getirerek oluşturulan bir ulusal arşiv, veri paylaşımı, karar alma süreçleri, ekonomik gelişim ve bilimsel araştırmalara katkı sağlayacaktır. Ancak büyük ölçekli bir ulusal veri arşivi için önemli finansal ve teknik kaynaklar gereklidir. Yüksek kapasiteli sunucular, bulut altyapıları, veri güvenliği çözümleri gibi yüksek maliyetli yatırımlar yapılmalıdır. Ancak bu yapı

oluşturulduğunda daha sonraki veri yönetim süreçlerini merkezi bir yapıya oturtarak maliyet tasarrufu sağlayacaktır. Ayrıca verilerin düzenli bir şekilde sınıflandırılması ve güvenli bir ortamda saklanması, yönetim sürecini daha verimli hale getirerek özellikle kriz durumlarında hızlı karar alma süreçlerini destekleyecektir. Örneğin doğal afetler, salgınlar veya ekonomik krizlerde, farklı kurumların verilerine hızla erişim sağlanarak, en doğru kararların alınması mümkün olacaktır.

Sonuç olarak, ulusal veri arşivi oluşturma süreci, verilerin standardizasyonu, güvenliği, teknik altyapıların seçimi ve verilerin işlenmesi gibi birçok bileşeni içermektedir. Bu ölçekte bir arşivin oluşturulması verilerin analiz edilmesi, farklı kurumlar arasında güvenli bir şekilde paylaşılması ve uzun vadeli olarak sürdürülebilir bir yapı kurulması açısından kritik öneme sahiptir. Verilerin ulusal ölçekte etkili bir şekilde kullanılması ve kamu hizmetlerinde iyileştirmeler sağlanması için bu konunun çalışılması önerilmektedir.

KAYNAKÇA

10 Years of Hadoop and its Israeli Pioneering Researchers. (2016, 18 Kasım). Erişim adresi: https://developer.yahoo.com/blogs/153336735536/

Abadi, D., Ailamaki, A., Andersen, D., Bailis, P., Balazinska, M., Bernstein, P., ... ve Zaharia, M. (2020). The Seattle report on database research. ACM SIGMOD Record, 48(4), 44-53.

About schema.org. (t.y.). Erişim adresi: https://schema.org/docs/about.html

Adeyeye, O. J. ve Akanbi, I. (2024). A review of data-driven decision making in Engineering Management. *Engineering Science & Technology Journal*, 5(4), 1303-1324.

Advisera. (t.y.). What is ISO 27001? A quick and easy explanation. Erişim adresi: https://advisera.com/27001academy/what-is-iso-27001/

Akmon, D., Zimmerman, A., Daniels, M. ve Hedstrom, M. (2011). The application of archival concepts to a data-intensive environment: Working with scientists to understand data management and preservation needs. *Archival Science*, 11(3-4), 329-348.

Al, U. (2002). İnternet'te veri güvenliği. Erişim adresi: http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/10049/datasecurity.pdf?sequence=1&isAllowed=y

Alhadad, S. S. J. (2018). A framework for using learning analytics to support multiple learning theories. *Journal of Learning Analytics*, *5*(1), 55–73.

Alharthi, A., Krotov, V. ve Bowman, M. (2017). Addressing barriers to big data. *Business Horizons,* 60(3), 285-292.

Al-Kassab, J., Ouertani, Z. M., Schiuma, G. ve Neely, A. (2014). Information visualization to support management decisions. *International Journal of Information Technology & Decision Making*, *13*(2), 407–428. Erişim adresi: https://doi.org/10.1142/S0219622014500497

Alonso, O., Strötgen, J., Baeza-Yates, R. A. ve Geriz, M. (2011). Temporal information retrieval: Challenges and opportunities. *Twaw,* 11, 1-8.

Altun, T., Şahin, F. ve Öztaş, N. (2017). Kamu politikalarının belirlenmesi ve uygulanmasında büyük veri. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *22* (Kayfor 15 Özel Sayısı), 2021-2044.

Amundsen-io. (2024). Amundsen - Data discovery and metadata engine. GitHub. Erişim adresi: https://github.com/amundsen-io/amundsen

Apache Hadoop. (t.y.). Modules. Erişim adresi: https://hadoop.apache.org/

Aperta - Türkiye Açık Arşivi. (t.y.). Aperta - Türkiye Açık Arşivi hakkında. Erişim adresi: https://aperta.ulakbim.gov.tr/about

Ardagna, D., Barbierato, E., Gianniti, E., Gribaudo, M., Pinto, T. B. M., da Silva, A. P. C. ve Almeida, J. (2020). Predicting the performance of big data applications on the cloud. *The Journal of Supercomputing*, 77 (2021): 1321-1353.

Arısoy, Y. E. (2018). *Türkiye'de elektronik belge yönetiminde milli arşiv politikalarının geliştirilmesi* (Doktora tezi). Ulusal Tez Merkezi veri tabanından erişildi (502747).

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... ve Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-58.

Atlan. (2024a). Metadata: Definition, examples, benefits & use cases. Erişim adresi: https://atlan.com/what-is-metadata-to-test/

Atlan. (2024b). Why is metadata important for effective data management? Erişim adresi: https://atlan.com/why-is-metadata-important/

Attard, J., Orlandi, F., Scerri, S. ve Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly.* 32 (4), 399–418.

Australian Bureau of Statistics. (t.y.). Data. Erişim adresi: https://www.abs.gov.au/statistics/understanding-statistics/statistical-terms-and-concepts/data

Australian Data Archive. (t.y.). About us. Erişim adresi: https://ada.edu.au/about-ada/

AWS (t.y.). What is a data lake? - Introduction to data lakes and analytics. Erişim adresi: https://aws.amazon.com/what-is/data-lake/

Axelos. (2019). *ITIL Foundation: ITIL 4 Edition*. The Stationery Office.

Ayvaz, S. ve Salman, Y. B. (2020). Türkiye'de büyük veri kullanım olgunluğunun belirlenmesi. *Bilişim Teknolojileri Dergisi, 13*(3), 297-310.

Azmir, A. F. ve Wijayanti, L. (2022). Cloud computing opportunities and challenges in electronic document management. *Record and Library Journal*.

B3LAB - Bulut Bilişim ve Büyük Veri Araştırma Laboratuvarı. (t.y.). TÜİK büyük veri ileri analitik projesi. Erişim adresi: https://b3lab.org/sayfa/tuik_buyuk_veri_ileri_analitik_projesi-10

Baker, R. S. ve Inventado, P. S. (2014). Educational data mining and learning analytics. J. A. Larusson ve B. White (Yay. haz.). *Learning analytics: From research to practice* içinde (s. 61-75). Springer.

Baldin, A. ve Eliseev, D. (2020). *Processing archive information in digital university*. ITM Web of Conferences Konferansında sunulan bildiri.

Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. *Government Information Quarterly*, 31(3), 388-395.

Battle, L. ve Scheidegger, C. (2020). Urban data visualization: Research challenges and opportunities. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 385–395.

Beagrie, N. (2003). *National digital preservation initiatives: An overview of developments in Australia, the Netherlands, and the United Kingdom and of related international activity.* Council on Library and Information Resources, Washington, DC.; Library of Congress, Washington, DC. Erişim adresi: https://files.eric.ed.gov/fulltext/ED481029.pdf

Bell, D. (2018). Breaking the wall – building an infrastructure to enable multi-disciplinary analyses for social sciences and the Internet of Things. I*ASSIST 2018: Once upon a data point: Sustaining our data storytellers.* UK Data Archive, University of Essex. Erişim adresi: https://dam.ukdataservice.ac.uk/media/621251/breakingthewall_30may2018.pdf

Benjelloun, S., El Aissi, M. E. M., Loukili, Y., Lakhrissi, Y., Elhaj Ben Ali, S., Chougrad, H. ve El Boushaki, A. (2020). *Big data processing: Batch-based processing and stream-based processing*. Fourth International Conference On Intelligent Computing in Data Sciences (ICDS) Konferansında sunulan bildiri, 1-6.

Benson, P. (2008). *ISO 8000 the international standard for data quality*. The MIT 2008 Information Quality Industry Symposium Konferansında sunulan bildiri.

Berends, J., Carrara W ve Radu, C. (2020). *Analytical report 9: The economic benefits of open data*. Luxembourg: Publications Office of the European Union.

Bertino, E. (2016). *Data security and privacy: Concepts, approaches, and research directions.* 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC) Konferansında sunulan bildiri, Vol. 1, 400-407.

Bhimanpallewar, R. N., Jangid, S. K., Yadav, R. K., Arunkumar, D. T., Dehankar, P. ve Geetha, B. (2023). *Automating big data analysis using artificial intelligence.* International Conference on Smart Generation Computing, Communication and Networking Konferansında sunulan bildiri.

Bian, H. ve Ailamaki, A. (2022). *Pixels: An efficient column store for cloud data lakes*. 2022 IEEE 38th International Conference on Data Engineering (ICDE) Konferansında sunulan bildiri, 3078-3090. Erişim adresi: https://doi.org/10.1109/icde53745.2022.00276

Borges, M., Bernardino, J. ve Pedrosa, I. (2021). *Data-driven decision making strategies applied to marketing*. 16th Iberian Conference on Information Systems and Technologies (CISTI) Konferansında sunulan bildiri, 1-7.

Briney, K. A., Coates, H. L. ve Goben, A. H. (2020). Foundational practices of research data management. *Research Ideas and Outcomes*.

Brynjolfsson, E., Hitt, L. M. ve Kim, H. H. (2011). *Strength in numbers: How does data-driven decisionmaking affect firm performance?.*

Budapest Open Access Initiative. (2002). Read the Budapest Open Access Initiative. Erişim adresi: https://www.budapestopenaccessinitiative.org/read

Bulutistan. (t.y.). Bulut bilişimin sağladığı avantajlar nelerdir?. Erişim adresi: https://bulutistan.com/blog/bulut-bilisimin-sagladigi-avantajlar-nelerdir/

Buyya, R., Srirama, S. N., Casale, G., Calheiros, R., Simmhan, Y., Varghese, B., ... ve Toosi, A. N. (2018). A manifesto for future generation cloud computing: Research directions for the next decade. ACM Computing Surveys, 51(5), 1-38.

Cartlidge, A., Hanna, A., Rudd, C., Macfarlane, I., Windebank, J. ve Rance, S. (2007). *An introductory overview of ITIL® V3*. The UK Office of Government Commerce.

CDC Public Health Law. (2024). Health insurance portability and accountability act of 1996 (HIPAA). Erişim adresi: https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html

CERN. (t.y.). Zenodo (version 4.x). Erişim adresi: https://github.com/zenodo/zenodo

Chavan, V. D. ve Yalagi, P. S. (2023, April). *A review of machine learning tools and techniques for anomaly detection.* International Conference on Information and Communication Technology for Intelligent Systems Konferansında sunulan bildiri, 395-406.

Chessell, M., Scheepers, F., Nguyen, N., van Kessel, R. ve van der Starre, R. (2014). Governing and managing big data for analytics and decision makers. IBM.

Child & Family Data Archive. (t.y.). About us. Erişim adresi: https://www.childandfamilydataarchive.org/cfda/pages/cfda/about.html

Chinosi, M. ve Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards & Interfaces*, 34(1), 124-134.

CKAN Metadata. (t.y.). Erişim adresi: https://ckan.org/features/metadata

CKAN Overview. (t.y.). Erişim adresi: https://docs.ckan.org/en/2.11/

CKAN The World’s Leading Open Source Data Management System. (t.y.). Erişim adresi: https://ckan.org/

Cloudian. (t.y.). Data lake security: Challenges and 6 critical best practices. Erişim adresi: https://cloudian.com/guides/data-lake/data-lake-security-challenges-and-6-critical-best-practices/

CNRI. (2023). The handle system. Erişim adresi: https://www.handle.net

Codit Teknoloji. (2024). 2024 Bulut bilişim trendleri nelerdir?. Erişim adresi: https://codit.com.tr/blog/2024-bulut-bilisim-trendleri-nelerdir

Cointelegraph. (2024). Dağıtık sistemler nedir ve nasıl çalışır?. TradingView. Erişim adresi: https://tr.tradingview.com/news/cointelegraph:190faf50cd9e8:0/

Colavizza, G., Blanke, T., Jeurgens, C. ve Noordegraaf, J. (2021). Archives and AI: An overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage*, 1 - 15.

Collectiv. (t.y.). Change management: The other half of an enterprise data strategy. Erişim adresi: https://gocollectiv.com/blog/enteprise-data-strategy-change-management/

Collibra. (t.y.). Collibra data intelligence: Data governance and catalog solutions. Erişim adresi: https://www.collibra.com

Connolly, T. ve Begg, C. (2015). *Database systems: A practical approach to design, implementation, and management (Sixth edition, global edition).* Pearson.

Consultative Committee for Space Data Systems. (2012). Reference model for an open archival information system (OAIS). Erişim adresi: https://public.ccsds.org

Creating Value Through Open Data. (2015). Study on the impact of re-use of public data resources, Europen Commission, European Data Portal.

CRL Center for Research Libraries. (t.y.). TRAC metrics. Erişim adresi: https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. Erişim adresi: https://snap.stanford.edu/soma2010/papers/soma2010_16.pdf

Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2020). Bilgi ve iletişim güvenliği rehberi. Erişim adresi: https://cbddo.gov.tr/SharedFolderServer/Genel/File/bg_rehber.pdf

Çalım, A. Ş. ve Tuna, G. (2022). Büyük veri tabanlı arşivleme yönetim sistemi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 15(2), 161-170.

Çiftçi, M. ve Gülsoy, S. (2020). Endüstri 4.0 uygulamaları için büyük verinin yönetimi: Teknik bir bakış. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 3(1), 36–45.

DAMA International. (2017). DAMA-DMBOK2: Data management body of knowledge. Technics Publications.

Dang, T. K. ve Anh, T. D. (2021). An effective and elastic blockchain-based provenance preserving solution for the open data. *International Journal of Web Information Systems*, 17, 480-515.

Data Archiving and Networked Services. (t.y.). DANS. Erişim adresi: https://dans.knaw.nl/en/about/

Data Catalog Vocabulary (DCAT) - Version 3. (2024, 22 Ağustos). Erişim adresi: https://www.w3.org/TR/vocab-dcat/#toc

DataHub. (2024). DataHub: The #1 open source metaveri platform. Erişim adresi: https://datahubproject.io

DataShop@CMU. (t.y.). About / Frequently asked questions (FAQ). Erişim adresi: https://pslcdatashop.web.cmu.edu/about/faq.html

Dataverse Project. (t.y.). About. Erişim adresi: https://dataverse.org/about

DATAVERSITY. (t.y.). Data lake strategy: Its benefits, challenges, and implementation. Erişim adresi: https://www.dataversity.net/data-lake-strategy-its-benefits-challenges-and-implementation/

Date, C. J. (2004*). An introduction to database systems*. Addison-Wesley.

Davenport, T. H. ve Bean, R. (2018). Big companies are embracing analytics, but most still don't have a data-driven culture. *Harvard Business Review Digital Articles*, 2-4.

DCC, Because good research needs good data. (2009). PAIMAS. Erişim adresi: https://www.dcc.ac.uk/guidance/standards/diffuse/show?standard_id=154

Dekkers, M., Polman, F., Te Velde, R. ve De Vries, M. (2006). MEPSIR: Measuring European public sector information resources, Final report of study on exploitation of public sector information–benchmarking of EU framework conditions*.*

Deloitte. (2013). Market assessment of public sector information. Londra: Department for Business Innovation & Skills.

Demelius, L., Kern, R. ve Trügler, A. (2023). Recent advances of differential privacy in centralized deep learning: A systematic survey. ArXiv.

Deming, W. E. (1986). Out of the Crisis. MIT Press.

Dennis, A., Wixom, B. ve Tegarden, D. (2015). Systems analysis and design: An object-oriented approach with UML. John wiley & sons.

Devlet Arşiv Hizmetleri Hakkında Yönetmelik (2019, 18 Ekim). Resmî Gazete (Sayı: 30922). Erişim adresi: https://www.devletarsivleri.gov.tr/varliklar/dosyalar/mevzuat/arsivhizmetleri.pdf

Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi (2018, 16 Temmuz). Resmî Gazete (Sayı: 20480). Erişim adresi: https://www.devletarsivleri.gov.tr/varliklar/dosyalar/mevzuat/19.5.11.pdf

Diamantini, C., Potena, D. ve Storti, E. (2024). Analytic processing in data lakes: A semantic query-driven discovery approach. *Information Systems Frontiers*, 1-19.

Dijital Arşiv - Sayısal Arşiv Sistemi. (2015, 21 Nisan). Devlet Arşivleri Genel Müdürlüğü Dijital Arşiv Yönetmeliği. Erişim adresi: https://sayisalarsiv.com/devlet-arsivleri-genel-mudurlugu-arsiv-yonetmeligi

Dixon, J. (2010, October 14). Pentaho, hadoop, and data lakes. *James Dixon's Blog*. Erişim adresi: https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/

DMP Tool. (2023). Data management general guidance. Erişim adresi: https://dmptool.org/general_guidance#types-of-data

Documentation of data.europa.eu. (t.y.). Who we are. Erişim adresi: https://dataeuropa.gitlab.io/data-provider-manual/

Doğan, K. ve Arslantekin, S. (2016). Büyük veri: Önemi, yapısı ve günümüzdeki durum. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi,* 56(1).

Dong, W., Liao, S. ve Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. *Journal of Management Information System*s, 35(2), 461–487.

DPC Digital Preservation Coalition. (t.y.). Digital preservation handbook, audit and certification. Erişim adresi: https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification

Dryad. (t.y.). Who we are. Erişim adresi: https://datadryad.org/stash/about

Dublin Core. (t.y.a). DCMI history. Erişim adresi: https://www.dublincore.org/about/history/

Dublin Core Metadata Initiative. (2020). DCMI metadata terms. Erişim adresi: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Duval, E. (2001). Metadata standards: What, who & why. *Journal of Universal Computer Science,* 7(7), 591-601.

Dzikria, I., Narulita, L. F., Hermanto, A. ve Kusnanto, G. (2022). *ISO 15489 attributes prioritization in electronic document management system of the first level healthcare facilities.* 2022 Seventh International Conference on Informatics and Computing (ICIC) Konferansında sunulan bildiri.

E-Devlet Kapısı Kamu Uygulamaları Merkezi. (t.y.). Kamu Uygulamaları Merkezi'ne hoş geldiniz. Erişim adresi: https://kamu.turkiye.gov.tr/

El Assouri, A. (2024, May 2). Data lake revolution in higher education. Stalks. Erişim adresi: https://www.stalks-app.com/en/2024/05/02/data-lake-revolution-in-higher-education/

Elmasri, R. ve Navathe, S. B. (2016). *Fundamentals of database systems*. Pearson.

Enfobilişim. (t.y.). Dijital arşiv yönetimi nedir ve nasıl yapılır? https://www.enfobilisim.com/blog/dijital-arsiv-yonetimi-nedir-ve-nasil-yapilir

Erbaş, E. P. (2018). *Karamanoğlu Mehmetbey Üniversitesi Türk Dili ve Edebiyatı Bölümü ile Bakü Devlet Üniversitesi Azerbaycan Dili ve Edebiyatı Bölümü öğrencilerinin okuma alışkanlıkları üzerine karşılaştırmalı bir inceleme* (Yüksek lisans tezi). Yükseköğretim Kurulu Ulusal Tez Merkezi veri tabanından erişildi.

Eroğlu, Ş. ve Külcü, Ö. (2014). TS 13298 çerçevesinde kurumsal bilgi sistemleri ve elektronik belge yönetimi standartlarının değerlendirilmesi: İçişleri Bakanlığı örneği. *Bilgi Dünyası*, 15(2), 327-352. Erişim adresi: https://doi.org/10.15612/BD.2014.439

Esds. (2021). Introduction to the concept of data lake and its benefits. Erişim adresi: https://www.esds.co.in/blog/introduction-to-the-concept-of-data-lake-and-its-benefits/

Esri ArcGIS Pro. (t.y.). Create ISO 19115 and ISO 19139 metadata. Erişim adresi: https://pro.arcgis.com/en/pro-app/latest/help/metadata/create-iso-19115-and-iso-19139-metadata.htm

Estuary. (t.y.). Real-time data lake. Erişim adresi: https://estuary.dev/real-time-data-lake/

European Data Portal. (2016). Warszawski Ninja application. Erişim adresi: https://data.europa.eu/sites/default/files/use-cases/use_case_poland_-ninja.pdf

Eurostat. (t.y.a). Structural business statistics - Quality. Erişim adresi: https://ec.europa.eu/eurostat/web/structural-business-statistics/quality

Eurostat. (t.y.b). Who we are. Erişim adresi: https://ec.europa.eu/eurostat/en/web/main/about-us/who-we-are

Eyüpoğlu, C. (2018). *Büyük veride etkin gizlilik koruması için yazılım tasarımı* (Doktora tezi). Ulusal Tez Merkezi veri tabanından erişildi (Tez no: 512702).

Fang, H. (2015). *Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem.* 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems Konferansında sunulan bildiri. Erişim adresi: https://doi.org/10.1109/CYBER.2015.7288049

Fayyad, U.M., Piatetsky-Shapiro, G. ve Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining.* Cambridge, MA: MIT Press.

Ferenstein, G. (2013). We must choose privacy or medical breakthroughs: Statisticians ID anonymous study participants. Erişim adresi: https://techcrunch.com/2013/01/18/we-must-choose-privacy-or-medical-breakthroughs/

FGDC.gov. Federal Geographic Data Committee. (t.y.). Geospatial metadata. Erişim adresi: https://www.fgdc.gov/metadata/

Ficek, J., Wang, W., Chen, H., Dagne, G. ve Daley, E. (2021). Differential privacy in health research: A scoping review. *Journal of the American Medical Informatics Association: JAMIA.*

Figshare. (t.y.). Erişim adresi: https://knowledge.figshare.com/about

Finnish Social and Health Data Permit Authority Findata. (t.y.). Finnish social and health data permit authority findata. Erişim adresi: https://findata.fi/en/

Flower Framework (t.y.a.). Differential privacy. Erişim adresi: https://flower.ai/docs/framework/explanation-differential-privacy.html

Flower Framework (t.y.b.). What is federated learning? Erişim adresi: https://flower.ai/docs/framework/tutorial-series-what-is-federated-learning.html

GDPR.EU. (t.y.).Complete guide to GDPR compliance. Erişim adresi: https://gdpr.eu/

Gelir İdaresi Başkanlığı. (t.y.). Hizmetlerimiz - GİB teknoloji. Erişim adresi: https://teknoloji.gib.gov.tr/teknoloji/hizmetlerimiz.html

GeoNetwork Opensource. (t.y.). Geographic information -- Metadata (iso19115-3.2018). Erişim adresi: https://docs.geonetwork-opensource.org/3.12/annexes/standards/iso19115-3.2018/

Gharehchopogh, F. S. ve Khalifelu, Z. A. (2011). *Analysis and evaluation of unstructured data: Text mining versus natural language processing*. International Conference on Application of Information and Communication Technologies (AICT) Konferansında sunulan bildiri.

Ghasemaghaei, M. (2019). Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management*, 57, 102055.

Giebler, C., Gröger, C., Hoos, E., Eichler, R., Schwarz, H. ve Mitschang, B. (2021). *The data lake architecture framework: A foundation for building a comprehensive data lake architecture.* 19. Fachtagung für Datenbanksysteme für Business, Technologie und Web (BTW 2021) Konferansında sunulan bildiri, 351-370. Gesellschaft für Informatik, Bonn.

Giebler, C., Gröger, C., Hoos, E., Schwarz, H. ve Mitschang, B. (2019). Leveraging the data lake: Current state and challenges. T. Welzer, J. Eder, V. Podgorelec ve A. K. Latific (Yay. haz.). *Big data analytics and knowledge discovery*, *21st International Conference, DaWaK 2019*, Konferansında sunulan bildiri, 179-188. Springer. Erişim adresi: https://doi.org/10.1007/978-3-030-27520-4_13

Giebler, C., Gröger, C., Hoos, E., Schwarz, H. ve Mitschang, B. (2020). *A zone reference model for enterprise-grade data lake management*. IEEE 24th

International Enterprise Distributed Object Computing Conference (EDOC) Konferansında sunulan bildiri, 57-66.

Gökalp, E., Gökalp, M. O., Çoban, S. ve Eren, P. E. (2018a). Analysing opportunities and challenges of integrated blockchain technologies in healthcare. S. Wrycza ve J. Maślankowski (Yay. haz.). *Information systems: research, development, applications, education. SIGSAND/PLAIS 2018. Lecture Notes in Business Information Processing,* vol 333, içinde (s. 399-416). Springer, Cham. Erişim adresi: https://doi.org/10.1007/978-3-030-00060-8_13

Gökalp, M. O., Kayabay, K., Çoban, S., Yandık, Y. B. ve Eren, P. E. (2018b). *Büyük veri çağında işletmelerde veri bilimi.* 5th International Management Information Systems Conference Konferansında sunulan bildiri, 94-97. Ankara.

Grosser, T., Bloeme, J., Mack, M. ve Vitsenko, J. (2016). Hadoop and data lakes: Use cases, benefits and limitations. *Business Application Research Center (BARC GmbH)*.

Gtech. (t.y.). Veri ambarı ve veri gölü. Erişim adresi: https://www.gtech.com.tr/veri-ambari-ve-veri-golu

Gupta, B. B., Yamaguchi, S. ve Agrawal, D. P. (2017). Advances in security and privacy of multimedia big data in mobile and cloud computing. *Multimedia Tools and Applications*, 76(21), 22391-22398.

Gurajada, S., Sattler, K., Tinnefeld, C. ve Höfner, P. (2017). Data pools vs. data warehouses: *Efficient data management in the age of big data*. *Journal of Big Data Management*, 3(2), 45-57.

Haberer, B. ve Schnurr, D. (2020). Open government data for digital services: Effects on innovation, competition and societal benefits. *Social Science Research Network (SSRN)*.

Hai, R., Geisler, S. ve Quix, C. (2016). *Constance: An intelligent data lake system.* 2016 International Conference on Management of Data Konferansında sunulan bildiri.

Hai, R., Quix, C. ve Jarke, M. (2021). Data lake concept and systems: A survey. *ArXiv*. Erişim adresi: https://www.researchgate.net/publication/352506022_Data_lake_concept_and_systems_a_survey

Halevi, G. ve Moed, H. (2012). The evolution of big data as a research and scientific topic: Overview of the literature. *Research Trends,* 30(1), 3-6.

Halevy, A., Ferrer, C. C., Ma, H., Ozertem, U., Pantel, P., Saeidi, M., Silvestri, F. ve Stoyanov, V. (2022). Preserving integrity in online social networks. Communications of the ACM, 65(2), 92-98.

Halevy, A., Korn, F., Noy, N. F., Polyzotis, N., Roy, S. ve Whang, S. E. (2016). Goods: Organizing Google's datasets. *SIGMOD*, 795-806.

Han, J., Kamber, M. ve Pei, J. (2012). *Data mining concepts and techniques.* Morgan Kaufmann.

Harris, R. (2012). ICSU and the challenges of big data in science. *Research Trends*, 30(1), 11-12.

Hazen, B. T., Kung, L., Cegielski, C. G. ve Jones-Farmer, L. A. (2014). Performance expectancy and use of enterprise architecture: Training as an intervention. *Journal of Enterprise Information Management.*

Hernandez, M. A. ve Roberts, C. (2022). Data architecture: A comprehensive guide to data management and strategy. *Journal of Information Systems*, 36(4), 45-67. Erişim adresi: https://doi.org/10.1016/j.jis.2022.01.003

Hiatt, J. M. (2006). ADKAR: A model for change in business, government and our community. Prosci.

Hitachi Vantara. (t.y.). What are the different types of data?. Erişim adresi: https://www.hitachivantara.com/en-us/insights/faq/what-are-different-data-types.html

Holom, R. M., Rafetseder, K., Kritzinger, S. ve Sehrschön, H. (2020). Metadata management in a big data infrastructure. *Procedia Manufacturing,* 42, 375-382.

Hossain, Q., Yasmin, F., Biswas, T. R. ve Asha, N. B. (2024). Data-driven business strategies: A comparative analysis of data science techniques in decision-making. *Scholars Journal of Economics, Business and Management*, 9, 257-263.

Hotz, N. (2024a). What is TDSP?. Erişim adresi: https://www.datascience-pm.com/tdsp/

Hotz, N. (2024b). OSEMN data science life cycle. Erişim adresi: https://www.datascience-pm.com/osemn/

Hu, V. C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R. ve Scarfone, K. (2015). Guide to attribute based access control (ABAC) definition and considerations. NIST Special Publication, 800(162), 1-54.

IBM Documentation. (2021). Data pool. Erişim adresi: https://www.ibm.com/docs/en/imdm/11.5?topic=gds-data-pool

ICA. (t.y.). Discover ICA, our mission and objectives. Erişim adresi: https://www.ica.org/discover-ica/our-mission-our-objectives/

ICPSR. (t.y.a). About ICPSR. Erişim adresi: https://www.icpsr.umich.edu/web/pages/about/

ICPSR. (t.y.b). Data management & curation. Erişim adresi: https://www.icpsr.umich.edu/web/pages/datamanagement/index.html

IDC Türkiye. (2024). The future of IT: Rethinking digitalization for an AI everywhere world. Erişim adresi: https://www.idc.com/mea/events/71258-idc-turkiye-cio-summit-2024

Informatica. (t.y.). Informatica: Data integration and data management solutions. Erişim adresi: https://www.informatica.com

Inmon, B. (2016). *Data lake architecture: Designing the data lake and avoiding the garbage dump.* Technics Publications.

Inmon, W. H. (2005). *Building the data warehouse (Fourth edition)*. Wiley Publishing.

Inmon, W. H. ve Linstedt, D. (2014). *Data architecture: A primer for the data scientist: Big data, data warehouse and data vault*. Morgan Kaufmann.

International DOI Foundation. (2022). The DOI® system. Erişim adresi: https://doi.org/

International Open Data Best Practices. (2017, Ocak). Initiativa aporta. Erişim adresi: https://datos.gob.es/sites/default/files/doc/file/international_open_data_best_practices.pdf

ISACA. (t.y.a). COBIT 5: A business framework for the governance and management of enterprise IT. Erişim adresi: https://www.isaca.org/resources/cobit

ISACA. (t.y.b). COBIT 2019 framework: Introduction and methodology. Erişim adresi: https://www.isaca.org/resources/cobit

Islam, M. T., Borovica-Gajic, R. ve Karunasekera, S. (2022*). A multi-level caching architecture for stateful stream computation*. 16th ACM International Conference on Distributed and Event-Based Systems Konferansında sunulan bildiri.

ISMS.online. (2024). ISO 27701, The Privacy Information Management Standard. Erişim adresi: https://www.isms.online/iso-27701/

ISO 8000-1 (2022) Data quality - Part 1: Overview. Erişim afresi: https://www.iso.org/standard/81745.html

ISO 16175 (2020). Information and documentation – Principles and functional requirements for records in electronic office environments. Erişim adresi: https://www.iso.org/standard/74294.html

ISO 23081-1 (2017). Information and documentation - records management processes - metadata for records. Second edition. Erişim adresi: https://www.sis.se/api/document/preview/922676/

ISO 30300 (2011). Information and documentation – Management systems for records – Fundamentals and vocabulary. Erişim adresi: https://www.iso.org/standard/74291.html

ISO 30301 (2011). Information and documentation – Management systems for records – Requirements. Erişim adresi: https://www.iso.org/standard/74292.html

ISO 30302 (2020) Information and documentation – Management systems for records - Guidelines for implementation. Erişim adresi: https://www.iso.org/standard/81595.html

ISO Online Browsing Platform (OBP). (2014). ISO 19115-1:2014(en) Geographic information - Metadata - Part 1: Fundamentals. Erişim adresi: https://www.iso.org/obp/ui/en/#iso:std:iso:19115:-1:ed-1:v1:en

ISO/DIS 16363 (2012). Space data and information transfer systems – Audit and certification of trustworthy digital repositories. Erişim adresi: https://www.iso.org/standard/87472.html

ISO/IEC TR 10032 (2003). Information technology — Reference model of data management. Erişim adresi: https://www.iso.org/standard/38607.html

ISO/IEC 11179-1 (2023). Information technology — Metadata registries (MDR) — Part 1: Framework. Erişim adresi: https://www.iso.org/standard/78914.html

ISO/IEC 19510 (2013) Information technology — Object management group business process model and notation. Erişim adresi: https://www.iso.org/standard/62652.html

ISO/IEC 19583-1 (2019). Information technology — Concepts and usage of metadata. Erişim adresi: https://www.iso.org/standard/67365.html

ISO/IEC 19763 (2020) Information technology – Metamodel framework for interoperability (MFI). Erişim adresi: https://www.iso.org/standard/84749.html

ISO/IEC 19773 (2011). Information technology - Metadata registries (MDR) modules. Erişim adresi: https://www.iso.org/standard/41769.html

ISO/IEC 20000 (2019). IT service management – A practical guide. Erişim adresi: https://www.iso.org/publication/PUB100441.html

ISO/IEC 27002 (2022) Information security, cybersecurity and privacy protection - Information security controls. Erişim adresi: https://www.iso.org/standard/75652.html

ISO/IEC 27014 (2020). Information security, cybersecurity and privacy protection - Governance of information security. Erişim adresi: https://www.iso.org/standard/74046.html

ISO/IEC 27701 (2019). Security techniques - Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management - Requirements and guidelines. Erişim adresi: https://www.iso.org/standard/71670.html

ISO/IEC 29100 (2024) Information technology - Security techniques - Privacy framework. Erişim adresi: https://www.iso.org/standard/85938.html

ISO/IEC 38500 (2015): Information technology - Governance of IT for the organization. Erişim adresi: https://www.iso.org/standard/81684.html

ISO/IEC 42001 (2023) Information technology — Artificial intelligence — Management system. Erişim adresi: https://www.iso.org/standard/81230.html

ISO/IEC/IEEE 42010 (2022). Software, systems and enterprise — Architecture description. Erişim adresi: https://www.iso.org/standard/74393.html

ISO/IEC/IEEE 42010. (t.y.). ISO/IEC/IEEE 42010 website. Erişim adresi: https://www.iso.org/standard/74393.html

ISO/IEC JTC 1/SC 42 (2017) Artificial intelligence. Erişim adresi: https://www.iso.org/committee/6794475.html

ISO/TC 46/SC 11. (t.y.). Management systems for records (ISO 30300 series). Erişim adresi: https://committee.iso.org/sites/tc46sc11/home/projects/published/management-systems-for-records-i.html

Jia, L. (2023). *Development and application of smart marketing systems based on big data technology*. International Conference on Artificial Intelligence, Systems and Network Security Konferansında sunulan bildiri.

Jiang, K. ve Zheng, Y. (2013, Aralık). *Mining Twitter data for potential drug effects.* Advanced Data Mining and Applications Konferansında sunulan bildiri, 434-443. Springer.

Jisc Archives Hub. (t.y.). ISAD(G). Erişim adresi: https://archiveshub.jisc.ac.uk/isadg/

John, T. ve Misra, P. (2017). *Data lake for enterprises: Lambda architecture for building enterprise data systems*. Packt Publishing.

Jörg, B. (t.y.). *CERIF: The common European research information format model.* Erişim adresi: https://www.dfki.de/fileadmin/user_upload/import/4319_CERIF_The_Common_European_Research_Information_Format_Model_DSJ.pdf

Kahn, R. E. ve Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries*, 6(2), 115–123. Erişim adresi: https://doi.org/10.1007/s00799-004-0109-8

Kaisler, S., Armour, F., Espinosa, J. A. ve Money, W. (2013). *Big data: Issues and challenges moving forward.* 46th Hawaii International Conference on System Sciences IEEE Konferansında sunulan bildiri, 995-1004.

Kamalakkannan, S., Yasmin, A., Arunkumar, R. ve Kavitha, P. (2023). *A model for the analytical performance of data lake in stock market analysis with Databricks Delta Lak*e. 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS) Konferansında sunulan bildiri, 1065-1071.

Kaptan, S. (1995). *Bilimsel araştırma ve istatistik teknikleri.* Ankara: Gazi Üniversitesi.

Karaarslan, E., ve Akbaş, M. F. (2017). Blok zinciri tabanlı siber güvenlik sistemleri. *Uluslararası Bilgi Güvenliği Mühendisliği Dergisi*, 3(2), 16-21.

Karkošková, S. (2023). *Architectural design of metadata management tool*. 9th International Conference on Control, Decision and Information Technologies (CoDIT) Konferansında sunulan bildiri, 65-70.

Kassen, M. (2019). Open data and e-government–related or competing ecosystems: A paradox of open government and promise of civic engagement in Estonia. *Information Technology for Development,* 25(3), 552-578.

Katuu, S. (2018). The utility of enterprise architecture to records and archives specialists. *2018 IEEE International Conference on Big Data (Big Data)* (s. 2702-2710) içinde. IEEE.

Kawashita, I., Baptista, A. A. ve Soares, D. (2022). *Open government data use by the public sector - An overview of its benefits, barriers, drivers, and enablers.* 2022 Hawaii International Conference on System Sciences (HICSS) Konferansında sunulan bildiri, 1-10.

Kayabay, K., Gökalp, M. O., Akyol, M. A., Eren, P. E. ve Koçyiğit, A. (2016). Geleceğin kuruluşları için büyük veri: Mevcut durum ve eğilimler. Erişim adresi: https://www.researchgate.net/profile/Kerem-Kayabay/publication/309635166_Big_Data_for_Future_Enterprises_Current_State_and_Trends/links/581afda608aeffb29414489b/Big-Data-for-Future-Enterprises-Current-State-and-Trends.pdf

Khan, N., Uddin, M. F. ve Gupta, N. (2014). Seven v's of big data understanding big data to extract value. *American Society for Engineering Education*, 1-6.

Khurshid, M., Zakaria, N., Rashid, A., Ahmad, M., Arfeen, M. ve Shehzad, H. M. F. (2020). Modeling of open government data for public sector organizations using the potential theories and determinants – A systematic review. *Informatics*, 7(3), 24.

Kimball, R. ve Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons.

Kini, S. ve Pai, K. (2024). Exploring real-time data processing using big data frameworks. *Communications on Applied Nonlinear Analysis*.

Kişisel Verileri Koruma Kurumu. (2023). 6698 sayılı Kişisel Verilerin Korunması Kanunu.

Kişisel Verilerin Korunması Kanunu. (2016). T.C. Resmî Gazete (Sayı: 29677). Erişim adresi: https://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf

Koo, J., Kang, G. ve Kim, Y. G. (2020). Security and privacy in big data life cycle: A survey and open challenges. *Sustainability*.

Kutlutürk, L. ve Özdemirci, F. (2023). Elektronik belge yönetim sistemleri ve kullanılabilirlik. *Bilgi Dünyası*, 15(1), 102-124.

Külcü, Ö. (2009). Elektronik belge yönetim sistemleri ve organizasyonel uyum. *Türk Kütüphaneciliği Dergisi.*

Külcü, Ö. (2018). *Kurumsal bilgi sistemleri ve belge yönetimi: Organizasyonlarda bilgi ve belge yönetiminin temel ilkeleri*. Hiperyayın.

Külcü, Ö., Çakmak, T. ve Özel, N. (2015). *Kamusal bilgi ve elektronik belge yönetimi: Organizasyonlar ve üniversitelere yönelik koşulların analizi.* Türk Kütüphaneciler Derneği.

KVKK Kişisel Verileri Koruma Kurumu. (t.y.). Veri güvenliğine ilişkin yükümlülükler. Erişim adresi: https://www.kvkk.gov.tr/Icerik/2040/Veri-Guvenligine-Iliskin-Yukumlulukler

Ladley, J. (2020). *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press.

LaPlante, A. ve Sharma, B. (2016). *Architecting data lakes data management architectures for advanced business use cases.* O'Reilly Media Inc.

Laufs, D., Peters, M. ve Schultz, C. (2022). Data platforms for open life sciences – A systematic analysis of management instruments. *PLOS ONE*, 17.

Learn Azure. (2024). The team data science process lifecycle. Erişim adresi: https://learn.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle

LeCun, Y., Bengio, Y. ve Hinton, G. (2015). Deep learning. *Nature,* 521(7553), 436-444. Erişim adresi: https://doi.org/10.1038/nature14539

Lee, C. A. (2010). Open archival information system (OAIS) reference model. *Encyclopedia of Library and Information Sciences,* Third Edition, 4020-4030. doi: 10.1081/E-ELIS4-120044377.

Leetaru, K. H. (2012). A big data approach to the humanities, arts, and social sciences. *Research Trends,* 30(1), 17-30.

Li, H. ve Zhang, G. (2022). *Optimization of an archives information system based on intelligent information sorting algorithm*. Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) Konferansında sunulan bildiri, 977-980.

Li, Y. F., Kennedy, G., Ngoran, F., Wu, P. ve Hunter, J. (2013). An ontology-centric architecture for extensible scientific data management systems. *Future Generations Computer Systems*, 29(3), 641–653.

Lin, Z., Yin, Y., Liu, L. ve Wang, D. (2023). SciSciNet: A large-scale open data lake for the science of science research. *Scientific Data*, 10(1), 315. Erişim adresi: https://doi.org/10.1038/s41597-023-02198-9

Losee, R. (2006). Browsing mixed structured and unstructured data. *Information Processing and Management*, 42(2), 440–452.

Ma, J., Abie, H., Skramstad, T. ve Nygård, M. (2011). Development and validation of requirements for evidential value for assessing trustworthiness of digital records over time. *Journal of Information.*

Madera, C. ve Laurent, A. (2016). The next information architecture evolution: The data lake wave. *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)* Konferansında sunulan bildiri, 191-197. Erişim adresi: https://doi.org/10.5220/0005749801910197

MapReduce vs. Pig vs. Hive. (2022, 18 Şubat). Erişim adresi: https://www.projectpro.io/article/mapreduce-vs-pig-vs-hive/163#:~:text=Hadoop%20MapReduce%20is%20a%20compiled,compared%20to%20Pig%20and%20Hive.

MARC standards. (2006). Frequently asked questions. Erişim adresi: https://www.loc.gov/marc/faq.html#definition

Marine Data Archive. (t.y.). Introduction. Erişim adresi: https://marinedataarchive.org/introduction.php

Martins, B. ve Calado, P. (2010, February). Learning to rank for geographic information retrieval. *Proceedings of the 6th workshop on geographic information retrieval* içinde (s. 21).

Mathis, J. (2017). Data lakes and data governance: Implementing the data lake. *TDWI Whitepaper*. Transforming Data with Intelligence.

Metin, O. ve Ünal, Ş. (2022). İçerik analizi tekniği: İletişim bilimlerinde ve sosyolojide doktora tezlerinde kullanımı. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 22(Özel Sayı 2), 273-294.

Michael, C. I., Ipede, O. J., Adejumo, A. D., Adenekan, I. O., Adebayo, D., Ojo, A. S. ve Ayodele, P. A. (2024). Data-driven decision making in IT: Leveraging AI and data science for business intelligence. *World Journal of Advanced Research and Reviews* 23(01), 432-439.

Mikroarea Digital. (2024). Bulut depolama avantajları ve dezavantajları. Erişim adresi: https://mikroarea.com.tr/blog/bulut-depolama-avantajlari-ve-dezavantajlari

Milić, P., Veljković, N. ve Stoimenov, L. (2021a). Comparative analysis of metadata models on e-government open data platforms. *IEEE Transactions on Emerging Topics in Computing*, 9(1), 119-130.

Milić, P., Veljković, N. ve Stoimenov, L. (2021b). Comparative analysis of metadata models for open government data. *Electronic Government, an International Journal*, 17(1), 118-141.

Miloslavskaya, N. ve Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science, 88*, 300-305.

Mishra, S ve Misra, A. (2017). *Structured and unstructured big data analytics*. IEEE International Conference on Current Trends in Computer, Electrical, Electronics and Communication Konferansında sunulan bildiri.

Mitchell, W. (1996). Architectural archives in the digital era. *The American Archivist,* 59(2), 200-204.

Mohammed, A. G., Eram, A. ve Talburt, J. R. (2017). I*SO 8000-61 Data Quality Management Standard, TDQM compliance, IQ principles.* MIT International Conference on Information Quality Konferansında sunulan bildiri.

Monte Carlo. (t.y.). 5 data lake examples. Monte Carlo data. Erişim adresi: https://www.montecarlodata.com/blog-5-data-lake-examples/

Nargesian, F., Zhu, E., Miller, R. J., Pu, K. ve Arocena, P. C. (2019). Data lake management: Challenges and opportunities. *Proceedings of the VLDB Endowment*, 12(12), 1986–1989.

NASA Earthdata. (t.y.). ISO 19115 Geographic metadata information. Erişim adresi: https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/iso-19115

National Archives and Records Administration. (2007). *Trustworthy repositories audit & certification: Criteria and checklist.* ABD: OCLC ve CRL. Erişim adresi: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/trac_0.pdf

National Law Review. (2021). Thinking beyond the law: What is ISO 29100 privacy framework. Erişim adresi: https://www.natlawreview.com/article/thinking-beyond-law-what-iso-29100-privacy-framework

NCAR Research Data Archive. (t.y.). About the research data archive. Erişim adresi: https://rda.ucar.edu/support/about-the-rda/

NDA. (t.y.). Welcome to the NIMH data archive. Erişim adresi: https://nda.nih.gov/

Nelson, B. ve Reuben, J. (2019). SoK: Chasing accuracy and privacy, and catching both in differentially private histogram publication. *Transactions on Data Privacy,* 13, 201–245.

NII Inter-University Research Institute Corporation Research Organization of Information and Systems. (t.y.). Mission & Strategies. Erişim adresi: https://www.nii.ac.jp/en/about/introduction/mission/

Nikiforova, A. ve McBride, K. (2021). Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*, 58, 101539.

Object Management Group. (t.y.). Business process model and notation (BPMN) Version 2.0. Erişim adresi: https://www.omg.org/spec/BPMN/2.0/

Open Archives Initiative. (t.y.). OAI-PMH. Erişim adresi: https://www.openarchives.org/pmh/

Open Data Driving Growth, Ingenuity and Innovation. (2012). London: The Creative Studio at Deloitte. Erişim adresi:

https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

OpenAIRE (t.y.). Research data management handbook. Erişim adresi: https://www.openaire.eu/research-data-management-handbook

Opendata.swiss. (t.y.). Portal. Erişim adresi: https://opendata.swiss/en/about

Oracle Help Center. (t.y.). Veri havuzları. Erişim adresi: https://docs.oracle.com/cloud/help/tr/content-cloud/CECDA/GUID-33E7F38E-E5F9-4D32-8077-A47876096FAE.htm#

Oracle Türkiye. (t.y.). Veritabanı nedir?. Erişim adresi: https://www.oracle.com/tr/database/what-is-database/

Özdemirci, F. (2001). Belge üretimi ve kurumsal bilgi yönetimi. T. Fenerci ve O. Gürdal (Yay. haz.) *21. Yüzyıla Girerken Enformasyon Olgusu Sempozyumu, 19-20 Nisan 2001, Hatay*, *Bidiriler* içinde (s. 1-9). Ankara: Türk Kütüphaneciler Derneği.

Özdemirci, F. (2004). Bir disiplin olarak belge yönetimi. S. Arslantekin ve F. Özdemirci (Yay. haz.). *Kütüphaneciliğin Destanı Uluslararası Sempozyumu, 21-24 Ekim 2004, Ankara, Bildiriler* içinde (s. 191-199). Ankara: A.Ü. DTCF Bilgi ve Belge Yönetimi Bölümü.

Özdemirci, F. (2019a). Milli e-arşiv bilgi sistemi ağı ve veri merkezi yapılanma önerisi: Yenilikçi teknolojiler-Yeni nesil arşivciler-Yapay zekâ ve ötesi…. *Bilgi Yönetimi*, 2(2), 169-176.

Özdemirci, F. (2019b). Kurumlar için EBYS ve e-arşiv sistemi idari yapılanma ve yönetim süreci: Bileşenler ve entegrasyonlar. B. Yalçınkaya, M. A. Ünal, B. Yılmaz ve F. Özdemirci. (Ed.) *Bilgi Yönetimi ve Bilgi Güvenliği: -eBelge-eArşiv-eDevlet-Bulut Bilişim-Büyük Veri-Yapay Zekâ.* içinde (s. 3-9). Ankara: Ankara Üniversitesi BİL-BEM.

Özdemirci, F. (2019c). “4. e-BEYAS Sempozyumu açış konuşması- Prof. Dr. Fahrettin Özdemirci”. Erişim adresi: http://2019.ebeyas.org/wp-content/uploads/2019/11/4_e_BEYAS-2019-Semp_Acis_konusmasi_FO_10_10_2019.pdf.

Özdemirci, F., Torunlar, M. ve Saraç, S. (2009). Belge yönetimi ve arşiv sistemi (BEYAS) terimler ve tanımlar sözlüğü. *Üniversiteler İçin Belge Yönetimi ve Arşiv Sistemi/İşlemleri (BEYAS) El Kitabı (Aralık 2009)* içinde (s. 353-360). Ankara.

Özkan, Ö. (Ed.). (2019). *Açık veri - Hukuk, düzenlemeler ve kamu ilişkileri çalışma grubu raporu.* Türkiye Bilişim Vakfı, Blockchain Türkiye Platformu. Erişim adresi: https://bctr.org/dokumanlar/Acik_Veri.pdf

Park, J., Kim, H. ve Lee, K. (2021). Visual data analytics for supporting healthcare decisions. *Journal of Data and Information Science*, *6*(2), 104–115. Erişim adresi: https://doi.org/10.2478/jdis-2021-0010

Pattanaik, S. N. ve Wiegand, R. P. (2021). *Data visualization*. Handbook of Human Factors and Ergonomics (5th ed.) içinde. Wiley

Pepper, J., Senin, A., Jebbia, D., Breen, D. E. ve Greenberg, J. (2022). *Metadata verification: A workflow for computational archival science*. IEEE International Conference on Big Data (Big Data) Konferansında sunulan bildiri, 2565-2571.

Pishgoo, B., Azirani, A. A. ve Raahemi, B. (2021). A hybrid distributed batch-stream processing approach for anomaly detection. *Information Sciences*, 543, 309–327.

Pisoni, G., Molnár, B. ve Tarcsi, Á. (2021). Data science for finance: Best-suited methods and enterprise architectures. *Applied System Innovation*.

Plume Labs. (t.y.). Erişim adresi: https://plumelabs.com/en/

Prabhugouda, A. ve Asra, S. (2024). A review on big data applications and their challenges. *Journal of Information & Knowledge Management*, 2430001.

PREMIS Editorial Committee (2015). Premis data dictionary for preservation metadata, version 3.0. Erişim adresi: https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

Prokhorenkov, D. (2022). *Anonymization level and compliance for differential privacy: A systematic literature review.* 2022 International Wireless Communications and Mobile Computing (IWCMC) Konferansında sunulan bildiri, 1119–1124.

Prosci. (2020). Best practices in change management. *Prosci Research Library*.

Qi, Q. ve Tao, F. (2018). Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access,* 6, 3585-3593.

Quix, C., Hai, R. ve Vatov, I. (2016). Metadata extraction and management in data lakes with GEMMS. *Complex Systems Informatics and Modeling Quarterly*, 9, 67–83.

QUT Manual of Policies and Procedures. (2013). Management of research data and primary materials. Erişim adresi: https://www.mopp.qut.edu.au/D/D_02_08.jsp

Raghupathi, W. ve Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3. Erişim adresi: https://doi.org/10.1186/2047-2501-2-3

Rahmanto, K. N. ve Riasetiawan, M. (2018). *Data preservation process in big data environment using open archival information system*. International Conference on Smart Technology and Community (ICSTC) Konferansında sunulan bildiri, 1-6. IEEE. Erişim adresi: https://doi.org/10.1109/ICSTC.2018.8528669

Rahul, K. ve Banyal, R. K. (2020). Data life cycle management in big data analytics. *Procedia Computer Science*, 173, 364-371.

Recker, J. (2010). Opportunities and constraints: The current struggle with BPMN. *Business Process Management Journal*, 16(1), 181-201.

Research Center for Open Science and Data Platform. (t.y.). Overview of the NII research data cloud. Erişim adresi: https://rcos.nii.ac.jp/en/service/

Resmî Gazete. (2016, 7 Nisan). Kişisel Verilerin Korunması Kanunu. Erişim adresi: https://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf

Resmî Gazete. (2018a, Temmuz 13). Devlet Arşivleri Başkanlığı Hakkında Cumhurbaşkanlığı Kararnamesi (Kararname No: 11). *Resmî Gazete*, 30473. https://www.resmigazete.gov.tr/eskiler/2018/07/20180716.htm

Resmî Gazete. (2018b, Temmuz 16). 11 numaralı Cumhurbaşkanlığı Kararnamesi. Erişim adresi: https://www.resmigazete.gov.tr/eskiler/2018/07/20180716-1.pdf

Resmî Yazışmalarda Uygulanacak Usul ve Esaslar Hakkında Yönetmelik. (2020, 10 Haziran). *Resmî Gazete* (Sayı: 31151). Erişim adresi: https://www.mevzuat.gov.tr/MevzuatMetin/21.5.2646.pdf

SafetyCulture. (t.y.). A quick guide to ISO 15489. Erişim adresi: https://safetyculture.com/topics/iso-15489/

Sajida, S. ve Ramakrishna, S. (2015). A study of extract-transform-load (ETL) processes. *International Journal of Engineering Research & Technology,* 3(18), 1-6.

Sak, R., Şahin Sak, İ. T., Öneren Şendil, Ç. ve Nas, E. (2021). Bir araştırma yöntemi olarak doküman analizi. *Kocaeli Üniversitesi Eğitim Dergisi*, 4(1), 227-250. Erişim adresi: http://doi.org/10.33400/kuje.843306

Sawadogo, P. ve Darmont, J. (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems,* 56(1), 97-120. doi: 10.1007/s10844-020-00608-7

Sawadogo, P., Kibata, T. ve Darmont, J. (2019). *Metadata management for textual documents in data lakes*. 21st International Conference on Enterprise Information Systems (ICEIS 2019) Konferansında sunulan bildiri, Heraklion, Crete-Greece.

Sdxcentral. (t.y.). What Is FISMA? An overview of the law. Erişim adresi: https://www.sdxcentral.com/security/definitions/data-security-regulations/what-is-fisma-an-overview-of-the-law/

Sever, H. ve Oğuz, B. (2002). Veri tabanlarında bilgi keşfine formel bir yaklaşım kısım 1: Eşleştirme sorguları ve algoritmalar. *Bilgi Dünyası*, 3(2), 173-204.

Sharma, U. (2024). OSEMN framework for data science. Erişim adresi: https://www.linkedin.com/pulse/osemn-framework-data-science-pronounced-awesome-udit-sharma-26blc/

Shi, W., Cao, J., Zhang, Q., Li, Y. ve Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal,* 3(5), 637-646.

Sivarajah, U., Kamal, M. M., Irani, Z. ve Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263-286.

Smart Global Governance (t.y.). ISO/IEC 29100. Erişim adresi: https://www.smartglobalgovernance.com/en/iso-iec-29100-2/

Snowflake. (t.y.). Cloud data lake. Erişim adresi: https://www.snowflake.com/guides/cloud-data-lake/

Sprehe, J. T. (2000). Integrating records management into information resources management in US government agencies. *Government Information Quarterly,* 17(1), 13-26.

Standart Dosya Planı ile İlgili 2005/7 Sayılı Başbakanlık Genelgesi (2005, 25 Mart). Resmî Gazete (Sayı: 25766). Erişim adresi: https://www.resmigazete.gov.tr/eskiler/2005/03/20050325-10.htm

Stein, B. ve Morrison, A. (2014). *Atlas: Metadata management for Hadoop*. 2014 International Conference on Big Data Konferansında sunulan bildiri.

Strollo, A., Cambaz, D., Clinton, J., Danecek, P., Evangelidis, C. P., Marmureanu, A., Ottemöller, L., Pedersen, H., Sleeman, R., Stammler, K., Armbruster, D., Bienkowski, J., Boukouras, K., Evans, P. L., Fares, M., Neagoe, C., Heimers, S., Heinloo, A., Hoffmann, M. ve Özener, H. (2021). EIDA: The European integrated data archive and service infrastructure within ORFEUS. *Seismological Research Letters*, 92(3). Erişim adresi: 1788-1795. https://doi.org/10.1785/0220200413

Suriarachchi, I. ve Plale, B. (2016). *Crossing analytics systems: A case for integrated provenance in data lakes*. 2016 IEEE International Conference on Big Data (Big Data) Konferansında sunulan bildiri. Erişim adresi: https://doi.org/10.1109/BigData.2016.7840676

Şeker, Ş. E. (2015). Büyük veri ve büyük veri yaşam döngüleri. *YBS Ansiklopedi*, 2(3), 10-17.

Şeker, Ş. E. (2018). CRISP-DM: Endüstriler arası standart işleme-Veri madenciliği için (Cross industry standard processing-Data mining). *YBS Ansiklopedi*, *5*(2). Erişim adresi: https://ybsansiklopedi.com/wp-content/uploads/2018/08/crispdm.pdf

Şeker, Ş. E. ve Erdoğan. D. (t.y.). *Knime ile uçtan uca veri bilimi.*

Tanenbaum, A. S. ve Van Steen, M. (2016). *Distributed systems: principles and paradigms.* Prentice-Hall.

Taylor, J. (2019). *ITIL Foundation: ITIL 4 edition*. The Stationery Office.

T.C. Cumhurbaşkanlığı. (2023). Gizlilik dereceli belgelerde uygulanacak usul ve esaslar hakkında yönetmelik. Erişim adresi: https://www.tccb.gov.tr/assets/dosya/resmiyazisma/dosyalar/gizlilik-yonetmeligi.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2019a). Veri sözlüğü oluşturma metodolojisi, ulusal veri sözlüğü sistemi. Erişim adresi: https://cbddo.gov.tr/SharedFolderServer/CMSFiles/VeriSozluguOlusturmaMetodolojisi.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2019b). E-yazışma teknik rehberi. Erişim adresi: https://bim.hku.edu.tr/wp-content/uploads/2019/09/e-yazisma-teknik-rehberi.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2020). Bilgi ve iletişim güvenliği rehberi. Erişim adresi: https://cbddo.gov.tr/SharedFolderServer/Genel/File/Bilgi_ve_Iletisim_Guvenligi_Rehberi.pdf

T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2024). Ulusal veri stratejisi. Erişim adresi: https://cbddo.gov.tr/ulusal-veri-stratejisi/

T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. (t.y.). Akıllı şehirler dönüşümü yol haritası. Erişim adresi: https://www.akillisehirler.gov.tr

T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü. (2024a, Temmuz 18). Veri ambarı ve büyük veri birimi. Erişim adresi: https://sbsgm.saglik.gov.tr/TR-32358/veri-ambari-ve-buyuk-veri-birimi.html

T.C. Sağlık Bakanlığı Sağlık Bilgi Sistemleri Genel Müdürlüğü. (2024b, Haziran 27). Büyük veri uygulamaları ve veri yönetimi koordinatörlüğü. Erişim adresi: https://sbsgm.saglik.gov.tr/TR-104968/buyuk-veri-uygulamalari-ve-veri-yonetimi-koordinatorlugu.html

T.C. Sanayi ve Teknoloji Bakanlığı. (t.y.). Hakkımızda. Erişim adresi: https://gbs.sanayi.gov.tr/

T.C. Ulaştırma Bakanlığı TUENA (1999). Türkiye ulusal enformasyon altyapısı anaplanı sonuç raporu. Ankara: TUENA Türkiye Ulusal Enformasyon Altyapısı Proje Ofisi, TÜBİTAK BİLTEN. Erişim adresi: http://www.bilgitoplumu.gov.tr/Documents/1/Yayinlar/991000_TuenaRapor.pdf

T.C. Ulaştırma ve Altyapı Bakanlığı. (2020, Şubat). 2019 yılı faaliyet raporu.

T.C. Ulaştırma ve Altyapı Bakanlığı (2024). Kamu Entegre Veri Merkezi Projesi. Erişim adresi: https://hgm.uab.gov.tr/kamu-entegre-veri-merkezi-projesi

Teradata. (t.y.). What is data lake security? Erişim adresi: https://www.teradata.com/insights/data-security/what-is-data-lake-security

Terrizzano, I., Schwarz, P., Roth, M. ve Colino, J. E. (2015). *Data wrangling: The challenging journey from the wild to the lake*. 7th Biennial Conference on Innovative Data Systems Research (CIDR'15) Konferansında sunulan bildiri.

The Federal Big Data Research and Development Strategic Plan. (2016). The networking and information technology research and development program. NITRD.

The Home of the U.S. Government's Open Data. (t.y.). About data.gov. Erişim adresi: https://data.gov/

The Library of Congress. (2022). Metadata object description schema (MODS). Erişim adresi: https://www.loc.gov/standards/mods/mods-overview.html

The Library of Congress. (t.y.). PREMIS, preservation metadata maintenance activity. Erişim adresi: https://www.loc.gov/standards/premis/

Tonta, Y. (2015). Açık bilim ve açık erişim. Erişim adresi: http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/11764/31-Tonta-235-250.pdf?sequence=1

Topcu, A. E. ve Işık, A. (2019). Açık devlet verisi süreçlerine dair bir model önerisi. *Avrupa Bilim ve Teknoloji Dergisi,* (17), 833-843.

Truică, C.-O., Apostol, E., Darmont, J. ve Pedersen, T. (2021). The forgotten document-oriented database management systems: An overview and benchmark of native XML DODBMSes in comparison with JSON DODBMSes. *ArXiv*, 30, abs/2102.02246.

TS ISO/IEC 11179-1 (2023). Bilgi teknolojisi - Metaveri kayıtları (MDR) - Bölüm 1: Çerçeve

TS ISO/IEC 19763 (2020). Bilgi teknolojileri - Birlikte çalışabilirlik için meta model çerçevesi (MFI). Erişim adresi: https://intweb.tse.org.tr/Standard/Standard/Standard.aspx?081118051115108051104119110104055047105102120088111043113104073087085056072075080107077112118115

TS ISO/IEC 27002 (2022). Bilgi teknolojisi - Güvenlik teknikleri - Bilgi güvenliği kontrolleri için uygulama kodu.

TS ISO/IEC 27014 (2022). Bilgi güvenliği, siber güvenlik ve gizlilik koruması - Bilgi güvenliği yönetimi

TSE K 523 (2016). Bilgi varlıklarının gizlilik derecelerine göre sınıflandırılması kriteri.

Tunks, J. (2024). 5 Data masking techniques and why you need them. Erişim adresi: https://pathlock.com/learn/5-data-masking-techniques-and-why-you-need-them/

Tupper, C. (2011). *Data architecture: From zen to reality*. Elsevier.

TÜBİTAK Açık Bilim Politikası. (t.y.). Erişim adresi: https://ulakbim.tubitak.gov.tr/sites/images/Ulakbim/tubitak_acik_bilim_politikasi-tr.pdf

TÜBİTAK ULAKBİM. (t.y.). Veri ambarı. Erişim adresi: https://ulakbim.tubitak.gov.tr/tr/hizmetlerimiz/veri-ambari-0

TÜİK. (2020). Büyük veri ileri analitik projesi. TÜİK proje belgeleri. Erişim adresi: http://www.sp.gov.tr/upload/xSPRapor/files/Jt8Ua+TUIK_21_PP.pdf

Türk Standartları Enstitüsü. (2024). Bilgi güvenliği yönetim sistemi (BGYS) belgelendirmesi. Erişim adresi: https://www.tse.org.tr/bilgi-guvenligi-yonetim-sistemi-bgys-belgelendirmesi-ts-iso-iec-27001/

Türkiye Cumhuriyeti Cumhurbaşkanlığı, Dijital Dönüşüm Ofisi. (2020). Bilgi ve iletişim güvenliği rehberi. Erişim adresi: https://cbddo.gov.tr/SharedFolderServer/Genel/File/bg_rehber.pdf

Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (t.y.a). AçıkVeri projesi. Erişim adresi: https://cbddo.gov.tr/projeler/acik-veri/

Türkiye Cumhuriyeti Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (t.y.b). Ulusal veri sözlüğü projesi. Erişim adresi: https://cbddo.gov.tr/projeler/ulusalverisozlugu/

Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı 100. Yıl Türkiye Planı. (2019). On birinci kalkınma planı (2019 - 2023). Erişim adresi: https://www.sbb.gov.tr/wp-content/uploads/2022/07/On_Birinci_Kalkinma_Plani-2019-2023.pdf

Türkiye İstatistik Kurumu. (t.y.a). Metaveri nedir?. Erişim adresi: https://www.tuik.gov.tr/Kurumsal/Meta_Veri_Nedir

Türkiye İstatistik Kurumu. (t.y.b). Mikro veri. Erişim adresi: https://www.tuik.gov.tr/Kurumsal/Mikro_Veri

Türkiye Yüzyılı. (2022). Kamu veri alanı. Erişim adresi: https://www.turkiyeyuzyili.com/proje-kamu-veri-alani

Türkyılmaz-van der Velden, Y. (2021). Research data management and FAIR data principles. *Creative Commons Türkiye Webinar Series*. Erişim adresi: https://creativecommons.org.tr/cctrwebinar-fair-veri-prensipleri-ve-arastirma-verilerinin-yonetimi/

UK Data Archive. (t.y.). About. Erişim adresi: https://www.data-archive.ac.uk/about/

Ularu, E. G., Puican, F. C., Apostu, A. ve Velicanu, M. (2012). Perspectives on big data and big data analytics. *Database Systems Journal,* 3(4), 3-14.

UltaHost. (2024). Bulut depolamanın avantajları ve dezavantajları nelerdir?. Erişim adresi: https://ultahost.com/blog/tr/bulut-depolamanin-avantajlari-ve-dezavantajlari/

United Nations Development Programme (UNDP). (2024). Data governance framework recommendation report for Türkiye. Erişim adresi: https://www.undp.org/sites/g/files/zskgke326/files/2024-04/data_governance_framework_recommendation_report_for_turkiye_2024-final-13_march.pdf

University of Nebraska Lincoln Center for Digital Research in the Humanities. (t.y.). What is EAD?. Erişim adresi: https://cdrh.unl.edu/articles/basicguide/EAD

Uslu, H. (2023). Dijital dönüşüm ve kamu hizmetleri yönetimde yenilikçi yaklaşımlar ve zorluklar. *Uluslararası Politik Araştırmalar Dergisi,* 9(3), 15-31. Erişim adresi: https://doi.org/10.25272/icps.1354693

U.S. Census Bureau. (2022). Building a modern, data-centric ecosystem. Erişim adresi: https://www.census.gov/newsroom/blogs/research-matters/2022/10/technology-transformation.html

Veri Paylaşımı Kurulu Yönetmeliği (2020, 8 Ağustos). *Resmî Gazete* (Sayı: 31207). Erişim adresi: https://www.resmigazete.gov.tr/eskiler/2020/08/20200808-3.htm

VERİM - Sabancı Üniversitesi Veri Analitiği Araştırma ve Uygulama Merkezi. (2024). Projeler. Erişim adresi: https://verim.sabanciuniv.edu/tr/projeler

Voigt, P. ve Von dem Bussche, A. (2017). The EU general data protection regulation (GDPR): A practical guide. Springer International Publishing.

Walker, C. ve Alrehamy, H. (2015). *Personal data lake with data gravity pull*. 2015 IEEE Fifth International Conference on Big Data and Cloud Computing (s. 160-167) Konferansında sunulan bildiri.

Watson, H. J. ve Wixom, B. H. (2007). The current state of business intelligence. *IEEE Computer*, 40(9), 96-99.

Weber, T., Kranzlmüller, D., Fromm, M. ve Tavares de Sousa, N. (2020). Using supervised learning to classify metadata of research data by field of study. *Quantitative Science Studies,* 1(3), 525-550.

Weibel, S. L. (2009). Dublin core metadata initiative: A personal history. M. J. Bates ve M. N. Maack (Ed.). *Encyclopedia of Library and Information Sciences*, 1655-1663. CRC Press.

Weller, M. (2011). The digital scholar: How technology is transforming scholarly practice. *A&C Black.*

Wieder, P. ve Nolte, H. (2022). Toward data lakes as central building blocks for data management and analysis. *Frontiers in Big Data.* doi: 10.3389/fdata.2022.945720

Yarımağan, Ü. (2016). *Veri tabanı sistemleri.* Akademi Yayıncılık.

Yaseen, H. ve Obaid, A. (2020). Big data: Definition, architecture & applications. *Journal of Online Innovation and Virtual Reality*, 4(1), 45-51.

Yıldız, M. ve Şeker, Ş. E. (2016). Veri madenciliği araçları (Data mining tools). *YBS Ansiklopedi,* 3(4).

Yılmaz, A. (2022). *Yapay zekâ*. İstanbul: KODLAB Yayın Dağıtım Yazılım ve Eğitim Hizmetleri San. ve Tic. Ltd. Şti..

Yılmaz, M. (2012). TS 13298 standardı ışığında elektronik belge yönetim sistemleri. *Türk Standartları Enstitüsü,* Ankara, 19. Erişim adresi: https://ab.org.tr/ab13/bildiri/71.pdf

Younas, M. (2019). Research challenges of big data. *Service Oriented Computing and Applications*, 1-3.

Yusof, Z. M. ve Chell, R. W. (2002). Towards a theoretical construct for records management*. Records Management Journal* 12(2), 55-64.

Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., Mccauley, M., ... ve Stoica, I. (2012). Fast and interactive analytics over Hadoop data with Spark. *Usenix Login*, *37*(4), 45-51.

Zenodo. (t.y.). About Zenodo. Erişim adresi: https://about.zenodo.org/

EK1 ORJİNALLİK FORMU

	HACETTEPE ÜNİVERSİTESİ **SOSYAL BİLİMLER ENSTİTÜSÜ**	Doküman Kodu *Form No.*	FRM-DR-21
		Yayım Tarihi *Date of Pub.*	04.01.2023
	FRM-DR-21 **Doktora Tezi Orijinallik Raporu** *PhD Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞINA

Tarih: 03/02/2025

Tez Başlığı: Türkiye'de Kurumsal İşleyiş ve Yönetsel Yapı Çerçevesinde Veri Gölü Tabanlı Veri Mimarilerinin Geliştirilmesi
Tez Başlığı (Almanca/Fransızca)*::.....

Yukarıda başlığı verilen tezimin a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 274 sayfalık kısmına ilişkin, 29/01/2025 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda işaretlenmiş filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 6 'dır.

Uygulanan filtrelemeler**:

1. ☒ Kabul/Onay ve Bildirim sayfaları hariç
2. ☒ Kaynakça hariç
3. ☒ Alıntılar hariç
4. ☐ Alıntılar dâhil
5. ☒ 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tezimin herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Ela Ankaralı

Öğrenci Bilgileri			
	Ad-Soyad	Ela Ankaralı	
	Öğrenci No	N18142213	
	Enstitü Anabilim Dalı	Bilgi ve Belge Yönetimi	
	Programı	Bilgi ve Belge Yönetimi	
	Statüsü	**Doktora** ☒	**Lisans Derecesi ile (Bütünleşik) Dr** ☐

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Özgür Külcü

*Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.
**Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları İkinci bölüm madde (4)/3'te de belirtildiği üzere: Kaynakça hariç, Alıntılar hariç/dahil, 5 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 5 words) filtreleme yapılmalıdır.

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-21
		Yayım Tarihi *Date of Pub.*	04.01.2023
	FRM-DR-21 **Doktora Tezi Orijinallik Raporu** *PhD Thesis Dissertation Originality Report*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

TO HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF INFORMATION MANAGEMENT

Date: 03/02/2025

Thesis Title (In English): Development of Data Lake Based Data Architectures Within the Framework of Institutional Operations and Administrative Structure in Türkiye
According to the originality report obtained by my thesis advisor by using the Turnitin plagiarism detection software and by applying the filtering options checked below on 29/01/2025 for the total of 274 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled above, the similarity index of my thesis is 6 %.

Filtering options applied**:

1. ☒ Approval and Decleration sections excluded
2. ☒ References cited excluded
3. ☒ Quotes excluded
4. ☐ Quotes included
5. ☒ Match size up to 5 words excluded

I hereby declare that I have carefully read Hacettepe University Graduate School of Social Sciences Guidelines for Obtaining and Using Thesis Originality Reports that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Ela Ankaralı

Student Information					
	Name-Surname	Ela Ankaralı			
	Student Number	N18142213			
	Department	Department of Information Management			
	Programme	Department of Information Management			
	Status	**PhD**	☒	**Combined MA/MSc-PhD**	☐

SUPERVISOR'S APPROVAL

APPROVED
Prof. Özgür Külcü

**As mentioned in the second part [article (4)/3]of the Thesis Dissertation Originality Report's Codes of Practice of Hacettepe University Graduate School of Social Sciences, filtering should be done as following: excluding refence, quotation excluded/included, Match size up to 5 words excluded.

EK 2 ETİK KURUL MUAFİYET FORMU

	HACETTEPE ÜNİVERSİTESİ SOSYAL BİLİMLER ENSTİTÜSÜ	Doküman Kodu *Form No.*	FRM-DR-12
		Yayım Tarihi *Date of Pub.*	22.11.2023
	FRM-DR-12 **Doktora Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for PhD Thesis*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞINA

Tarih:03/02/2025

Tez Başlığı: Türkiye'de Kurumsal İşleyiş ve Yönetsel Yapı Çerçevesinde Veri Gölü Tabanlı Veri Mimarilerinin Geliştirilmesi
Tez Başlığı (Almanca/Fransızca)*::.....
.....

Yukarıda başlığı verilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır.
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir.
4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırma niteliğinde değildir.
5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.

Hacettepe Üniversitesi Etik Kurullarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.
Gereğini saygılarımla arz ederim.

Ela Ankaralı

Öğrenci Bilgileri			
	Ad-Soyad	ELA ANKARALI	
	Öğrenci No	N18142213	
	Enstitü Anabilim Dalı	BİLGİ VE BELGE YÖNETİMİ	
	Programı	BİLGİ VE BELGE YÖNETİMİ	
	Statüsü	**Doktora** ☒	**Lisans Derecesi ile (Bütünleşik) Dr** ☐

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Özgür Külcü

* Tez **Almanca** veya **Fransızca** yazılıyor ise bu kısımda tez başlığı **Tez Yazım Dilinde** yazılmalıdır.

	HACETTEPE ÜNİVERSİTESİ **SOSYAL BİLİMLER ENSTİTÜSÜ**	Doküman Kodu *Form No.*	FRM-DR-12
		Yayım Tarihi *Date of Pub.*	22.11.2023
	FRM-DR-12 **Doktora Tezi Etik Kurul Muafiyeti Formu** *Ethics Board Form for PhD Thesis*	Revizyon No *Rev. No.*	02
		Revizyon Tarihi *Rev.Date*	25.01.2024

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF INFORMATION MANAGEMENT

Date:03/02/2025

ThesisTitle (In English): Development of Data Lake Based Data Architectures Within the Framework of Institutional Operations and Administrative Structure in Türkiye

My thesis work with the title given above:

1. Does not perform experimentation on people or animals.
2. Does not necessitate the use of biological material (blood, urine, biological fluids and samples, etc.).
3. Does not involve any interference of the body's integrity.
4. Is not a research conducted with qualitative or quantitative approaches that require data collection from the participants by using techniques such as survey, scale (test), interview, focus group work, observation, experiment, interview.
5. Requires the use of data (books, documents, etc.) obtained from other people and institutions. However, this use will be carried out in accordance with the Personal Information Protection Law to the extent permitted by other persons and institutions.

I hereby declare that I reviewed the Directives of Ethics Boards of Hacettepe University and in regard to these directives it is not necessary to obtain permission from any Ethics Board in order to carry out my thesis study; I accept all legal responsibilities that may arise in any infrigement of the directives and that the information I have given above is correct.
I respectfully submit this for approval.

Ela Ankaralı

Student Information					
	Name-Surname	ELA ANKARALI			
	Student Number	N18142213			
	Department	Department of Information Management			
	Programme	Department of Information Management			
	Status	**PhD**	☒	**Combined MA/MSc-PhD**	☐

SUPERVISOR'S APPROVAL

APPROVED
Prof. Özgür Külcü