VERİ İŞLEME SÜREÇLERİNDE TARTIŞMALI BİR ÇÖZÜM:

VERİ ANONİMLEŞTİRMESİ

Merve GÖZÜKÜÇÜK

111692026

İSTANBUL BİLGİ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

HUKUK YÜKSEK LİSANS PROGRAMI

(BİLİŞİM HUKUKU)

Danışman: Yard. Doç. Dr. Leyla BERBER

2014

VERİ İŞLEME SÜREÇLERİNDE TARTIŞMALI BİR ÇÖZÜM:

VERİ ANONİMLEŞTİRMESİ

A CONTROVERSIAL PROTECTION METHOD ON DATA PROCESSING :

DATA ANONYMIZATION

Merve GÖZÜKÜÇÜK

111692026

Yrd. Doç. Dr. Leyla BERBER :

Yrd. Doç. Dr. Nilgün BAŞALP :

Öğr. Görv. Bülent ÖNSOY :

Tezin Onaylandığı Tarih :

Toplam Sayfa Sayısı : 116

Anahtar Kelimeler (Türkçe)
1) Veri Anonimleştirmesi
2) Kişisel Verilerin Korunması
3) Gizlilik
4) Veri İşleme
5) Büyük Veri

Anahtar Kelimeler (İngilizce)
1) Data Anonymization
2) Personal Data Protection
3) Privacy
4) Data Processing
5) Big Data

Özet

Bilişim teknolojilerindeki hızlı gelişmeleri takiben hayatımızın her alanına giren büyük veri, sağladığı fayda ile enformasyonel toplumun hammaddesi haline dönüşmüştür. Veri odaklı yaklaşımlar, büyük veriden elde edilen faydanın ve kazanımın artmasına paralel olarak gelişmiş ve ticari, hukuki, idari, sosyal olmak üzere tüm faaliyetlerin merkezine yerleşmiştir. Bütün bu gelişmeler ve veri odaklı yaklaşımlar veri işleme, veri analizi, veri yönetimi süreçlerinin önemini arttırmış ve bu konuları ayrı birer uzmanlık seviyesine taşımıştır. Diğer taraftan ise, büyük veri, veri paylaşımı ve ifşa süreçlerini hızlandırmış ve bu durum kişisel verilerin korunması ve gizlilik hususlarında yeni tehditler ve yeni tartışmalar doğurmuştur. Büyük veri içeriğinde yer alan kişisel verilerin, paylaşım ve ifşa yoluyla daha geniş kitlelere yayılıyor olması, kişilerin kendi verileri üzerindeki denetimlerini büyük ölçekte düşürmüştür. Büyük veri ve veri odaklı yaklaşımların enformasyonel topluma sağladığı faydanın vazgeçilmezliği, uzmanları hem faydayı hem de gizliliği koruyacak mimari çözümler geliştirmeye itmiştir. Veri anonimleştirmesi, paylaşım ve ifşa süreçlerinde bu dengeyi sağlaması beklenen bir mimari çözüm olarak karşımıza çıkar. Veri anonimleştirmesi ile veri kümesi içinde, kişilerin kimliğini saptayan veya saptayabilme özelliği olan tüm veriler çıkartılarak, gizlenerek veya çeşitli istatistiksel metotlara tabi tutularak, veri kimliksizleştirilir. Böylece, verinin sağladığı çıktılar gelişime katkı sağlarken, veriyi üreten öznenin gizliliği korunmuş olur. Ancak, teoride dengeyi sağladığı düşünülen veri anonimleştirmesi, pratikte yaşanan bazı ihlallerle anonimleştirmenin güvenilirliği tartışmasını başlatmıştır. Yaşanan örneklerde, anonimleştirilmiş veri kümeleri, birden fazla dolaylı betimleyicinin bir araya gelmesiyle veya dışarıdan elde edilen ek bilgilerle birleştirilmesi sonucunda yeniden kişileri saptayabilir hale gelir ve böylece anonimleştirme bozarak veri öznelerinin kişisel bilgileri ifşa edilmiş olmaktadır.

Bu çalışma öncelikle büyük verinin genel çerçevesini, veri işleme operasyonlarına etkisini ve akabinde gelişen veri anonimleştirme süreçlerini detaylandıracaktır. Sonrasında kişisel verilerin korunması ve gizlilik hususunda yasal mevzuatlar Türkiye ve AB özelinde incelenecek ve bu konudaki hukuksal yaklaşım netleştirilecektir. Takiben, veri anonimleştirilmesi etrafında gelişen teknik ve içeriksel tartışmalar yaşanmış ihlal örnekleriyle analiz edilerek, fayda ve gizlilik dengesinin kurulması hususundaki mevcut anlayış ele alınacaktır. Sonuç bölümünde ise kontrol edilemeyen parametreler, mevcut tartışmaların eksikleri tespit edilerek, anonimleştirmenin ilkelerine dair öneriler geliştirilecektir.

Abstract

Due to the rapid developments in the information science, big data has diffused to the every aspect of our lives and has become a substantial source of the information society. As the utility gained from big data has expanded in years, it has gradually leveraged the data-centered approaches in commercial, legal, administrative and social activities. Accordingly, all these progress increased the value of the data management steps in terms of data processing and analysis and created the per se developing expertise on this field. Apart from its utility, big data, by intensifying the data sharing and disclosure processes, posed new privacy and data protection threats and initiated new debates. With respect to the indispensible contribution of big data and data-centered approaches to the modern information age, researchers aimed to produce architectural solutions in the aim of balancing the privacy with utility. As a result, data anonymization processes have been designed and implemented in order to fulfill this balance. Data anonymization roughly means to dismiss or omit all the identified or identifiable patterns of the data that can disclosure the personal information of data subjects. Thus, with the help of anonymization, big data could still be utilized while preserving the personal information and privacy of individuals. However, although this architecture created succesful results in theory, it practically failed in some real-life cases by emerging the outside information with anonymized data or by applying statistical calculations on the combinations of non-identifiable variables. Therefore, these unexpected real-life de-anonymization cases led the researchers to revisit the liability of the anonymization processes and mischiefed the faith on anonymization.

This thesis firstly aims to analyse the big data concept and its interactions with the data processing operations, specifically focusing on sharing and disclosure steps, and data anonymization methods. Afterwards, the privacy and data protection legislations that are currently in force in EU and Turkey will be investigated in order to draw the frame of the legal approaches. Following, technical and contextual debates regarding the faith on anonymization practices will be

examined in detail including the comprehensive researches conducted by both the advocates and the detractors of anonymization. Finally, by studying the divergent approaches, it will be concluded by mentioning the deficiencies in the prevailing debates and developing new suggestions in the aim of clarifying the fundamentals of anonymization.

İÇİNDEKİLER

KISALTMALAR

SQL	:	Search Query Language
CDR	:	Call Data Record
AB	:	Avrupa Birliği
ABD	:	Amerika Birleşik Devletleri
RFID	:	Radio-Frequency Identification
POS	:	Point of Sale
PRAM	:	Post-Randomization Method
AOL	:	America Online

KAYNAKÇA

29. Madde Veri Koruma Grubu	29. Madde Veri Koruma Grubu , Opinion 4/2007 on the concept of personal data , 2007,bkz. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf
Avrupa Veri Koruma Denetçisi	Avrupa Veri Koruma Denetçisi ,EDPS opinion on privacy in the digital age: "Privacy by Design" as a key tool to ensure citizens' trust in ICTs, Brussels, 2010,bkz. https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/PressNews/Press/2010/EDPS-2010-06_Privacy%20in%20digital%20age_EN.pdf
Bacak	Ahmet Bacak, Gizliliği Koruyarak Veri Yayınlamak İçin K-Anonimity ve L-Diversity Metodları, 2013, bkz. https://www.bilgiguvenligi.gov.tr/siniflandirilmamis/gizliliği-koruyarak-veri-yayinlamak-icin-k-anonimity-ve-l-diversity-metodlari.html
Barbaro/ Zeller	Michael Barbaro, Tom Zeller, A Face is Exposed for AOL Searcher No. 4417749, New York Times, bkz. http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0
Brown/Marsden	Ian Brown, Christopher T. Marsden, Regulating Code: Good Governance and Better Regulation in the Information Age, The MIT Press, 2013
Castells	Manuel Castells, Ağ Toplumunun Yükselişi, Birinci Cilt, çev. Ebru Kılıç, İstanbul Bilgi Yayınları, 2005
Christen/ Alfano/ Bangenter/ Lapsley	Markus Christen, Mark Alfano, Endre Bangerter, Daniel Lapsley, Ethical IIssues of Morality Mining: Moral Identity as a Focus of Data Mining, Ethical Data Mining Applications for Socio-Economic Development, IGI Global, 2013
Chunara/ Andrews/ Brownstein	Rumi Chunara , Jason R. Andrews, John S. Brownstein, Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak, The American Society of Tropical Medicine and Hygiene,2010,bkz. http://healthmap.org/documents/Chunara_AJTMH_2012.pdf
Couvakian	Ann Couvakian, Privacy By Design…Take the Challange, Canada, 2009
Demirci	İlkay Demirci, T-Closeness Metodu Gizliliği Koruyarak Veri Yayınlamak İçin, 2014 bkz. http://www.phphocam.com/t-closeness-metodu-gizliliği-koruyarak-veri-yayinlamak-icin/#sthash.z70qZ2sb.dpuf
Digital Rights Ireland and Seitlinger	Judgment in Joined Cases C-293/12 and C-594/12,Digital Rights Ireland and Seitlinger and Others, Court of Justice of the European Union , Press Release No 54/14, Luxembourg, 8.4.2014

Directive 95/46/EC	Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Oficial Journal of the European Communities of 23 November 1995, No L. 281, s. 31.
Directive 2002/58/EC	Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and protection of privacy in the electronic communications sector OJ L201/37
Directive 2006/24/EC	Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the Retention of data generated or processed in connection with the provision of publicly available electronic communications service sor of public communication Networks and amending Directive 2002/58/EC, OJ L 105
Enformasyon Komiserliği Ofisi	Enformasyon Komiserliği Ofisi, Privacy by Design, 2008, bkz. http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/documents/pdb_report_html/PRIVACY_BY_DESIGN_REPORT_V2.ashx
Enformasyon Komiserliği Ofisi	Enformasyon Komiserliği Ofisi, Anonymization: Managing Data Protection Risk Code of Practice, 2012.bkz. http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
Gür	İkbal Gür, Kişisel Verilerin Korunması Hususunda AB ile ABD Arasında Çıkan Uyuşmazlıklar, Turhan Kitabevi, 2010
Gürses/ Danezis	Seda Gürses, George Danezis, A Critical Review of Ten Years of Privacy Technology, UK, 2012
Gürses/ Troncoso/ Diaz	Seda Gürses, Carmela Troncoso, Claudia Diaz, Engineering Privacy by Design, International Conference on Privacy and Data Protection (CPDP) Book, 2011
Hilbert	Martin Hilbert, Big Data for Development: From Information- to Knowledge Societies, United Nations ECLAC, 2013
Honer	Jason Honer, U.S. government commits big R&D money to 'Big Data', , bkz.http://www.zdnet.com/blog/btl/u-s-government-commits-big-r-andd-money-to-big-data/72760
ESSNet	Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, Handbook on Statistical Disclosure Control Version 1.2, ESSNet, 2010
ESSNet-Project/ µ- ARGUS version 4.2	Anco Hundepool, Aad van de Wetering, Ramya Ramaswamy, Luisa Franconi, Silvia Polettini, Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo, Vicenc Torra, Ruth Brand, Sarah Giessing, µ- ARGUS version 4.2 User's Manuel,

	ESSNet-Project, 2008
Hunter/Letterie	Jenny Hunter, Jelmer Letterie, IBM harnesses power of Big Data to improve Dutch flood control and water management systems, bkz. http://www-03.ibm.com/press/us/en/pressrelease/41385.wss
Hurwitz/ Nugent/ Halper/ Kaufman	Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, Big Data For Dummies, Wiley & Sons, 2013
IHSN	International Household Survey Network, Anonymization Principles, bkz. http://www.ihsn.org/home/node/137
IHSN	International Household Survey Network, Reducing the Disclosure Risk, bkz. http://www.ihsn.org/home/node/201
Irzık	Gürol Irzık, "Bilgi Toplumu mu, Enformasyon Toplumu mu? Analitik-Eleştirel Bir Yaklaşım", Bilgi Toplumuna Geçiş Sorunlar Görüşler Yorumlar Yorumlar Eleştiriler Ve Tartışmalar, Tüba Yayınevi, 2005
Koot	Matthijs R. Koot, Measuring and Predicting Anonymity, Gildeprint Drukkerijen, 2012
Korff	Douwe Korff, Comperative Study on Different Approaches to New Privacy Challanges, In Particular in the Light of Technological Developments, Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in meeting the challanges posed by global social and technical developments, London Metropolitan University, 2010,bkz. http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf
Krishnan	Krish Krishnan , Data Warehousing in the Age of Big Data, Newnes, 2013
Küzeci	Elif Küzeci, Kişisel Verilerin Korunması, Turhan Kitabevi, 2010
Laney	Doug Laney, 3D Data Management: Controlling Data Volume, Velocity and Variety, META Group, 2001. Bkz. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
Lessig	Lawrance Lessig, *Code Version 2.0*, Basic Books, 1996
Levine/Roos	Joel H. Levine, Homas B. Roos, Introduction to Data Analysis:The Rules of Evidence, bkz. http://www.dartmouth.edu/~mss/docs/Volumes_1-2.pdf
Li/Li/ Venkatasubramanian	Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, t-Closeness: Privacy beyond k-Anonymity and l-Diversity, Data Enginering (ICDE) IEEE 23rd International Conference, 2007
Machanavajjhala/ Gehrke/ Kifer	Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, l-Diversity: Privacy Beyond k-Anonymity,

	Cornell University, 2007
Moore	Richard A. Moore, Jr, Controlled Data-Swapping Techniques for Masking Public Use Microdata Sets, US Bureau of the Census Washington, 1996
Morozov	Evgeny Morozov, The Net Delusion: How not to Liberate World, Penguin Books, 2011
Narayanan/ Shmatikov	Arvind Narayanan, Vitaly Shmatikov, How to Break Anonymity of the Netflix Prize Dataset, The Universtiy of Texas, 2008
Ohm	Paul Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, UCLA Law Review, Vol 57, 2010
Oram	Andrew Oram, The Information Technology Fix For Health, OReilly, 2014
Özdemir	Hayrunnisa Özdemir, Elektronik Haberleşme Alanında Kişisel Verilerin Özel Hukuk Hükümlerine Göre Korunması, Seçkin Yayıncılık, 2009
Özmen	Şule Işınsu Özmen, Ağ Ekonomisinde Yeni Ticaret Yolu: E-Ticaret, İstanbul Bilgi Üniversitesi Yayınları, 2012
Phitzmann/ Hansen	Andreas Pfitzmann, Marit Hansen, Anonymity, Unobservability, Pseudonymity, and Identity Management:A Proposal for Terminology, bkz. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.18.pdf
Schmarzo	Bill Schmarzo, Big Data:Understanding How Data Powers Big Business, Wiley, 2013
Simon	Phil Simon, Too Big To Ignore:The Business Case for Big Data, Wiley, 2013
Spiekerman/ Cranor	Sarah Spiekerman, Lorrie Faith Cranor, Engineering Privacy, IEEE Transactions on Software Engineering, Vol. 35, Nr. 1, 2009
Bulletin of IEEE	Alain Biem, Eric Bouillet, Hanhua Feng, Anand Ranganathan, Anton Riabov, Olivier Verscheure, Haris Koutsopoulos,Mahmood Rahmani, Barış Güç, Real-Time Traffic Information Management using Stream Computing, bkz. http://sites.computer.org/debull/A10june/Anand.pDf
Sweeney	Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, Carnegie Mellon University, 2002
Swire/ Ahmad	Peter P. Swire, Kenesa Ahmad, Foundations of Information Privacy and Data Protection, IAPP,2012
Şimşek	Oğuz Şimşek, Anayasa Hukukunda Kişisel Verilerin Korunması, Beta Basım, 2008
Yakowitz	Jane Yakowitz, Tragedy of Data Commons, Harvard Journal of Law and Technology, Vol.25, 2011
Warren/Brandeis	Samuel D. Warren, Louis D. Brandeis, The Right to Privacy, Harvard Law Review, 1890

Wolfe/ Gunesekara/ Bogue	Nathan Wolfe, Lucky Gunasekara, Zachary Bogue, Crunching Digital Data can help the World, 2011, bkz. http://edition.cnn.com/2011/OPINION/02/02/wolfe.gunasekara.bogue.data/index.html?_s=PM:OPINION
Wu	Felix T. Wu, Defining Privacy and Utility in Data Sets, University of Colorado Law Review 1117 (2013)

TABLOLAR

Veri İşleme Süreçlerinde Tartışmalı Bir Çözüm:

Veri Anonimleştirmesi

1. Giriş

1960'lı yıllar sonrasında büyük bir hızla gelişen bilgi ve iletişim teknolojileri hayatın her alanında devrimsel nitelikte farklılıklara ve süreçlere yol açmıştır. ABD Savunma Bakanlığı'nın, olası bir nükleer savaşta iletişim ağlarının çökertilmesini engellemek için kurduğu gayri merkezi ağ topolojileri, sonrasında kurgulanan iletişim standartları ve protokolleriyle birbirine bağlanarak Internet adını verdiğimiz global bir ağ meydana getirmiş ve fiziksel mesafelere bağımlı kalmadan iletişim kurabilme maliyetlerini benzersiz şekilde düşürmüştür. Böylelikle, bilgi teknolojilerine yapılan altyapı yatırımları, yaygın Internet ağları, gelişmiş iletişim ve otomasyon uygulamaları, verilerin depolanması ve işlenmesini kolaylaştıran donanım ve yazılımlar ile bilgi ve iletişim teknolojileri, hayatın vazgeçilmez bir parçası haline dönüşmüştür. Bugün gelinen noktada, bireyler nezdinde resmi veya gayri resmi nitelikte olan ve gündelik hayatımızdan kariyerimize, devletle olan ilişkimizden sosyal çevremize kadar geniş bir yelpazede çeşitlilik gösteren süreçlerimiz, kurumlar nezdinde ise tüm altyapı ve otomasyon ağları, bilgi ve iletişim teknolojileri aracılığıyla sağlanmaktadır. Castells, süreci kalkınma biçilerindeki değişimle ilişkilendirip açıklamaktadır. Endüstriyel kalkınma biçimlerini, 20. yüzyılın sonuna doğru gelişen bilgi teknolojileri ile enformasyonel kalkınma biçimi takip etmiştir ve bu yapıda "üretkenliğin kaynağı, bilgi üretme, bilgi işleme ve sembollerle iletişim teknolojisindedir, […], enformasyonel kalkınma biçimine özgü olan şey, bilginin üzerine bilgi gelmesi eyleminin bizzat üretkenliğin ana kaynağı olmasıdır[1]." Bu

[1] Manuel Castells, Ağ Toplumunun Yükselişi, Birinci Cilt, çev. Ebru Kılıç, İstanbul Bilgi Yayınları, 2005, s.20

noktada bilgi ve enformasyon terimlerinin farklılıklarına değinmek faydalı olacaktır. Irzık, bu ayrımı yaparken bilginin özneyle olan ilişkisine vurgu yapar; "bilgi, her şeyden önce, bilen özne ile bilinen şey arasında iki terimli bir ilişkidir, […], öznesiz bilgi olmaz[2]". Enformasyon ise "bilginin hammaddesidir[3]" ve "bilgi, tanımı gereği yalnızca bilen öznenin zihninde varolabilir, […], enformasyon ise çeşitli biçimlerde, örneğin bilgisayar ortamında varolabilir[4]". Bu tanımlara göre, bilginin insanlık tarihi kadar eski olduğunu, ancak özneye bağlı olmadan işlenen, saklanan, paylaşılan enformasyonun teknoloji ile mümkün kılındığını söylemek yanlış olmaz. Enformasyonun sayısal işleme ortamlarında saklandığı, işlendiği, paylaşıldığı şekli ise veridir.

Enformasyonel kalkınma biçimlerinin yarattığı köklü dönüşümlerin çok boyutlu sonuçları olmuştur. İletişim teknolojileri ile yürüttüğümüz hayatlarımız bireysel, toplumsal ve hatta uluslar arası boyutta enformasyon ve akabinde veri üretmektedir. Parçası olduğumuz her ağ veya sistem kendi içinde veri kümelerinden oluşmaktadır ve her yeni katılımcının varlığı ve hareketleri sürekli ve çok çeşitli yeni veri kümeleri doğurmaktadır. Bu veri kümelerinin birleşmesi sonucunda da ölçeklenmesi zor bir yapı olarak değerlendirebileceğimiz bir veri evreni olan *büyük veri*[5] kavramıyla karşılaşmış oluruz. Kişisel verilerimiz, varlıklarımız, sağlık tarihçemiz, sosyal ve profesyonel tercihlerimiz, yaşama alanımız gibi hayatımıza dair pek çok detay hayatımızın her alanında karşımıza çıkan bilgi teknolojileri vasıtasıyla büyük veriyi beslemektedir. Ayrıca enformasyonel kalkınma biçiminin dinamiklerine uygun olacak şekilde gelişim, kendi kendini besleyen bir düzlemde ilerlemektedir. Bilgi ve iletişim teknolojileri, sistemler ve altyapılar üzerinden büyük veriyi düzenli olarak beslerken, diğer

---

[2] Gürol Irzık, "Bilgi Toplumu mu, Enformasyon Toplumu mu? Analitik-Eleştirel Bir Yaklaşım", Bilgi Toplumuna Geçiş Sorunlar Görüşler Yorumlar Yorumlar Eleştiriler Ve Tartışmalar, Tüba Yayınevi, 2005, s. 54

[3] Elif Küzeci, Kişisel Verilerin Korunması, Turhan Kitabevi, 2010, s. 10

[4] Gürol IRZIK, s. 56

[5] Büyük veri kavramı ve esaslarına, ilk olarak Doug Laney tarafından hazırlanmış bir raporda değinilmiştir. Hacim, hız ve çeşitlilik vurguları ile rapor, bugünkü büyük veri tanımın oluşmasındaki temelleri atmıştır. *3D Data Management: Controlling Data Volume, Velocity and Variety,* META Group, 2001. Bkz. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

yandan, büyük veriyi girdi olarak kullanarak elde edilen sonuçlar ekonomik, sosyal ve politik açılardan ihtiyaçların netleştirilmesini sağlar ve kişiye veya kuruma özel iş modelleri, esnek uygulamalar ve gelişmiş teknolojilerin tasarlanmasına ön ayak olur. Bu sebeple büyük veri, pazarın vazgeçilmez hammaddesi haline dönüşmüştür. Küreselleşme politikaları ile temelleri atılmış olan ve Internet ile akıl almaz bir hızla bağlantılar kurabilen ağ toplumu, sürekli veri üreterek pazarın ve politik gündemin yönlendiricisi olmaktadır. Büyük verinin teknolojik ve akademik gelişimler, ekonomik faaliyetler, siyasi ve kültürel farkındalıklar nezdinde toplumlara katkısı yadsınamaz. Ancak büyük veri, içeriği ve ölçeği ile sağladığı faydaların yanında, veri öznesinin gizliliğinin ve akabinde güvenliğinin korunması hususunda yeni tehditlere sebep olmuş ve yeni tartışmaları başlatmıştır. Büyük veri olarak değerlendirelim ya da değerlendirmeyelim, verinin saklanması, sınıflandırılması, yorumlanması, ifşası, birleştirilmesi, paylaşılması gibi veri işleme ve yönetimi süreçleri, bilgi teknolojilerinin gelişen yapısıyla hayati önem kazanmıştır. Özel sektör veya kamu sektörü fark etmeksizin, veriyi işleyen ve yöneten partilerin veriyi kullanım amaçları, paylaşım noktaları, saklama modelleri ve veri yönetimindeki şeffaflık anlayışları, veriyi üreten ve asıl sahibi olan veri özneleri nezdinde hayati önem teşkil eder. Ancak bu noktada, büyük verinin toplumu ve pazarı yönlendirici kapasitesi tüm dikkatlerin üzerine çekilmesini sağlamış ve bu durum güvenlik açıklarının ve gizlilik ihlallerinin daha büyük risklere yol açmasını sağlamıştır.

Büyük verinin faydalı sonuçlar üretebilmesi için ilk adım verinin işlenmesidir. İşleme pek çok alt fonksiyondan oluşmaktadır ve veri yönetimi için gereklidir. İşlemenin önemli iki fonksiyonunda biri olan verinin paylaşılması ve ifşası ise sebep oldukları güvenlik açıkları ve tehditler sebebiyle özel olarak ele alınmalıdır. İşte tam da bu noktada, büyük verinin katkılarından faydalanmakla, bireylerin güvenliklerini ve gizliliklerini korumak adına bazı denge metotları tasarlanmıştır. Bu yöntemlerle hem veriden faydalanmak isteyen partilerin talepleri karşılanırken hem de bireylerin temel hak ve özgürlüklerinin korunması hedeflenmektedir. Mevcut süreçler incelendiğinde görülmektedir ki fayda ve güvenlik dengesinin

sağlanması için üretilen çözümlerden biri veri anonimleştirilmesi süreçleridir. Anonimleştirme uygulamaları ile veriler, veriyi üreten veri öznelerini betimleyici özelliklerinden arındırılır ve kimliksizleştirilir. Böylece anonim veriye bakılarak veri öznelerinin kişisel verilerine erişilemez. Bu sayede, aynı anda veriden hem faydalı çıktılar üretebilmek hem de veri öznelerinin kişisel bilgilerini ve gizliliğini korumak mümkün olur. Ancak, teoride bu dengeyi sağladığı düşünülen anonimleştirme süreçleri pratikte varsayımların dışında sonuçlar üretebilmekte ve anonimleştirilmiş veriler sayesinde kişisel bilgiler ortaya çıkartılarak bireylerin kimlikleri saptanabilmektedir. Kişilerin kendi rızalarıyla da pek çok kişisel detaylarını çeşitli uygulamalar vasıtasıyla ifşa ettiği gerçeğini de göz önünde tutarsak, dışarıdan edinilen bilgilerin anonim verilerle birleştirilmesi sonucunda anonimleştirme bozulabilmekte ve kişilerin hassas verilerine bile ulaşılabilmektedir. Bu örnekler ise anonim veri ve anonimleştirme süreçlerinin koruyucu etkisine olan güveni sarsmaktadır.

Bu çalışma, büyük veriden fayda sağlanması amacıyla gerçekleştirilen veri işleme süreçlerinin iki önemli alt başlığı olan veri ifşası ve veri paylaşımı fonksiyonlarına odaklanmakta, güvenlik açıklarını ve gizlilik ihlallerini engellemesi beklenen anonimleştirme süreçlerinin teknik altyapısını, Avrupa Birliği ve Türkiye mevzuatlarındaki referanslarını incelemekte, anonimleştirmenin tartışmalı örneklerini analiz ederek sonuç bölümünde anonimleştirmeye dair ilkeleri kurgulamayı hedeflemektedir. Böylelikle anonimleştirmenin güvenilirliğine dair tartışmaların odak noktaları ve bu noktalardaki eksikler değerlendirilecek ve hukuki bakış açısı analiz edilecektir. İlk bölümde tanımı, kapsamı, kaynaklarıyla veri ve büyük veri kavramı detaylandırılıp, veri işleme süreçlerinin aktörleri ve gerekliliği incelenecek ve gizlilik konusu işlenecektir. İkinci bölüm anonimleştirme kavramını, metotlarını, amacını ve büyük veri içindeki duruşunu belirleyecektir. Üçüncü bölüm anonimliğin güvenilirliğine dair tartışmaları, anonimliğin ölçülmesine dair araştırmaları içerecektir. Sonuç bölümünde ise anonimleştirme süreçlerine dair tartışmanın boyutu ve algısı netleştirilecek, eksikler tespit edilecek ve ilkeler belirlenerek öneriler geliştirilecektir.

2. Veri, Büyük Veri, Veri İşleme ve Gizlilik

Anonimleştirme, özünde büyük verinin yarattığı tehditlere ve gizlilik sorunlarına istinaden üretilmiş bir çözümdür. Bu bağlamda, anonimleştirme süreçlerinin geliştirilmesi büyük veri süreçleriyle dolaylı olarak ilişkilidir. Bu durumda, büyük veri neden oluşuyor, içeriği neler, hangi boyutlarıyla faydalı, ne ölçüde ve çeşitlilikte tehditler içeriyor? Bu bölüm bu sorulara cevap arayabilmek için öncelikle veri kavramını ve verinin bileşenlerini netleştirmeyi, büyük veri olarak adlandırdığımız veri evreninin tanımlamayı hedeflemektedir. Böylece, büyük verinin dinamik yapısını, kapsamını, kaynaklarını ve olumlu-olumsuz etkilerini incelemek daha sağlıklı olacaktır. Ek olarak, büyük verinin anlamlı çıktılar üretmesini sağlayan veri işleme süreçlerine değinilerek, bu süreçlerden paylaşım ve ifşa fonksiyonlarına odaklanılacaktır. Bu çalışma boyunca, paylaşım ve ifşa anonimleştirme süreçlerinin kurgulanmasında temel dayanaklar olarak değerlendirilmektedir. Çünkü anonimleştirme, amaç olarak güvenli payaşım ve ifşa yapılmasını sağlamayı hedefleyen bir çözüm olarak sunulmuştur. Bu bağlamda anonimleştirmenin gerekli güvenliği sağlayamadığı ve gizliliği koruyamadığı hallerde paylaşım ve ifşanın durdurulması veya engellenmesi gerektiği kanısı doğmaktadır. Böyle bir çözümün sunulamayacağını göstermek adına da büyük verinin ve veri odaklı mekanizmaların çerçevesini netleştirmek gerekmektedir.

I. Veri Nedir?

Veri en basit haliyle enformasyonun en ham ve küçük parçasıdır[6]. Veriler, insanlar, nesneler, işlemler, uygulamalar, olaylarla ilgili gerçekleri yansıtan

[6] Krish Krishnan , *Data Warehousing in the Age of Big Data*, Newnes, 2013, s. 3

niceliksel ve niteliksel değerlerdir ve işlenerek enformasyonu oluştururlar[7]. Veri, otomatik yollarla kayıt altına alınmadan önce de manuel yöntemlerle tutularak ve saklanarak da yine toplumların hayatında büyük öneme sahipti. Ancak gelişen otomasyon sistemleri veriye çok yönlü ticari ve toplumsal anlamlar katmıştır. Ancak veri, yönetebildiğimiz ölçüde hayatlarımıza yön verebilmektedir. Kayıt altına alınmayan, ölçeklenemeyen, sınıflandırılamayan verilerin araştırmalara, yeni tasarımlara veya kurgulara yol göstermesi mümkün değildir. Verinin yorumlanabilmesi ve işaret ettiği konularda bilgi verebilmesi için verinin işlenmesi ve analiz edilmesi şarttır. Verinin işlenmesi ilerleyen bölümlerde ayrıca ele alınacaktır. Bunun yanında, "veri analizi, olayları açıklayan, kalıpları belirleyen, tanımlar geliştiren ve hipotezleri test eden yöntemler topluluğudur[8]". Veri analizi adımında ortaya çıkan en genel sınıflandırma verinin formatına göre yapılabilir. Buna göre veri yapılandırılmış, yapılandırılmamış ve üst veri olmak üzere 3 temel başlık altında incelenebilir.

A. Yapılandırılmış Veri

Yapılandırılmış veri, geleneksel otomasyon uygulamalarıyla kolaylıkla işlenebilen, ilişkisel veri tabanlarında büyük ölçekli maliyet ve performans ihtiyaçları yaratmadan tutulabilen ve düzenli bir yapıya sahip veri kümeleridir. Örneğin; özel sektör açısından değerlendirirsek bir bankanın sistemlerinde tuttuğu müşteri numarası, müşteri adı ve soyadı, erişim bilgileri, cinsiyeti, ürün bilgileri gibi düzenli, ilişkili veri kümeleri yapılandırılmış veri olarak değerlendirilir. Bu veriler geleneksel ilişkisel veri tabanı yönetim sistemleri acılığıyla işlenebilir ve saklanabilir. Böylelikle günümüzde kullanılan SQL uygulamaları ve ilişkisel veri tabanı sistemleri yapılandırılmış verilerin yönetimini, raporlanmasını ve birbiriyle ilişkilendirilmesini sağlamıştır. Böylelikle "yazılımcılar yeni ilişkiler kurarak yeni

[7] Şule Işınsu Özmen, Ağ Ekonomisinde Yeni Ticaret Yolu: E-Ticaret, İstanbul Bilgi Üniversitesi Yayınları, 2012, s. 408

[8] Joel H. Levine, homas B. Roos, *Intrıduction to Data Analysis:The Rules of Evidence*, bkz. http://www.dartmouth.edu/~mss/docs/Volumes_1-2.pdf

veri kümeleri üretebilir, […], yöneticiler verileri envanter bazında ve karşılaştırılmalı olarak inceleyerek karar verebilir, […], farklı coğrafyalara ait müşteri bilgileri karşılaştırılabilir[9]".

B. Yapılandırılmamış Veri

Yapılandırılmamış veri, yapılandırılmışın tersine, ilişkilendirilemeyen, dağınık, metin içeren ve geleneksel veri depolama ve yönetim sistemleriyle kolaylıkla işlenemeyen yapıdaki verilerdir[10]. Bu tip verinin işlenmesi için performansı ve kapasitesi yüksek uygulamalar kullanılması gerekmektedir. Yapılandırılmamış veri, bugün anıldığı haliyle "büyük veri" olarak da tanımlanmaktadır ve ilerleyen bölümlerde ayrıca detaylandırılacaktır.

C. Üst veri

Üste veri en kısa tanımıyla veri hakkında veri demektir[11]. Bir veriyi üretirken her an o veriye ait üst veriler de oluşmaktadır. Bu haliyle üst veri, veri analizi yapılırken ek enformasyon sağlamakta ve araştırmayı kolaylaştırmaktadır. Bir örnekle açıklamak gerekirse, bir banka şubesi çalışanı o esnada işlemini gerçekleştirdiği müşterinin mail adresini öğrenirse bu bilgiyi banka müşterilerinin bilgilerinin yönetildiği ara yüz programıyla sisteme eklemek isteyecektir. Böyle bir durumda ara yüz programında ilgili müşterinin sayfasına erişerek bilgilerini görüntüler. Sonrasında edindiği mail adresi bilgisini sistemde ilgili alanlara girerek kaydeder. Böyle bir işlem sonucunda arka planda müşteriye ait veri tabanı tablolarında mail adresi verisi güncellenir. Bu örnekte girişi yapılan mail adresi

[9] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, *Big Data For Dummies*, Wiley & Sons, 2013
[10] Phil Simon, *Too Big To Ignore:The Business Case for Big Data*, Wiley, 2013, s. 35
[11] Phil Simon, s. 36

bilgisi yapılandırılmış veridir. Ancak sistem eş zamanlı olarak farklı tablolara bu kayıt işleminin yapıldığı tarihi, kaydı gerçekleştiren kullanıcın adını ve kaydın gerçekleştiği şubenin kodu gibi bilgileri de işleyecektir. İşte bu gibi işlemin kendisine dair tutulan bilgilere üst veri denmektedir. Bu veriler asıl veri olarak saydığımız mail adresi bilgisine istinaden verilerdir ve analiz edilmeleri halinde bankanın işleyişine ve çalışanlarına dair pek çok bilgi sağlamaktadırlar.

II. Yapılandırılmamış Veri Evreni: Büyük Veri

A. Tanımı ve Gelişimi

Yukarıda açıklandığı üzere yapılandırılmış veriler düzenli, ilişkili ve kolayca işlenebilir durumdayken yapılandırılmamış veriler bu özellikleri göstermez. Ölçeklenmesi ve kontrol edilmesi güç özelikler gösteren yapılandırılmamış veriler bir araya gelerek büyük bir veri evreni yaratırlar. Bu haliyle, kaynağını yapılandırılmamış verilerden alan bu büyük veri evrenini "büyük veri" olarak ifade edilmektedir. Bu tanımı detaylandıracak olursak, büyük veri, değişken derecelerde karmaşıklığa sahip, aşırı ölçekli hacmi olan, geleneksel teknolojiler, veri işleme yöntemleri, algoritmalar ve standart ticari çözümler ile işlenemeyecek kadar farklı hızlarda üreyip ve değişken derecelerde muğlaklık içeren veri kümesidir[12]. Özellikle, yakınsayan teknolojiler ile değişen veri yönetimi anlayışı sebebiyle büyük veri, aşırı büyük hacim, aşırı yüksek hız ve aşırı geniş çeşitlilik kavramlarıyla betimlenir[13]. Bilgi sistemlerindeki hızlı gelişmeler, verinin yapısı, ölçeği ve kapsamını direk etkileyerek hem özel sektörde hem kamu bünyesinde veri analizinin değerini arttırmıştır. Krishnan, teknolojik gelişimin büyük verinin oluşmasındaki kronolojik etkisini incelediğinde, süreci 5 temel dönemle ele

[12] Krish Krishnan, s. 5
[13] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, s. 10

almaktadır. Öncelikle 1980'li yıllarda geliştirilen karar destek ve veri madenciliği uygulamaları ile yeni eğilimler oluşturulmuş, tarihsel analizler gerçekleştirilmiş, mantıksal analiz yapılabilmiş ve yüksek ölçekli metrikler yaratılarak pek çok yeni çözümün sağlanması ve şirketin kurulması ile başlı başına bir sektör ortaya çıkmıştır. İkinci dönem ise 1995 yılından itibaren gelişen e-ticaret uygulamaları ve Internet'teki ticari yapılanmalar dönemidir. Bu periyotta yeni bir tüketim dünyası keşfedilmiş ve hayatımıza noktadan noktaya iletişim süreçleri girmiştir. Bu yapılanmalar verinin hacmine ve çeşitliliğine büyük bir ivme kazandırmıştır. Bu dönemi takip eden 5 ile 7 yıl arasındaki ticari anlamda pek çok yeni iş modeli gelişmiş ve bu modeller hızlıca kendi standartlarını geliştirmiştir. Bu yeni ticari ortamda, tüketicilerin tecrübelerine dair geri dönüşleri, anketler ve ağızdan ağıza pazarlama yöntemleri ile ortaya çıkan veriler sürece o zamana kadar görülmemiş bir hacim katmıştır. Krishnan'a göre, bir diğer belirleyici dönem ise 1997 ve 2002 yıllarında yükselen bir eğilim gösteren mobil çözümler dönemidir. Hücresel iletişim sayesinde sesli mesajlar, metin mesajları önem kazanmış ve bu durum iletişimi arttırırken topluma yönelik servis ve ürünlerin geliştirilmesine ön ayak olmuştur. Bütün bu gelişmeler ise sektöre ve veri yönetimi süreçlerine hacim, çeşitlilik, karmaşıklık ve kullanım açıcından yeni bir boyut kazandırmıştır. Son olarak ise 2000'lerden sonraki dönemi yazar şöyle açıklamaktadır;" […] arama motorlarının ortaya çıkması (Google, Yahoo), müziğin kişiselleştirildiği (iPod), tablet bilgisayalar (iPad), kapsamlı mobil çözümler (akıllı telefonlar, 3G ağlar, mobil genişbant erişim, Wi-Fi), sosyal medyanın ortaya çıkması (Facebook, Twitter, MySpace ve Blogger) ile veri dünyasının belirlendiği nokta[14]."

Yukarıda bahsedilen dönemlerde gelişen teknoloji, bireylerin bilgi ve iletişim uygulamalarına daha fazla dahil olup, daha farklı boyutlarda birleşik yapılı, dağınık, süreklilik arz eden veriler üretmelerine sebep olmaktadır. Bunun yanında, üreyen yüksek hacimli verinin depolanabilmesi, sınıflandırılması, raporlanabilmesi için ihtiyaç duyulan yüksek performanslı ve düşük maliyetli altyapılar olmadan, verimli sistemler kurabilmek ve veriden anlamlı sonuçlar

[14] Krish Krishnan, s. 5

üretebilmek mümkün olmayacaktır. İşletmeler, saklama ve işleme maliyetlerinin talep ettikleri her veriyi analiz etmelerini engellemesinden ötürü genellikle önemli buldukları bilgilerin anlık seyirlerini veya alt kümelerini saklayarak süreçten feragat ediyorlardı[15]. Martin Hilbert[16], büyük verinin enformasyon toplumu yaratmada sağladığı katkıyı tartıştığı makalesinde bütün bu ekstra verinin kaynaklarını 3 temel maddeye bağlamaktadır; veri akışı, veri depolama kapasitesi, veri işleme süreçleri. Veri akışı 2000'lerden sonra gelişen sosyal medya ve arama motoru uygulamaları ve bu tip uygulamaların kullanımını arttıracak altyapı yatırımları sayesinde inanılmaz ölçülerde artmıştır. 2012'de yapılan bir ölçüme göre, Google dakikada 2 milyon arama sorgusu alırken, Facebook kullanıcıları dakikada 700,000 adet içerik paylaşıyor ve Twitter kullancıları ise dakikada 100,000 tweet göndermektedir. Hilbert'ın bir diğer vurgusu veri depolama yöntemlerindeki gelişmelere yöneliktir. Makalesinde paylaştığı rakamlara göre, 1986 yılında tüm depolama araçları kullanılarak ancak iletişim verisinin %1 inden azı depolanabilmekteyken 2007 yılında bu değer %16'a çıkmıştır. Üçüncü olarak ise veri işleme kapasitelerindeki artıştır. Telekom ve depolama kapasiteleri geçtiğimiz on yıla oranla %25-30 oranında büyürken, veri işleme kapasiteleri %60-80 oranında gelişmiştir. Bütün bunların yanında kullanıcı bazlı üreyen verilerin dışında kullanıcılardan bağımsız olarak oluşan istatistik ve takip verileri de büyük verinin önemli kaynaklarından biridir. Yeni gelişen teknolojiler sadece daha gelişmiş mobil cihazlar, yaygın ve kolay kullanılabilir Internet uygulamaları veya yüksek performanslı otomasyon sistemleri geliştirmekle kalmamış aynı zamanda farklı modellere hizmet edebilen makine ve donanımlar da geliştirmiştir. Örneğin, radyo frekans göstergeleri (RFID) gibi fiziksel sensörler tarafından üreyen etiketler sayesinde fiziksel hayatın nesnelerine dair veri üretilebilmekte ve bu veriler çevrimiçi verilerle birleştirilebilmektedir[17]. Bütün bu iç içe geçmiş ve

[15] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, s. 13
[16] Martin Hilbert, *Big Data for Development: From Information- to Knowledge Societies*, United Nations ECLAC, 2013

[17] Ian Brown, Christopher T. Marsden, *Regulating Code: Good Governance and Better Regulation in the Information Age*, The MIT Press, 2013, s. 50

sınırları muğlak yapıyı hesaba katarak büyük verinin tanımını yeniden düşündüğümüzde görüyoruz ki, "büyük veri, tablolarda saklanabilmesi yeterince uygun olmayan, SQL işleyişine olumsuz yanıt dönen her tip veri parçasıdır" ve bu haliyle "görüntü dosyalarına Telekom şirketlerindeki CDR dosyalarına, ağ kayıtlarına, sosyal medya verilerine, RFID çıktılarına büyük veri diyebilirken müşteri ve çalışan kayıtları veya ürün listeleri büyük veri olarak adlandırılacak nitelikte değillerdir[18]." Ek olarak, ölçülemeyecek boyutlarda hacime ve çeşitliliğe sahip olan büyük verinin kullanım alanları ve çıktıları da aynı hacim ve çeşitlilikte sonuçlar üretmektedir. Her sektör ve iş modeli büyük veriyi beslerken aynı zamanda, yönetebildiği ölçüde büyük veriden fayda sağlayabilmekte ve büyük veriyi girdi olarak kullanmaktadır. Büyük verinin etkilediği ve geliştirdiği alanları aşağıdaki 3 temel başlık altında toplayabiliriz.

B. Etki Alanı

1. Inovasyon

Hayatın her katmanına nüfus eden uygulama ve servisler yaygın ağ yapısıyla beraber ağa katılan bireylerin sayısını arttırmış ve dolaylı olarak ağın değerini yükseltmiştir[19]. Bilgi teknolojileri her geçen gün daha çok kullanıcıyı kendisine çekerek, büyük kitlelere yayılmış ve günlük hayatın bilgi teknolojileri ile yürütülmesi süreci kendiliğinden gelişmiştir. İşletmeciler, son kullanıcıya benzeri olmayan platformlar, ürünler ve servisler sağlamış, aynı zamanda kullanıcı hareketlerini ve tercihlerini gerçek zamanlı takip edebilme imkanı bulmuştur. Paralel olarak, veri analizi ve yönetimi altyapılarına yatırım yaparak büyük veriyi

[18] Phil Simon, s. 71

[19] Robert Metcalce'in geliştirdiği Metcalfe Kanununa göre, bir ağın değeri ağa dahil olan ekipman sayısının(n) karesi(n^2) oranında artar. Kanunun ana fikri, ağlar yayıldığında, büyümeleri de katlanır, daha fazla bağlantı olduğundan ağa dahil olmanın yararları da katlanarak artar, ağın dışında kalmanın cezası da ağın büyümesiyle birlikte büyür, çünkü ağın dışında kalan başka unsurlara erişim fırsatları azalır temelleri üzerine kuruludur.

doğru sonuçlarla yorumlayabilen kuruluşlar, hem stratejilerini hem de şirket içi yapılanmalarını revize ederek tüketici ve kar odaklı yeni tasarımlar geliştirebilmişlerdir. Bill Schmarzo, büyük verinin özel sektör üzerindeki etkilerini incelediği kitabına[20] ambalajlı tüketim malları sektörünün, POS teknoloji ile yaşadığı gelişimi ve büyük veri sayesinde 1988 öncesi ve sonrasındaki pazar araştırması ve strateji geliştirme süreçlerindeki değişiklikleri inceleyerek başlamaktadır. POS teknolojisinde üreyen veri Proctar&Gamble, Unilever, Frito Lay ve Kraft gibi üreticilerle, Walmart, Tesco ve Vons gibi perakendecilerin arasındaki güçler dengesini değiştirmeye bile yaramıştır. POS teknolojisi öncesinde, üreticilerin dayattıkları miktarlarda, fiyatlarda ve promosyonlarla satış yapan tüketiciler, bu teknolojinin sağladığı verileri analiz ederek müşterilerinin tercihlerini daha iyi gözlemleyebilmiş, tüketicilerin hangi ürünleri ve fiyatları tercih ettiği, hangi promosyonları tercih ettiği, hangi ürünleri aynı anda aldığı gibi bilgileri edinmiş ve dengeleri tersine çevirerek satmak istedikleri ürün miktarını, fiyat ve promosyon taleplerini netleştirerek üreticilere baskı yapar konuma geçmişlerdir. Kitapta, büyük verinin işletmeleri, iş takibi, iş anlayışı, optimizasyon, verinin ticarileştirilmesi ve süreçlerin başkalaşımı hususunda etkilediği vurgulanmaktadır. Yazar, işletmeciler için kritik olan en kıymetli müşterilerin ve en verimli çalışanların kimler, en önemli ürünlerin, en başarılı kampanyaların, en başarılı satış kanallarının hangileri olduğu gibi soruların hala önemini aynı şekilde koruduğunu, ancak, büyük verinin işletmecilere yeni metrikler ve oranlar kazandırarak performansı arttıracak daha doğru tahminler yaptırdığını savunmaktadır. Bunların yanı sıra kitapta bir diğer önemli vurgu da, verinin kendisinin ticari değeri ve verinin şirket içi organizasyonlarda ve sektörde yarattığı değişimlerdir. Şirketler, organizasyon yapılarında veri analizi ve veri madenciliği yapılanmalarına yatırım yaparak kendi veri yönetimi yaşam süreçlerini oluşturmuşlardır. Böylece, "verinin özelliklerinden eğilimlerini anlamaya, çok büyük miktarlardaki veriden yeni ve anlamlı bilgiler üretmeye, verinin modellemelerle enformasyon ve bilgiye

[20] Bill Schmarzo, *Big Data:Understanding How Data Powers Big Business*, Wiley, 2013

dönüşmesini sağlamaya ve bilgiyi eyleme yönelik değerlendirmeye yarayan[21]" veri madenciliği, başlı başına bir sektör haline dönüşmüştür. Bu durum ise yeni sektörleri, organizasyon yapılanmalarını ve iş modellerini doğurarak inovasyonu sağlamıştır. Davranışsal pazarlama süreçleri, web sayfalarında alan ve tıklama sayıları üzerinden yürütülen yeni reklam stratejileri, e-ticaret firmaları ile değişen tüketici ve tüketim anlayışı, lokasyon bazlı uygulamalar, sosyal medya, konaklama ve ulaşım planlama uygulamaları, akıllı ev teknolojileri, günlük hayatı ölçmeye yarayan takip uygulamaları (yeme alışkanlıkları, spor, günlük takvim...vs.), e-kitaplar gibi pek çok örnek büyük verinin yarattığı ortamda geliştirilen inovasyonlardır. Aynı şekilde inovasyon da tüm ürün ve servisleriyle, kullanıma bağlı olarak yeni veriler üretir ve büyük veriyi geri besler.

2. Politika Belirleme

Büyük verinin toplum hakkında sağladığı veri çeşitliliği toplumu anlamak, sosyal politikalar geliştirmek ve toplumsal çözümler üretmek açısından da katkılar sağlar. Çevre ve şehircilik faaliyetleri, sağlık, eğitim ve enerji sektörü gibi toplum hayatını direk etkileyen idari faaliyetler, büyük veri sayesinde farklı boyutlarda ele alınabilir ve ihtiyaçlar daha net belirlenebilir. Toplumu ve doğayı daha iyi ölçümleyen ve daha hassas tahminler üretebilen sistemler idarecilerin yönetim politikalarını da etkilemiştir.

2008 yılında Stockholm'de hayata geçirilen akıllı ulaşım sistemler projesi ile GPS teknolojisi, sensörler ve mobil şebeke kullanılarak, şehirdeki araç trafiğinin yoğunluğu gerçek zamanlı ölçeklenebilir hale gelmiştir[22]. Kurulan sistem ile saniyede 120.000 GPS noktasından akan veri işlenebilmekte ve bu bilgiler 600.000 üzerinde bağlantının bulunduğu haritaya anlık olarak eklenmekte, düzenli

[21] Şule Işınsu Özmen, s. 434

[22] Alain Biem, Eric Bouillet, Hanhua Feng, Anand Ranganathan, Anton Riabov, Olivier Verscheure, Haris Koutsopoulos,Mahmood Rahmani, Barış Güç, *Real-Time Traffic Information Management using Stream Computing*, bkz. http://sites.computer.org/debull/A10june/Anand.pdf

olarak trafik istatistikleri hesaplanarak kullanıcıların sorgularına cevap üretilebilmektedir. Taksi ve kamyonları da içeren projede taksiler her 60 saniyede bir lokasyon ve kimlik bilgileriyle beraber GPS verisi üretirken, kamyonlar her 30 saniyede bir lokasyon, kimlik ve hız bilgileriyle beraber GPS verisi üretmektedir. Veriler sayesinde en kısa yol hesaplamaları, günün saatlerine göre trafik yoğunluğu ölçümleri, yol bazlı ortalama hız tespitleri yapılabildiği gibi elde edilen veriler, hava tahminleri, kamera görüntülerinden gelen veriler, trafik kazası bildirimleri gibi verilerle birleştirilerek daha başarılı ve etkili sonuçlar sağlanabilmektedir.

Benzer bir örnek de 2013 yılında Hollanda'da hayata geçirilen ve sel kontrolü ve tüm su kaynaklarının yönetimini dönüştürecek olan Delta Projesi'dir[23]. Bu proje yağış ölçümlerini, su seviyesi ve su kalitesi ölçümlerini, barajlardan ve radarlardan gelen verileri, model tahminlerini ve aynı zamanda mevcut duruma ve geçmişe ait bent kapakları, pompalama istasyonları ve kanallardan gelen bakım bilgilerini içermektedir. Hollanda'da nüfusun %55'nin sel bölgesinde yaşadığı düşünüldüğünde, su kontrolü hem vatandaşların günlük hayatı hem tarım hem de sektörel anlamda hayati öneme sahiptir. Ek olarak sistemdeki verilerle hava tahminleri verileri birleştirilerek suyun saklanmasında, suyun deniz seviyesinde alçak arazilere yönlendirilmesinde, tuzlu suyun içme suyuna karışmasının engellenmesinde, kanalizasyon ve su kirliliğinin yönetilmesinde daha doğru kararlar üretilecektir.

Mevcut idari sorunlara çözüm bulmasının yanında büyük verinin sağladığı imkanlar yeni sosyal oluşumlara ve sivil toplum anlayışlarına da ön ayak olmuştur. Global Viral gibi misyonlarını "mikrobik dünya hakkındaki bilinci, keşifleri ve gelişimi teşvik etmek[24]" olarak belirlemiş sivil toplum örgütleri büyük veri sayesinde varlıklarını koruyabilmektedir. Örgüt, dünyadaki salgınlarla ilgili yayınlanan karmaşık raporları takip edebilmek için gelişmiş mühendislik ve

[23] Jenny Hunter, Jelmer Letterie, *IBM harnesses power of Big Data to improve Dutch flood control and water management systems*, bkz. http://www-03.ibm.com/press/us/en/pressrelease/41385.wss
[24] *About Global Viral*, http://globalviral.org/about.php

yazılım teknikleri kullanarak ağdaki açık veriyi analiz etmektedirler[25]. Büyük ölçekli perakendecilerin normal değerlerin üzerinde verdikleri ilaç siparişleri, arama motorlarında yapılan arama sorguları içinde geçen ifadeler hangi coğrafyalarda ne gibi salgınların başlamak üzere olduğunu ya da riski gözlemleyebilmek için imkan sağlamaktadır. Benzer bir durum 2010 yılında Haiti'de yaşanan kolera salgınının sosyal medya sitesi olan Twitter üzerinden hızla yayılmasıyla yaşanmıştır. Yapılan istatistik araştırmaları sonucunda "bir salgının erken aşamalarında resmi olmayan kaynaklardan elde edilen verilerin, sadece salgının gerçekleştiğine dair bilgi vermesinin yanında, salgın hastalıkların anahtar parametrelerinden biri olan hastalığın yayılma oranlarına ait tahminler üzerinden hastalığın dinamikleri hakkında da bilgi sağlayabilmektedir[26]".

Bu ve benzeri konularda büyük veri, toplumların yaşam koşullarını, eğitim politikalarını, idari işlerini, enerji kaynaklarını yönetim biçimlerini daha etkili bir noktaya taşıyacak önemli bir kaynak haline dönüşmüştür. Bütün bu süreçler düşünüldüğünde büyük verinin yalnızca özel sektöre hizmet eden ve sermaye sahiplerinin karlarını arttırmak için kullandıkları kritik bir koz olmasından öte, sosyal ve toplumsal anlamda belirleyici bir niteliği olduğunu da görmekteyiz.

3. Akademik Çıktılar ve Ar-Ge Çalışmaları

Büyük verinin bir diğer etkilediği alan ise akademik çalışmalar ve araştırma-geliştirme süreçleridir. Büyük verinin sınırsız içeriğinin, çeşitliliğinin ve çıktılarının başarısı kabul gördükçe, büyük veri yatırımları ülke politikaları haline

[25]Nathan Wolfe, Lucky Gunasekara, Zachary Bogue, *Crunching Digital Data can help the World*, 2011, bkz. http://edition.cnn.com/2011/OPINION/02/02/wolfe.gunasekara.bogue.data/index.html?_s=PM:OPINION

[26] Rumi Chunara , Jason R. Andrews, John S. Brownstein, *Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak*, The American Society of Tropical Medicine and Hygiene, 2010, s.44 bkz. http://healthmap.org/documents/Chunara_AJTMH_2012.pdf ,

dönüşerek veri madenciliği, akademik çalışmaların ve araştırma geliştirme süreçlerinin en önemli girdisi olarak değerlendirilmiştir. 2012 yılında Amerika'da 6 büyük federal kurumun katılımıyla büyük veriye 200 milyon dolay yatırım yapılmıştır. Beyaz Saray'ın Bilim ve Teknoloji Politikaları Direktörü John Holdren, "bilgi teknolojilerinin araştırma geliştirme projelerine yapılan geçmiş federal yatırımların Internet'in yaratılmasıyla sonuçlanmasına benzer şekilde, bugün hayata geçirdiğimiz inisiyatif de bilimsel keşifler, çevresel ve biyomedikal araştırmalar, eğitim ve ulusal güvenlik konularında büyük veriyi kullanma kabiliyetlerimizi dönüştürmeyi vaat etmektedir[27]" diyerek Amerikan hükümetinin, büyük veriyi kullanarak yürüteceği araştırma geliştirme çalışmalarına verdiği ehemmiyeti vurgulamıştır.

Pazar ekonomisine ve idari kadrolara çok çeşitli girdi sağlayan büyük veri aynı zamanda toplumu anlamaya çalışan, modern dünyada değişen normları ve toplumsal dinamikleri incelemeyi hedefleyen tüm bilim dallarına da girdi oluşturmaktadır. "Ahlak Madenciliği[28]" adıyla büyük veri içinden kimlik analizleri yapılmasını etik açılardan inceleyen bir çalışmaya göre, büyük veri imkanları (yeni analiz teknolojileri ve düşük maliyetli programcılık ile erişilebilen geniş bir dijital metin, söylem ve görüntüler arşivi sayesinde) ahlaki değerler üzerine araştırmalar yapan psikologlara, sosyologlara ve ahlak bilimcilere bireylerin ahlaki kimlikleri ve davranışları arasındaki bağlantıyı incelemek adına oldukça zengin imkanlar sağlamaktadır. Araştırmada aktarılan çarpıcı örneklere göre, sosyal bilimcilerin insana dair hemen hemen her konuda araştırma yapması büyük veri ile rahatlıkla sağlanmaktadır. Linkedin gibi profesyonel ilişki ağlarındaki dijital itibar ile iş ortamlarındaki etik değerlere uyma konusundaki başarısızlıklar arasındaki çakışmaların analiz edilmesi; bireylerin davranış biçimlerini tahmin etmede rolü olan temel ahlaki eğilimlerin tespit edilmesi; kişilerin sosyal ağlardaki hareketlerini inceleyerek ahlaki olarak ikiyüzlü olarak

[27] Jason Honer, *U.S. government commits big R&D money to 'Big Data',* 2012, bkz. http://www.zdnet.com/blog/btl/u-s-government-commits-big-r-and-d-money-to-big-data/72760

[28] Markus Christen, Mark Alfano, Endre Bangerter, Daniel Lapsley, *Ethical Issues of Morality Mining: Moral Identity as a Focus of Data Mining,* 2013

kabul edilen davranışlara (aldatma v.b.) karşı potansiyellerinin takip edilmesi gibi çalışma konuları örnekler arasındadır.

Bunların yanında sağlık sektöründe üreyen büyük veri vasıtasıyla hastalıklara ve tedavilere yönelik incelemelerin yanında, sağlık sistemine dair de detaylı bilgi edinilmekte, bütün bu süreçler araştırmacılar tarafından incelenerek optimize edilmekte veya yeni çözümler sunulmaktadır. Bunların yanında "araştırmacılar, makalelerinde sadece vardıkları sonuçları paylaşmayıp aynı zamanda o sonuca ulaşmalarını sağlayan veri kümelerini de paylaşarak çalışmalarının kontrol edilmesini, teyit edilmesini ve ilerletilmesini sağlayabilmektedirler[29].

Büyük veri, topluma, doğaya ve insana dair barındırdığı zengin içerikle, bilimin her alanına ve her uzmanlıktaki araştırmacıya sınırsız katkı sağlamaktadır.

C. Çalışma Kapsamındaki Önemi

Yukarıda detaylandırıldığı üzere büyük veri, çok geniş bir yelpazede ve oldukça kritik etki alanına sahiptir. Çağımızda veri odaklı gelişen hayat, büyük veriden elde edilecek çıktılar sayesinde yön bulur. Diğer taraftan da, büyük verinin sınırsız büyüklüğü ve çeşitliliği ise yeni sorunları hayatımıza taşımaktadır. Bu çalışma kapsamında büyük verinin öneminin ve etki alanının bu detayda ele alınmasının sebebi büyük verinin tetikleyici pozisyonundan kaynaklanmaktadır. Bir önceki bölümlerde görüdüğümüz üzere, büyük veri toplumsal fayda ve gelişmeyi tetiklemektedir. Ancak ilerleyen bölümlerde göreceğimiz üzere aynı ölçekte, ihlal ve güvenlik açıklarını da tetiklemektedir. Büyük veriden faydalanılmasını sağlayacak veri işleme mekanizmaları çok çeşitli ve geniş ölçekli verileri saklayarak, paylaşarak, analiz ederek yeni güvenlik ve gizlilik ihlallerine ön ayak olabilmektedir. Bu haliyle büyük veri, fayda ve gizlilik arasındaki tartışmanın bugünkü boyutuna ulaşmasında temel nedendir.

[29] Andrew Oram, *The Information Technology Fix For Health*, OReilly, 2014, s.3

III. Veri İşleme

Bu bölüm, veri işlemenin dinamiklerine, bilginin yaşam döngüsüne, büyük verinin işlenmesi sürecindeki farklılıklara, ortak noktalara ve en önemlisi de işlemenin iki önemli alt başlığı olan paylaşım ve ifşa süreçlerine odaklanacaktır. Burada hedef, veri işlemenin altyapısını tanıttıktan sonra büyük verinin işlenmesiyle paylaşım ve ifşa süreçlerinin ne ölçüde ilişkili olduğunu analiz edebilmektedir. Çalışmanın amacına uygun olarak paylaşım ve ifşa büyük veri süreçlerinin yaygınlaştırdığı ve gerekli kıldığı alt fonsiyonlar olarak karşımıza çıkmaktadır.

A. Tanım ve İçerik

Bilginin yaşam döngüsünu oluşturan temel adımlar "toplama, kullanma, ifşa, depolama ve yok etme[30]" olarak karşımıza çıkmaktadır. Veri işleme ise bilginin hayat döngüsünde yer alan tüm bu fonksiyonları içeren işlemlerdir. Veriden anlamlı çıktılar üretebilmek, veriyi kaydetmek, saklamak, raporlamak, paylaşmak, aktarmak, analiz etmek, yorumlamak gibi işlemlerin hepsini hayata geçirebilmeyi gerektirir. Bu bağlamda, ilerde inceleyeceğimiz gibi veri işleme süreci, hukuksal metinlerde de oldukça kapsamlı fonksiyonlar kümesinden oluşmaktadır.

Bilgi ve enformasyon kavramlarına değinilenz önceki bölümlerde, enformasyonun bilginin hammaddesi olduğu işlenmişti. Buna istinaden, bilgiye ulaşmanın yolunun enformasyondan geçtiğini iddia etmek yanlış olmayacaktır. "Enformasyon işlenmiş verilerdir, […], verilerin toplandıktan sonra sınıflandırılması, ortalama, mod, standart sapma ve benzeri istatistiksel ölçümlerle özetlenmesi, grafiksel olarak sunulması, istatiksel ve matematiksel yöntemlerle

[30] Peter P. Swire, Kenesa Ahmad, *Foundations of Information Privacy and Data Protection*, IAPP,2012, s. 13

analiz edilerek anlamlandırılması, çeşitli değişkenlerin birbiriyle ilişkisi olup olmadığının tespit edilmesidir[31].” Yani enformasyone ulaşabilmek için verinin işlenmesi gerektiği sonucu çıkar. Özetle, bilgiye de enformasyon sayesinde ulaşıyorsak, bilgi ancak ve ancak verinin işlenmesi sayesinde elde edilebilecektir.

Verini işlenmesi, veriden elde edilecek bilgi için ön şart olduğundan, özellikle günümüzün teknoloji çağında en önemli süreçlerden biri haline gelmiştir. Veri türlerinin incelendiği önceki bölümlerde, verinin yapılandırılmış veya yapılandırılmamış özellikler göstermesinin veri yönetimi ve veri işlemesi süreçlerini etkilediği tartışılmıştı. Yapılandırılmış ve daha kontrol edilebilir özellikler gösteren veriler alışılagelmiş ilişkisel veri tabanları sayesinde ve kapasitesi daha düşük altyapılarla işlenebilirken, yapılandırılmamış veriler için daha karmaşık altyapılar gerekmektedir. Özellikle işleme hızı, hacmi ve çeşitliliği ile benzersiz özellikler gösteren büyük veri için veri tabanları yerine “tüm operasyonel işlemlerin en alt düzeydeki verilerine kadar inebilen, analiz yapabilmek için özel olarak modellenen, tarihsel derinliği olan, fiziksel ve mantıksal olarak operasyonel sistemlerden farklı ortamdaki yapı üzerinde gerçekleşen[32]” süreçlerden oluşan veri ambarlarına ihtiyaç bulunur.

Basit bir örnekle açıklayabilecek olursak, bir bankanın tüm şubelerinden gelen müşteri bilgileri güncellemeleri ile bir otoban üzerinde kurulan ve trafikteki hız ölçümlerini sağlamak için araçların anlık hızlarını hesaplayan bir sensörün ürettiği verilerin işlenmesini aynı altyapılarla sağlayamayız. Ne kadar yoğunluk olursa olsun, bankalarda müşterilerin bilgileri personel tarafından belli ara yüzlerden ve kişi odaklı olması sebebiyle belli bir hızda gerçekleşebilir. Bu durumda üreyen veri, bankanın müşteri sayısıyla da orantılı olarak depolama kapasitesi yüksek olan ama yine de geleneksel olarak ilişkisel yapıda çalışan veri tabanlarında işlenir. Ancak akan trafikte, özellikle de yoğun saatler düşünüldüğünde, araçların hız değerlerini anlık hesaplayan ve işleyen bir sensörün ürettiği veri çok yüksek hızlara ve hacime ulaşabilir. Bu bağlamda ilişkisel veri tabanları yerine öncelikle

[31] Şule Işınsu Özmen, s. 409
[32] Şule Işınsu Özmen, s. 411

üreyen veriyi işlenebilecek anlamlı formata sokabilen ve sonra yüksek hacimli depolama sistemlerinin olduğu altyapılar gerekecektir.

B. Büyük Verinin İşlenmesi

Büyük verinin hacmi, hızı ve çeşitliliği öncelikle üreyen verinin anlamlı bir formata getirilmesi sorununu doğurmaktadır. Milyonlarca satırdan oluşan ve içinde pek çok anlamsız karakteri barındırabilen verileri, geleneksel bir yöntemle ayrıştırarak veri tabanlarına eş zamanlı olarak aktarmaya çalışmak başarısız sonuçlar üretecektir. Bu sebeple öncelikle büyük verinin toplanması, küçük parçalara ayrıştırılması ve eş zamanlı olarak gruplanabilmesini sağlayan yüksek kapsiteli yazılımlar geliştirilmiştir. Bunlara bir örnek Hadoop yazılımı verilebilir. Hadoop açık kaynak kod projesi olarak geliştirilmiştir ve dağıtılmış dosya sistemi ve MapReduce motoru diye adlandırılan iki bileşenden oluşmaktadır. Dağıtılmış dosya sistemi "makinalar arası ilşkili dosyaların yönetimini sağlayan güvenilir, yüksek bant genişliğine sahip, düşük maliyetli veri depolama sunucusu" iken, MapReduce motoru "yüksek performanslı paralel/dağıtılmış veri işleme algoritmasıdır"[33]. "Hadoop, düğüm noktaları arasındaki veri işleme sürecini paralelliştirmeyi, işlemeyi hızlandırmayı ve gecikmeleri gizlemeyi sağlamak için tasarlanmıştır ve hatalar da dahil olmak üzere değişiklikleri tespit edip ayarlayarak operasyonu kesintisiz devam ettirmektedir[34]."

Veriden sonuç alabilmek için veri analizi ve veri madenciliği yapılması şarttır. Ancak bu noktada veri madenciliğinin gerçekleştirebilmek için veri ambarının "ilişkili durumları bağlayabilecek kadar düzenlenmiş yapıda olması, uygulanabilir tüm perasyonları içermesi, güncel ve tutarlı bilgiye sahip olması[35]" gerekmektedir. Bu durumda büyük verinin yapılandırılmamış özellikler gösteriyor olması öncelikle Hadoop gibi uygulamalarla veri madenciliğine uygun hale

[33] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, s. 112
[34] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, s. 112
[35] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, s. 130

getirilmesini gerektirir. Bu sebepledir ki, büyük verinin işlenmesi öncelikle verinin formatının düzenlenmesini gerektiren bu ara katman uygulamalara ihtiyaç duyar. Daha sonrasında ise hibrit mekanizmalarla, büyük verinin işlendiği kaynak sunucular, diğer ilişkisel veri tabanları ile birleştirilirek veri ambarı yapısı oluşturulur ve veri madenciliği yapılır. Görüldüğü üzere büyük verinin işlenmesindeki fark, öncelikle hacim, hız ve çeşitliliği yüksek olan veri evreninin işlenebilir ve anlamlı çıktılar üretilebilecek bir formata dönüştürebilmesi ve gruplandırılabilmesidir. Daha sonrasında talebe göre tek başına ele alınabilir veya diğer yapılandırılmış veri kümeleriyle birleştirilebilir.

C. İşlemenin İki Önemli Fonksiyonu: Paylaşım ve İfşa

Büyük verinin işlenmesini sağlayan sistemler ve veri madenciliğindeki gelişmeler veriden anlamlı çıktılar üretilmesinin en kritik nedenidir. Veri madenciliği ile yapılandırılmış veya yapılandırılmamış olsun tüm veri kümeleri birleştirilerek yorumlanabilir bilgiye dönüşmektedir. Bu noktada da veri işlemenin iki önemli alt fonksiyonu daha fazla önem kanmaktadır; yorumlanabilir bir hale getirilmiş verinin paylaşım ve ifşası. Paylaşım ve ifşa araştırmaların kontrol edilmesi, yeni araştırmaların önünün açılması, yeni ürün ve servislerin geliştirilebilmesi, yeni politikaların yaratılması gibi tüm fayda sağlayıcı aktiviteler için gerekli olduğu kadar, kurumların, devlet organlarının, işletmelerin devamlılığı için ihtiyaç duyulan veri akışını da sağlamaktadır. Veri ancak paylaşım ve ifşa yoluyla kullanıma açılabilir.

Burada vurgulanması gereken, paylaşım ve ifşanın büyük veri süreçleriyle kazandığı önemdir. Büyük veri tüm veri odaklı yaklaşımları tetiklediği gibi paylaşım ve ifşa süreçlerini de kaçınılmaz bir adım haline dönüştürmüştür. Büyük verini sağladığı geniş bilgi yelpazesi, bu bilginin farklı partiler arasında paylaşılması veya genele ifşa edilmesi sürecinin önemini hızla geliştirmiştir. Bu

sebeple bu bölüm, veri işlemenin bu iki önemli alt fonksiyonuna ve sürecin gerçekleştiği taraflara odaklanacaktır.

1. Sektörel bazlı paylaşımlar

Kurumlar hizmet verdikleri sektöre göre (kamu veya özel sektör fark etmeksizin) önemli veri üreticileri konumuna gelmişlerdir. Özellikle kamu kurumları, bankalar, iletişim şirketleri (operatörler, internet servis sağlayıcılar, altyapı sağlayıcılar v.b), e-ticaret şirketleri (amazon, ebay v.b) gibi günlük operasyonu oldukça yoğun olan kuruluşların ürettiği ve sakladığı operasyonel veriler büyük verinin önemli parçalarını oluşturmaktadır. Bu şirketler pek tabi ki üretilen veriden maksimum faydayı sağlayabilmek için çeşitli katmanlarda bu veriyi paylaşır veya ifşa ederler. Bu paylaşımlar kurum içinde olabildiği gibi, kurumlar arasında da gerçekleşebilmektedir.

a) Birimler Arası

İşletmenin veya kuruluşun kendi müşterilerine, kullanıcılarına, operasyon kayıtlarına, ürünlerine v.b dair verilerini organizasyon içi birimleri arasında farklı sebeplerle paylaştığı durumlara karşılık gelmektedir. Veri, bir şirket için kendini ölçekleyebilmesi ve devamlılığını sağlayacak stratejileri geliştirmesi için en önemli kriterdir. Bu bağlamda veri güvenliği ve gizliliği sağlanarak verilerin şirket içindeki farklı birimler arasında paylaşılması esastır. Örnekleyecek olursak, bir havayolu şirketinin etkin pazarlama stratejileri geliştirmesi için müşterileri profillerini, tercih ettikleri coğrafyaları, tercih edilen fiyat aralıklarını kullanıcı bazlı olarak pazarlama ve strateji birimleriyle paylaşması gerekir. Diğer taraftan başarılı müşteri yönetimi ve ihtilaf çözüm mekanizmaları da yine şikâyet

konularının, içeriklerinin, yoğun şikayet eden müşteri profillerinin ilgili birimlerle paylaşılması ile gerçekleşir. Ek olarak, şirketin iç süreçlerinin iyileştirilmesi, çalışan profilleri ve performanslarının ölçeklenmesi gibi süreçleri gibi çalışanlara dair her türlü verinin insan kaynakları ve organizasyon birimleriyle paylaşılması gerekir.

Hatta, veri madenciliği uygulamlarında "departmanların kendi kullanım amaçlarına hizmet edecek şekilde ayrılmasıyla "data mart" olarak isimlendirilen her departmana özel veri tabanları[36]" bile oluşturulabilmektedir.

b) Şirketler Arası

Şirket birleşmeleri, ürün ve hizmet bazlı ortak projeler, 3. parti ilişkileri ve sektör/tüketici yararı gibi sebeplerle şirketler arasında veri paylaşımları sıklıkla gerçekleşmektedir. Şirket birleşimleri ve satın almaları süreçlerinde şirketlerin birbirlerine dair personel bilgileri, ürün ve hizmet bilgileri, finansal tabloları, ticari sırları, pazar ve müşteri verileri gibi pek çok verisi kendi içlerinde paylaşılabilmektedir. Küreselleşmiş ticari dünyada farklı ülkelere ait şirketlerin birleşmesi veya ortak projeler yürütmesi halinde paylaşılan verilerin kültürel ve coğrafi özellikler de içerebilmektedir. Şirketler arası veri paylaşımının en yoğun yaşandığı süreçler 3. parti ilişkileridir. Günümüzde şirketler, kurulum ve yönetim maliyetlerine katlanmak istemedikleri süreçleri, ilgili konuda hizmet veren 3. parti şirketlere proje olarak atayabilmektedir. Personel yönetimi, finansal veya hukuksal danışmanlık, ürün tasarımı, pazarlama stratejileri geliştirilmesi gibi pek çok konuda 3. parti şirketlerle ortak projeler yürütülebilmektedir.

Bu çalışmanın konusuna en uygun örnek ise veri madenciliği ve büyük veri yönetimi konusunda çalışan şirketlerin, özellikle veri işleme ve saklama hacmi yüksek büyük ölçekli işletmelerle kurduğu 3. parti ilişkilerdir. 3. parti süreçlerinde

[36] Şule Işınsu Özmen, s. 407

asıl şirket, sahip olduğu verileri, hizmet almayı kabul ettiği 3. parti şirket ile paylaşarak arada yapılan iş anlaşmasına uygun şekilde bu şirketten hizmet almayı bekler. Örneğin, bir banka yüksek hacimli işlemlerinin yorumlanması ve yeni bankacılık ürünleri tasarlanması için bu konuda uzmanlaşmış bir pazarlama şirketiyle anlaşarak, belli periyotlara ait müşteri ve ürün bilgilerini bu şirketle paylaşabilir ve bu veriden anlamlı çıktılar üreterek kendi müşterilerine en uygun hizmet veya ürünü tasarlaması için anlaşabilir.

c) Hukuksal Yükümlülüklere İstinaden Paylaşımlar

İlerleyen bölümlerde detaylandırılacak olan veri koruması, saklaması, gizliliği, güvenliği gibi hususlardaki hukuksal yükümlülüklere istinaden özel sektör işletmecileri ve kamu kurumları talep edilmesi halinde resmi makamlarla her türlü veriyi paylaşma yükümlülüğündedir. Ne tip verilerin, hangi koşullarda isteneceği ulusal ve uluslar arası hukuk mevzuatında veya konuyla ilgili yürürlükte olan veri koruması hukukunda düzenlenir. Günümüzdeki en güncel örnek mahkeme veya savcılık talepleriyle telefon görüşmesi kayıtlarının mobil veya sabit operatörlerden talep edilmesidir. Bunlara ek olarak, sektörel araştırmalar için gerekli olan, müşteri, ürün, servis, altyapı, finansan veriler gibi şirkete dair pek çok detaylı idari otoritelerle düzenli olarak paylaşılmaktadır.

2. Kamu Geneline Yapılan İfşalar

Özellikle araştırma geliştirme süreçlerinin devamlılığı adına, verilerin kamuya ifşa edilmesi büyük önem taşımaktadır. İfşa edilen veriler herhangi bir başka kurum, işletme veya enstitü tarafından kullanılarak yeni tasarım ve buluşların önünü açacaktır. Amerika'da her yıl sağlık bakanlığı sağlık verilerini uygun formatlarda

ifşa etmektedir. Böylece bu verileri kullanan aktivistlere veya girişimcilere enstitüye direk iletişime geçme ve başka verileri de ekleme imkanı sağlamaktadır. Sonrasında bu veriler devlet seviyesinde yüksek riskli alanları belirleme, hastalara en uygun sağlık hizmeti sağlama gibi alanlarda kullanılmaktadır. Bu tip uygulamalara en güzel örnek acil servislerdeki bekleme sürelerini trafik raporlarıyla birleştirerek en hızlı hizmet alınabilecek acil servis birimlerinin tespit edildiği uygulamadır[37].

3. Uluslararası Güvenlik Gerekçeleri

Uluslararası güvenlik gerekçesiyle veri paylaşımlarının en belirgin örneği, terörle mücadele kapsamında AB ile ABD arasında paylaşılan havayolu taşımacılığına dair verilerdir. 11 Eylül saldırıları sonrasında da yeniden tasarlanan mevzuata göre "yabancı firmalar ABD Gümrükler Komisyonu'na iki çeşit yolcu verisi aktarmakla yükümlüdürler: […] havayolu şirketleri ve bunların uçuşları ile listelenen yolcuların kimlikleri ve spesifik güzergahları ile hava yolu firmalarınca daha iyi müşteri hizmeti sunmak amacıyla toplanan verilerden oluşan yolcu isim kayıtlarıdır[38]". Bu tip paylaşımlar, hassas veriler içeriyor dahi olsa ulusal ve uluslar arası güvenlik gerekçeleriyle yapılmaktadır.

IV. Kişisel Verilerin Korunması ve Özel Hayatın Gizliliği

Gizlilik ve kişisel verilerin korunmasına istinaden en temel metinlerden biri sayılan, 1890 tarihli "Gizlilik Hakkı[39]" isimli makaleye Warren ve Brandeis şu ifadeyle başlar, "Kişinin şahsen ve tüm özellikleriyle koruma altına alınıyor

[37] Andrew Oram, *The Information Technology Fix For Health*, OReilly, 2014, s. 21
[38] İkbal Gür, *Kişisel Verilerin Korunması Hususunda AB ile ABD Arasında Çıkan Uyuşmazlıklar*, Turhan Kitabevi, 2010, s. 223
[39] Samuel D. Warren, Louis D. Brandeis, *The Right to Privacy,* Harvard Law Review, 1890

olması *örf adet hukuku*[40] kadar eski bir prensiptir; ancak zaman zaman bu korumanın niteliğini ve kapsamını yeniden tanımlamak gerekli hale gelmiştir[41]". Makalede, basının ve yeni fotoğraf teknolojilerinin hızlandırdığı fotoğraf ve dedikodu sirkülasyonu ile değişen gizlilik ve ihlal anlayışı tartışılmakta ve yaşama hakkı artık "hayatın tadını çıkartmak, yalnız kalabilme hakkı, kapsamlı sivil önceliklerin icra edilmesini güvence altına alan özgürlükler[42]" olarak tanımlanırken, "mülkiyet tanımı her tip aidiyeti içeren[43]" bir tanım olarak ifade edilmektedir. Çalışmadaki diğer önemli vurgu ise, "yeni oluşan kültürün etkisiyle herkesçe bilinme ve tanınmaya daha duyarlı hale gelen kişi için, gizliliğin daha gerekli hale geldiği ve modern kurumların ve müdahalelerin kişiyi fiziksel zarardan çok daha fazla etkileyecek olan psişik acı ve strese sürüklediği gerçeğidir[44]". Bugün gelinen noktada ise büyük veri sayesinde süreci hızlanan paylaşım ve ifşa mekanizmaları hayatlarımıza yeni gizlilik sorunlarını taşımıştır. Buna ek olarak, e-devlet uygulmalarıyla kişilerin vatandaşlık haklarına ve bilgilerine dair pek çok işlem otomasyon sistemleri sayesinde ve birbirine bağlı ağlar üzerinden yürütülmektedir. E-devlet girişimleri, çoğunlukla daha önceleri ayrık ortamlarda tutulan ve vatandaşların kişisel verilerini içeren veri tabanlarını bağlayıp merkezileştirmekte ve kişiye ait daha detaylı bir profil oluşturma imkanını elde etmektedir[45]. Bütün bu yeni uygulamalar ve otomasyon sistemleri, kişinin daha rahat takip edilebilmesine, gözetlenebilmesine, hayatına dair daha fazla detayı depolabilmeye yaramaktadır. Lessig, bu noktada fiziksel hayattaki hareketler ile bilgi ve iletişim teknolojilerinin yarattığı ağlar içindeki hareketler arasında bir karşılaştırma yapar. Fiziksel hayatta kamuya karıştığımız her durumda izlenebildiğimizi ancak bu izleme sonucunda sadece bizi izleyen kişilerin akıllarında bir kayıt oluşturduğumuzu, diğer taraftan, ağ içindeki her hareketimizin depolanması sonucu farklı veri tabanlarında veya ortamlarda, daha sonra üzerinde araştırma yapılabilecek kayıtlar ürediğini vurgular. Burada ifade

[40] "Common Law" ifadesinin çevirisi olarak kullanılmıştır.
[41] Samuel D. Warren, Louis D. Brandeis, s. 193
[42] Samuel D. Warren, Louis D. Brandeis, s. 193
[43] Samuel D. Warren, Louis D. Brandeis, s. 193
[44] Samuel D. Warren, Louis D. Brandeis, s. 196
[45] Ian Brown, Christopher T. Marsden, s. 49

edilmek istenen, fiziksel hayattaki eylemlerimize dair bizi izleyenler tarafından sonradan araştırılabileceği ortamlar oluşmazken, ağ içindeki hareketler kaydedildiğinden, sonradan analiz edilebilecek veri kümelerine dönüşebilirler. Ağ daha çok hareketi sadece izlenebilir yapmanın dışında, daha çok hareketi de araştırılabilir hale getirmektedir[46]. Kendisi hakkında daha fazla verinin kendi takibi ve kontrolü dışında paylaşılıyor ve ifşa ediliyor olması ise kişinin özerkliğine tehdit anlamına gelmektedir. Küzeci, buradaki önemi vurgulamak için şu ifadeyi kullanır, "bize ilişkin olanın üzerindeki denetimimizi kaybettiğimiz noktada kendimiz olmamız ve bireysel özerkliğimizi korumamız da olanak dışına çıkacaktır […] sürekli izlenen, gözlenen, davranışları yönlendirilen bireyin özgürlüğünden de söz edilemez[47]." Morozov, "Ağ Çözülmesi: Dünyayı Nasıl Liberalleştiremeyiz[48]" isimli kitabında ifade özgürlüğü konusunda benzer bir tartışma yürütmektedir. Buna göre, Internet ve bilgi teknolojilerinin sağladığı ortamlar ile toplumların daha demokratik ve özgür olabileleceğini ve özellikle otoriter hükümetlerin görevde olduğu ülkelerde köklü sosyal ve politik ilerlemelerin yaşanacağını iddia eden yaygın kanıya karşı çıkarak, aynı teknolojilerin hükümetlerin de faydasına çalışabileceğini çarpıcı örneklerle gösterir. Böylece toplumu özgürleştirmesi beklenen teknoloji, güçlüyü daha güçlü hale getirerek, benzeri olmayan gözetleme, sansür ve propoganda mekanizmalarıyle daha kısıtlayıcı bir toplum yaratabilir. Böyle bir durum gizlilik tartışması için de geçerlidir. Veri, faydalı pek çok gelişime ön ayak olabileceği gibi, içinde bulundurduğu kişisel veriler yanlış gruplar tarafından yanlış kullanılmaları halinde kişilere benzersiz zararlar verebilir ve kişinin en önemli değerlerinden biri olan özel alan gizliliğini ihlal edebilirler.

Bu noktada kişisel veri kavramına ve büyük verinin içerdiği kişisel veriye değinmekte fayda var. Kişisel veri kavramının farklı ülke mevzuatlarında farklı tanımlamalarına rastlasak da, AB mevzuatındaki en geniş kapsamlı tanımı "kişinin kimliğini saptayan veya saptayabilen bütün veriler" olarak geçmektedir.

---

[46] Lawrance Lessig, The Code Version 2.0, Basic Books, 1996, s. 203
[47] Elif Küzeci, Giriş
[48] Evgeny Morozov, *The Net Delusion: How not to Liberate World*, Penguin Books, 2011

İlerideki bölümlerde daha detaylı işleyeceğimiz bu kavram, kişiyle ilişkilendirilebilen her tip veriyi kapsar. Büyük verinin geniş hacmi ve çeşitliği düşünüldüğünde ise içinde kişisel veri barındırması kaçınılmazdır. Ancak büyük verinin bir diğer özelliği ve yarattığı tehdit, içeriğinde çok çeşitli kaynaklardan çok türde veri bulundurmasından ötürü, bu verilerin bir araya gelmesinden kişinin kimliğini saptayabilecek kombinasyonlar oluşabilmektedir. Buradan büyük verinin her durumda kişisel veri barındırdığı sonucuna varamayız. Ancak odaklanmamız gereken, çeşitliliğin kişiel veri özelliği göstermeyen verileri bile birleştirerek kişisel veri özelliğine taşıyabildiği gerçeğidir. Bu haliyle büyük veri paylaşım ve ifşa süreçlerini tetiklediği gibi, içeriğiyle de kişisel verilerin ve gizliliğin ihlalini de tetikleyecek bir boyut kazanmaktadır.

Kişisel verilerin ve gizliliğin ihlal edilmesi halnde oluşabilecek zararlar arasında, kişinin rahatsız hissetmesi, hareketlerinin ve kararlarının etkilenmesi veya otokontrol, kişinin başkaları tarafından yargılanması, hükümet nezdinde kimliğinin açığa çıkması veya anonimlik hakkının ihlali, kimlik hırsızlığı ve finansal kayıplar, verilerinin gelecekteki kullanımlarına dair korku ve güvensizlik geliştirme, ifade özgürlüğünün ihlali, kişinin kendi geçmişine ait kayıtlarına mahkum olması, kişinin mahremine ait müstehcen, utanç verici veya aşağılayıcı gerçeklerin yayılması ve haysiyetinin zarar görmesi, şantaja maruz kalması maddeleri sayılabilir[49].

Büyük veriden sağlanan fayda ve verinin yeni ekonomik düzlemde oynadığı temel rol, aynı ölçüde yaratacağı büyük gizlilik sorunları düşünülünce ortaya dengelenmesi gereken bir tablo çıkmaktadır. Veri odaklı çalışan ve hayatımızı kolaylaştıran, kamu yararına pek çok kritik proje üreten, toplumları daha iyi anlamamızı sağlayan büyük verinin, aynı ölçekte sebep olabileceği gizlilik ve güvenlik tehditleri düşünülünce, ancak doğru kişisel veri koruması politikaları ve uygulamalarıyla toplumsal refah ve güven sağlanabilir. Günümüz bilgi toplumunda kişisel verilerin koruması politikaları kaçınılmazdır ve hem kişileri hem veriden elde edilecek faydayı koruyan bir nitelikte olmaları gerekir. Bu

[49] Daniel J. Solove, *A Taxonomy of Privacy*, Pennsylvania Law Review, Vol. 154, No.3, 2006

politikalar, bir yönüyle "etki olanağını kaybetmiş bireye, yeniden denetim hakkını verir[50]" ve "verilerin korunması temelde verilerin değil, bu verilerle ilişkili olduğu kişilerin korunmasını hedef alır[51]", diğer taraftan veri işleme yöntemlerindeki esasları belirleyerek hem ticari dinamiklerin hem de demokratik toplumun temellerini korumayı hedefler.

A. Kişisel Verilerin Korunmasında Avrupa Birliği Ve Türkiye'deki Yasal Düzenlemelere Genel Bakış

Bu bölümde incelenen kanunlar, yönetmelikler ve yönergeler Avrupa Birliği'nde ve Türkiye'de yürürlükte olan ve kişisel verilerin korunması hususunda temelleri belirleyen mevzuattır.

1. Avrupa Birliği

1970'li yıllardan itibaren kişisel verilerin korunmasına istinaden Avrupa'da yapılmaya başlanan düzenlemelerin, 1990'lı yıllara gelindiğinde gelişen iletişim teknolojileri ile önemi daha da artmıştır. Buna istinaden "farklı ulusal yasaların tek bir iç pazar oluşumunu engelleme tehlikesinin önüne geçilmesi[52]" amacıyla "zorlayıcı[53]" nitelikte yasal düzenlemeler yapılma yoluna gidilmiştir.

Köklü ve sürekli kendini revize eden bir sistematik içinde Avrupa Birliği yönergeleri kişisel veri kavramını, kişisel veirlerin korunmasına dair usul ve esasları belirleyerek veri aktarımı ilişkilerinden ötürü dünyanın diğer ülkelerindeki yapılanmaları da etkilemektedir. İlerleyen bölümde tarih sırasıyla veri koruması hususundaki kritik AB direktifleri genel olarak incelenmiştir.

[50] Elif Küzeci, s. 13
[51] Elif Küzeci, s. 13
[52] Elif Küzeci, s.164
[53] Elif Küzeci, s.166

a) 95/46/AT sayılı Kişisel Verilerin Korunması Yönergesi

Avrupa Birliği'nde kişisel verilerin işlenmesi, korunması, paylaşımı hususundaki en temel metin "Kişilerin verilerin işlenmesi ve bu türdeki verilerin serbest dolaşımı bağlamında bireylerin korunmasına ilişkin 24 Ekim 1995 tarihli ve 95/46/AT sayılı Avrupa Parlamentosu ve Konseyi Yönergesi[54]"dir. Yönergenin kapsamına genel olarak bakıldığında, koruma konusunun yalnızca gerçek kişiler olduğu ve tüzel kişilerin hariç bırakıldığını[55],veri işleme sürecine dair aktörlerin tanımlandığını[56], kişisel verilerin kısmen veya tamamen, otomatik olan veya olmayan veri işleme süreçlerinden geçirildiği süreçleri kapsadığını[57] , özel kategorilerle hassas veri kavramının tanıtıldığı ve bu kategorideki verilerin işleme süreçlerinin ayrıca belirlendiğini[58], üçüncü ülkelere veri aktarımında asgari önlemlerin vurgulandığını[59] görmekteyiz.

Yönergenin tanımlar bölümünde öncelikle kişisel veri, veri işleyen, veri işlem sorumlusu , ilgili kişi, üçüncü parti kavramları netleştirilmiştir. Buna istinaden kişisel veri "bir kimlik numarasına istinaden veya fiziksel, psikolojik, psişik, ekonomik, kültürel veya sosyal kimliğe bağlı olarak direk veya dolaylı yollarla kimliği saptanmış veya saptabilen bir kişiye ait olan tüm verilerdir[60]". Görüldüğü üzere yönergede, kişisel veri kavramı oldukça geniş tanımlanmıştır. Kişinin

---

[54] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Oficial Journal of the European Communities of 23 November 1995, No L. 281, s. 31.
Bundan böyle "95/46/AT sayılı Yönerge" olarak anılacaktır.
[55] 95/46/AT sayılı Yönerge, m. 1/1
[56] 95/46/AT sayılı Yönerge, m. 2
[57] 95/46/AT sayılı Yönerge, m. 3/1
[58] 95/46/AT sayılı Yönerge, m. 8
[59] 95/46/AT sayılı Yönerge, m. 25
[60] 95/46/AT sayılı Yönerge, m. 2/a

belirlenebilir olması kişinin dosya numarası, işlem numarası, kişisel işareti vs. gibi tanımlama bilgileriyle kimliğinin ortaya çıkarılabilmesidir[61].

Yönergenin en önemli bölümleri veri işleme fonksiyonlarının ve istisnalarının tanıtıldığı kısımlardır. Buna göre veri işleme sürecinin çerçevesi belirlenmiş olur. Yönergeye göre veri işleme fonksiyonu, otomatik bir sistemle olsun veya olmasın, tüm toplama, kaydetme, organize etme, depolama, uyarlama veya değiştirme, kurtarma, danışma, kullanma, aktarım yoluyla ifşa, dağıtma veya başka yollarla erişilebilir kılma, ayarlama veya birleştirme, engelleme, silme veya yok etme işlemlerinin hepsini içerir[62]. Yönerge uyarınca kişisel veriler ancak ilgili kişinin açık rızası ile ilgili kişinin taraf olduğu bir sözleşmenin veya veri işlem sorumlusunun hukuksal bir yükümlülüğünün yerine getirilmesi, ilgili kişinin hayati çıkarlarının korunması, kamu yararı bulunan bir görevin yerine getirilmesi, veri işlem sorumlusu tarafından korunması gerekli olan haklı bir yararın bulunması durumlarında işlenebilecektir[63].

Yönerge'de aynı zamanda özel kategorilerde sıralanmış "ırksal ve etnik köken, politik düşünce, dini ve felsefi inançlar, ticari üyelikler, sağlık ve cinsel tercihlere yönelik bilgiler" gibi hassas verilerin işlenmesi yasaklanmıştır. Bu veriler ancak ilgili kişinin açık rızası, ilgili kişinin hayati fonksiyonlarını koruma amacı, yeterli güvence sağlanarak resmi faaliyetler amacıyla, ilgili kişinin açıkça kamuyla paylaşmasına istinaden veya hukuki iddiaları uygulama ve savunma amacıyla işlenebilir[64].

Yönergede, gerçek kişilerin kişisel verilerinin her türlü işlemeye karşı korunmasının yanı sıra, aynı zamanda, kişisel verilerin serbest dolaşımı da düzenlenmiş ve kişisel verilerin, sadece Birliğe üye ülkeler arasında değil, aynı zamanda Birliğe üye olmayan ülkeler arasında da gerçekleşecek ver

[61] Oğuz Şimşek, *Anayasa Hukukunda Kişisel Verilerin Korunması*, Beta Basım, 2008, s. 43
[62] 95/46/AT sayılı Yönerge, m. 2/b
[63] Oğuz Şimşek, s. 46
[64] 95/46/AT sayılı Yönerge, m. 8/b, 8/c, 8/d, 8/e

transferlerinde uyulacak kuralları detaylı bir şekilde düzenlemiştir[65]. Yönerge'ye göre, veri aktarımı ancak "yeterli düzeyde koruma[66]" sağlayan ülkeler arasında yapılabilir. Bu sıkı kural, ülkeler arası veri aktarımlarını imkansız bir hale getirmemek adına bazı istisnalarla gevşetilmiştir. Buna göre, ilgili kişinin açık rızası, bir sözleşmenin ifası için gereklilik, ilgili kişinin çıkarlarının korunması doğrultusunda bir sözleşmenin sonuçlandırılması ve uygulanması için gereklilik, önemli bir kamusal çıkarın korunması, ilgili kişinin hayati çıkarının korunması, aktarımın herkese açık olan kamu sicillerinden yapılıyor olması istisnaları kapsamında ülkeler arası veri aktarımları sağlanabilir[67]. Bu ilkenin benimsenmesindeki temel amaç, veri işlem sorumlulularının kişisel verilerin korunması ilkelerini uygulamaktan kaçınmak amacıyla veri işleme operasyonlarını daha düşük koruma getiren, veri cennetleri olarak adlandırılabilecek ülkelere kaydırmalarına engel olmaktır[68]. Yönergenin ülkeler arası veri aktarımlarını düzenleyen 25. ve 26. maddeleri özellikle Avrupa ve ABD arasında pek çok tartışmaya yol açmıştır. ABD, kişisel verilerin korunmasına ilişkin ilk tartışmaların gündeme geldiği günden beri federal düzeyde bir yasa çıkarmaya olumlu bakmamıştır[69]. Sistematik tek bir yasal kaynak tasarlamak yerine, ABD'nin özel hayatın gizliliğini koruyan düzenlemeleri, parça parça gelişmiştir ve kabul edilen yaygın görüşe göre, sağlık ve finans sistemlerinde olduğu gibi farklı sektörlerin farklı yasal gereksinimleri olduğundan ABD, özel hayatın gizlilğinin korunmasında sektörel bir bakış açısı benimsemiştir[70]. Bu sebeple Avrupa ve ABD arasında, 2000 yılında, Safe Harbor adıyla bilinen ve dilimize "Bağışıklı Anlaşması[71]" olarak çevirdiğimiz sözleşme imzalanmıştır. Sistemin işleyişinden müzakereleri de yürütmüş olan ABD Ticaret Bakanlığı

---

[65]Hayrunnisa Özdemir, *Elektronik Haberleşme Alanında Kişisel Verilerin Özel Hukuk Hükümlerine Göre Korunması*, Seçkin Yayıncılık, 2009, s. 30.
[66] 95/46/AT sayılı Yönerge, m. 25/1
95/46/AT sayılı Yönerge, m. 26
[68] Elif Küzeci, s. 172. Veri işlem sorumlusu orijinal cümlede yazar tarafından veri denetçisi olarak kullanılmıştır. Her iki kullanım da yönergenin orijinal metnindeki "data controller" ifadesine karşılık gelmektedir. Bu metnin gidişatına uyumlu olması adına bu cümlede de veri işlem sorumlusu ifadesi tercih edilmiştir.
[69] Elif Küzeci, s. 178
[70] Peter P. Swire, Kenesa Ahmad, s. 41
[71] Elif Küzeci, s. 180

sorumludur ve işleyiş proaktif olmayıp şikayet üzerne gerçekleştirilen bir uygulama sistemine dayanmaktadır[72]. Bu bağlamda bağışıklık sözleşmesine dahil olan ABD şirketleri bilgi verme, seçim, ileri transfer, güvenlik, veri bütünlüğü, erişim ve uygulama hususlarında taahhütlerde bulunmuş olmaktadır.

Görüldüğü üzere 95/46/AT sayılı yönerge kişisel verilerin tanımına, işlenmesine, aktarımına, içeriğine, ilgili kişilerin haklarına kadar pek çok temel usul ve esası içeren kapsamlı bir metindir. Özellikle veri aktarım hususunda belirlediği standartlar sayesinde sadece birlik üyelerini değil birlikle veri alış verişi yapmayı talep eden tüm ülkeleri etkiler niteliktedir.

b) 2002/58/AT sayılı Özel Yaşamın ve Elektronik İletişimin Korunması Yönergesi

Bu yönerge 2002 yılında "2002/58/AT sayılı Elektronik İletişim Alanında Özel Yaşamın Gizliliğinin Korunması ve Kişisel Verilerin İşlenmesine İlişkin Yönerge[73]" adıyla kabul edilmiştir. İsminden de anlaşılacağı gibi bu yönerge sektör bazlı hazırlanmıştır. İletişim teknolojilerinin hızlı gelişimi teknoloji odaklı yeni sorunları yaratmış buna istinade AB elektronik haberleşme sektörünü düzenleyecek bu yönergeyi yayınlamıştır. Komisyon elektronik haberleşme alanında bu düzenlemeyi yaparken modern haberleşme araşlarının teknik olarak çokluğunu ve gelişmişliğini göz önüne alarak geleceğe yönelik ve tarafsız bir düzenleme yapmaya çalışmıştır[74]. Yönerge genel hatlarıyla güvenlik, iletişimin gizliliği, veri işlemenin sınırlandırılması, istenmeyen iletiler, çerezlerin ve casus yazılımların kullanımı, yer bilgileri, verilerin saklanması hususlarını

[72] İkbal Gür, s.166

[73] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and protection of privacy in the electronic communications sector OJ L201/37. Bundan sonra "2002 Yönergesi" olarak anılacaktır.

[74] Oğuz Şimşek, s. 57

düzenlemektedir. Bu yönerge gerçek kişiler kadar tüzel kişileri de kapsamaktadır[75]. Yönergede, öncelikle sektöre özgü olan kullanıcı, trafik verisi, yer verisi, iletişim gibi başlıca ifadeler netleştirilmektedir. Yönerge'de trafik verileri "işletmeciler tarafından elektronik haberleşme ağında iletişimin sağlanması için faturalandırmanın yapılabilmesi için işlenen tüm verilerdir[76]" ifadesiyle geniş bir çerçevede tanımlanmıştır. Trafik verileri, haberleşmeyi sağladıkları için haberleşme özgürlüğünün korunması açısından büyük önem taşımaktadırlar[77]. Diğer taraftan bir diğer önemli tanım olan yer verisi elektronik haberleşme ağında bir kullanıcının terminal cihazının coğrafi konumunu göstermek için işlenen tüm verilerdir[78]" ifadesiyle tanımlanmıştır. Yönergenin temel katkısı bu iki kritik veri türünün hangi şartlarda işleneceğinin belirlenmesidir. Burada faturalama, tüketici sorunlarının çözülmesi, ara bağlantının sağlanması, katma değerli hizmetlerin tasarlanması, trafik yönetimi, dolandırıcılık tespiti gibi istisnalar dışında amacını aşan trafi verisi işleme süreci engellenmek istenir[79]. Yer verisi için ise ancak anonimleştirme veya kullanıcı rızası alınması halinde işleme yapılabileceği kısıtı tanımlanmıştır[80].

Bunların yanında yönerge istenmeyen iletiler ve çerezler hakkında da maddeler içermektedir. Burada istenmeyen iletiler ilgilinin "opt-in[81]" rızasıyla mümkündür[82], çerezlerin kullanımı ise ancak açık ve tam olarak bilgilendirilme yapılarak gerçekleştirilir[83].

Diğer önemli maddeler ise güvenlik ve iletişimin gizliliğinin düzenlendiği 4. ve 5. maddelerdir. Buna göre, işletmeciler güvenliği sağlayacak teknik ve organizasyonel tedbirleri almakla yükümlü tutulur ve iletişimin gizliliği adına

---

[75] 2002 Yönergesi, m. 1/3
[76] 2002 Yönergesi, m. 2/b
[77] Hayrunnisa Özdemir, s. 39
[78] 2002 Yönergesi, m. 2/c
[79] 2002 Yönergesi, m. 6
[80] 2002 Yönergesi, m. 9
[81] İzinli pazarlamaya ait bir terim olup, müşterilerin veya kullanıcıların rızaları alındıktan sonra pazarlama yapılacak liste içine dahil edilmesi anlamına gelmektedir. Önce izin alınır, sonra pazarlama yapılabilir. Ayrıca müşteri veya kullanıcıya her zaman bu liseden çıkma hakkı verilir.
[82] 2002 Yönergesi, m. 5/3
[83] 2002 Yönergesi, m. 13

dinleme, hatta girme, kaydetme gibi iletişimi gözetleyecek müdahalelerde bulunamazlar.

c) 2006/24/AT sayılı İletişim Trafik Verilerinin Saklanması Yönergesi

2006 yılında kabul edilen "2006/24/AT sayılı İletişim Trafik Verilerinin Saklanması Yönergesi[84]" genel olarak "terörizmle mücadele amacıyla trafik verilerine ulaşılmasını sağlamak için üye devletlerde veri saklama kurallarını uyumlaştırma[85]" amacı güdülmektedir. Yönergeye göre elektronik haberleme sektöründe hizmet veren işletmeciler, yönerge kapsamında belirlenmiş kategorilerdeki verileri, belirlenen süre boyunca "ciddi suçların soruşturma, tespit ve kovuşturması[86]" süreçlerinde kullanılmak üzere saklamakla yükümlüdür. Burada belirtilen veri kategorileri, "iletişimin kaynağını belirlemek, iletişimin hedefini belirlemek, iletişimin tarih,zaman ve süresini belirlemek, iletişimin türünü belirlemek ve iletişimin konumunu belirlemek[87]" için gerekli olan "arayan numara, abonenin veya kayıtlı kullanıcının adı ve adresi, aranan numara, IP adresi, kullanıcı adı, görüşme esnasında atanmış telefon numarası, konuşmanın tarih ve süre bilgileri, Internet'e erişimin tarih ve süre bilgileri, kullanılan telefon veya Internet hizmet türü, hem arayan hem de aranan tarafların IMSI ve IMEI bilgileri, coğrafi konumu belirleyecek hücre bilgisi[88]" verileridir. Ek olarak işletmecilerin bu verileri, "6 ay ile 2 yıl arasında[89]" tutması beklenmektedir.

---

[84] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the Retention of data generated or processed in connection with the provision of publicly available electronic communications service sor of public communication Networks and amending Directive 2002/58/EC, OJ L 105. Bundan sonra "2006 sayılı Yönerge" olarak anılacaktır.
[85] Elif Küzeci, s. 192
[86] 2006 sayılı Yönerge, m. 1/1
[87] 2006 sayılı Yönerge, m. 5/1
[88] 2006 sayılı Yönerge, m. 5/1
[89] 2006 sayılı Yönerge, m. 6

İletişime dair çok detaylı veirlerin saklanması öngörmesinden ötürü bu yönerge pek çok eleştiriye maruz kalmıştır. İşletmecilerin kendilerine bir suç gerekçesiyle başvurulduğunda, başvuruya cevap verebilmeleri için tüm abone ve kullanıcılarına ait iletişimin detayları saklaması gerekecektir. Bu durum da herhangi bir suça karışmamış vatandaşların iletişim bilgilerine dair yüklü miktarda verinin saklanması anlamına gelmektedir.

Bütün bu tartışmaların sonunda, Avrupa Birliği'nin yargı organı olan Avrupa Adalet Divanı 8 Nisan 2014 tarihli kararında direktifin geçersizliğini ilan etmiştir. Buna göre divan, "yönergede bahsi geçen verilerin saklanması ve bu verilere ulusal otoritelerin erişiminin sağlanmasının özel hayata saygı ve kişisel verilerin korunması hususundaki temel hak ve özgürlüklerle ciddi oranda çeliştiğini[90]" ifade ederek, direktifi geçersiz kılmıştır.

2. Türkiye

Türkiye'de bir "taslak[91]" bulunuyor olmasına rağmen hala kişisel veri korumasına istinaden usul ve esasları belirleyen bir veri koruması kanunu yürürlüğe girmemiştir. Bir çerçeve kanunun eksikliğine rağmen kişisel verilere istinaden oluşan ihlaller ve özel alan gizliliği anayasada, özel hukuk alt başlıklarında, ceza hukukunda, idare hukukunun alt başlıklarda, kolluk faaliyetlerinin mevzuatında yer almaktadır. Bu çalışmanın kapsamındaki ana düşünceden uzaklaşmamak adına Türk mevzuatındaki tüm kişisel veirlerle ilişkilili mevzuatı detaylı incelememiz mümkün değildir. Bu bağlamda, ilerleyen metinde çalışmanın geneliyle ilişkili başlıca kanun ve yönetmeliklerden bahsedilecektir.

1982 Anayasası'nın 20. maddesinin birinci fıkrasına göre "Herkes, özel hayatına ve aile hayatına saygı gösterilmesini isteme hakkına sahiptir. Özel hayatın ve aile

[90] Judgment in Joined Cases C-293/12 and C-594/12,Digital Rights Ireland and Seitlinger and Others, Court of Justice of the European Union , Press Release No 54/14, Luxembourg, 8.4.2014
[91] Adalet Bakanlığı'nca hazırlanan ve Bakanlar Kurulu'nca 7/4/2008 tarihinde kararlaştırılan *Kişisel Verilerin Korunması Kanun Tasarısı*, bkz. http://www2.tbmm.gov.tr/d23/1/1-0576.pdf

hayatının gizliliğine dokunulamaz.” Böylece, özel hayat ve aile hayatı kişilere anayasal bir hak olarak tanımlanmıştır.

Özel hukuk kapsamında kişinin özel alan gizliliğini ihlallerin önlenmesi anlamında Türk Medeni Kanununu ve Borçlar Kanunu kapsamında korumalar getirilmiştir. Buna göre MK’nın 24. Maddesine göre, “Kişilik hakkı zedelenen kimsenin rızası, daha üstün nitelikte özel veya kamusal yarar ya da kanunun verdiği yetkinin kullanılması sebeplerinden biriyle haklı kılınmadıkça, kişilik haklarına yapılan her saldırı hukuka aykırıdır.” Bu madde hangi durumların kişilik hakkı ihlali yaratmayacağını da listelemektedir. Diğer taraftan BK’nın 49. maddesi ile “hukuka aykırı bir fiille başkasına zarar veren, bu zararı gidermekle yükümlüdür” sorumluluğu tanınmıştır.

Ceza hukuku kapsamında “Türk Ceza Kanunu’nda kişisel verilerin korunması alanında değerlendirilen bazı eylemlerin suç olarak düzenlendiği görülmektedir[92]”. Bu kapsamda TCK’nın 134., 135. ve 136. maddeleri net bir şekilde suçun tanımını ve cezai yaptırımları içerir. TCK’nın 134. maddesinin 1. fıkrası “kişilerin özel hayatının gizliliğini ihlal eden kimse, bir yıldan üç yıla kadar hapis cezası ile cezalandırılır. Gizliliğin görüntü veya seslerin kayda alınması suretiyle ihlal edilmesi halinde, verilecek ceza bir kat artırılır” der. Aynı maddenin 2. fıkrasında ise ifşa ve basın yayın yoluyla yayma eylemleri dikkate alınmış ve “kişilerin özel hayatına ilişkin görüntü veya sesleri hukuka aykırı olarak ifşa eden kimse iki yıldan beş yıla kadar hapis cezası ile cezalandırılır” denmektedir. TCK’nın 135. maddesi kişisel verilerin kaydedilmesi hakkında hükümleri içerir ve “hukuka aykırı olarak kişisel verileri[93]” ve “kişilerin siyasi, felsefi veya dini görüşlerine, ırki kökenlerine; hukuka aykırı olarak ahlaki eğilimlerine, cinsel yaşamlarına, sağlık durumlarına veya sendikal bağlantılarına ilişkin bilgileri[94]” kaydeden kişilere bir yıldan üç yıla kadar hapis cezası verir. TCK’nın 136. maddesi ise verilerin yayılması ve ele geçirilmesi hususlarını düzenler ve “kişisel verileri, hukuka aykırı olarak bir başkasına veren, yayan veya

[92] Elif Küzeci, s. 286
[93] TCK, m. 135/1
[94] TCK, m. 135/2

ele geçiren kişi, iki yıldan dört yıla kadar hapis cezası ile cezalandırılır" der. Bunların yanında TCK'nın 138. maddesinde "verileri sistem içinde yok etmekle yükümlü olanlara görevlerini yerine getirmediklerinde[95]" oluşan yaptırımlar, 132., 133. ve 134. maddelerinde ise sırasıyla "haberleşme gizliliğinin ihlali", "kişiler arasında konuşmaların dinlenmesi ve kayda alınması" ve "özel hayatın gizliliğini ihlal"e ilişkin esaslar ve yaptırımlar düzenlenmiştir.

İdare hukuk kapsamında bu çalışmayla en ilişkili mevzuat 5429 sayılı Türk İstatistik Kurumu Kanunu ve 2006 yılında yayınlanmış Resmi İstatistiklerde Veri ve Gizli Veri Güvenliğine İlişkin Usul ve Esaslar Hakkında Yönetmelik'tir. Türk İstatistik Kanunu'nun 1. maddesinde kanunun amacı "resmî istatistiklerin üretimine ve organizasyonuna ilişkin temel ilkeleri ve standartları belirlemek; ülkenin ihtiyaç duyduğu alanlarda veri ve bilgilerin derlenmesini, değerlendirilmesini, gerekli istatistiklerin üretilmesini, yayımlanmasını, dağıtımını ve Resmî İstatistik Programında istatistik sürecine dâhil kurum ve kuruluşlar arasında koordinasyonu sağlamak üzere, Türkiye İstatistik Kurumunun kuruluş, görev ve yetkilerine ilişkin esasları düzenlemektir" şeklinde tanımlanmıştır. Bunların yanında kişisel veri yerine veri, bireysel veri ve gizli kavramları geçmektedir. Kanunun 2. maddesinde yer alan tanımlara göre, veri "anket veya idarî kayıtlar yoluyla elde edilen nicel ve/veya nitel istatistikî bilgiler[96], bireysel veri "hakkında bilgi toplanan istatistikî birimlerin, özellikleri ile birlikte tanımlandığı veriyi[97], gizli veri ise "istatistik birimin doğrudan veya dolaylı bir şekilde özellikleri ile birlikte tanınabilmesine ve bu şekilde bireysel bilgilerin açığa çıkarılmasına imkân sağlayan bireysel veya tablo hâlinde saklı tutulan veriyi[98]" göstermektedir. Kanunun ikinci bölümünde bilgilerin toplanması, dağıtımı ve gizliliği hususularının düzenlenmektedir ve bu bölüm bilgi isteme, bilgilerin doğruluğunu araştırma, kontrol ve saklama, cevap verme yükümlülüğü ve sınırları, ulusal kayıt sistemleri, sınıflamalar, istatistik sonuçlarına erişim, gizli veriler, bireysel verilerin kullanımı, istatistikî birimlerin hakları başlıklarından

[95] TCK, m. 138/1
[96] Türkiye İstatistik Kanunu, m. 2/n
[97] Türkiye İstatistik Kanunu, m. 2/o
[98] Türkiye İstatistik Kanunu, m. 2/s

oluşmaktadır. Buna göre 13. ve 14. maddelerinde düzenlenen "istatistik üretiminde görev alanlar resmî istatistik üretiminde görev alanlar, görevlerini yerine getirebilmek için ihtiyaç duydukları ölçüde erişebilirler[99]" ve aynı şekilde gizli veri "istatistik amacı dışında kullanılamaz[100]" ve "idarî, adlî ve askerî hiçbir organ, makam, merci veya kişiye verilemez[101]" ifadeleriyle verilerin kullanım amacı ve paylaşım kısıtları netleştirilmiştir. Kanunun üçüncü bölümü Türk İstatistik Kurumuna dair görev ve teşkilat yapısı esaslarını içerirken, dördüncü bölüm "programın hazırlanmasına, uygulanmasına, resmî istatistiklerin gelişimine ve işlevlerine ilişkin tavsiyelerde bulunmak, resmî istatistik ihtiyaçlarını tespit etmek, değerlendirmek ve ileriye yönelik görüş ve önerileri kapsayan çalışmalar yapmak" görevlerinin yerine getirilmesi amacıyla İstatistik Konseyinin kurulmasını öngürür. Kanuna ek olarak 2006 yılında yayınlanmış Resmi İstatistiklerde Veri ve Gizli Veri Güvenliğine İlişkin Usul ve Esaslar Hakkında Yönetmelik, kanunla aynı tanımlara sahip olmakla beraber gizli veriye dair gizlilik ve güvenlik hususlarında detayları, kurum ve kuruluşların yükümlülüklerini belirlemektedir. Yönetmeliğin 5. ve 6. maddelerinde gizli verinin istisnaları ve tablolaştırılmış verinin gizliliğine dair esaslar detaylandırılır. Küzeci (2010), yönetmleikte 5. Maddede yer alan "iş kayıtları sistemi kapsamında bulunan istatistikî birimlerin bu nitelikteki unvan, faaliyet ve adres bilgileri"nin de istisna olarak tanımlandığına dikkat çeker. Böylelikle kanunda belirtilen "herkese açık kaynaklardan edinilen veriler" istisnası genişletilmiş olur ve "temel hak ve özgürlüklerin ancak yasa ile sınırlandırılabileceği hükmüne aykırılık oluşturduğu gibi, normlar hiyerarşisine ve yönetmeliklerin amacına uygun değildir[102]". Yönetmeliğin üçüncü bölümü kurum ve kuruluşların nezdinde gizli verinin korunması, görevlilerin tespiti, gizli/bireysel verinin paylaşım sınırları, gizli verinin bilimsel amaçlı kullandırılması, gizliliği sağlama taahhüdü, bilgi taleplerinin karşılanması, istatistiki birimlerin hakları hususlarında sorumlulukları

[99] Türkiye İstatistik Kanunu, m. 13, parag. 1
[100] Türkiye İstatistik Kanunu, m. 13, parag. 3
[101] Türkiye İstatistik Kanunu, m. 13, parag. 3
[102] Elif Küzeci, s. 329

belirlemektedir. Ayrıca yönetmelik "gelişmeleri takip etmek[103]", "kurumsal stratejileri belirlemek[104]" ve "görüş oluşturmak[105]" görevlerini yerine getirmesi amacıyla Veri Gizliliği İhtisas Komisyonu'nun oluşturulmasını öngörmektedir.

Görüldüğü üzere, pek çok mevzuatta kişisel verilerin ihlaline ve korunmasına yönelik düzenleme olmasına rağmen, bu durum kişisel verilerin korunması kanunu gibi çerçeve bir kanunun gerekliliğini ortadan kaldırmamaktadır. Mevcut mevzuatlarda ya ihlal gerçekleştikten sonra yaptırımları belirleyen ardıl önlemler sıralanır (MK, TCK vs.) ya da sadece belli bir amaca hizmet eden istatistiksel veriler (İstatistik Kanunu) dikkate alınmaktadır. Kişisel verinin tanımını, esaslarını, işlenmesine dair hususları belirleyecek çerçeve bir kanun bütün mevzuatı tamamlayıcı nitelikte olacaktır.

Türk mevzuatında çerçeve kanun kapsamında olmasa bile ve sektörel bir bakış açısı getirmesine rağmen kişisel verilere dair hususları işleyen ve yürülükte olan tek metin "Elektronik Haberleşme Sektöründe Kişisel Verilerin İşlenmesi ve Gizliliğin Korunması Hakkında Yönetmelik[106]"tir. Gelişen teknoloji ve iletişim araçlarının yrattığı yeni tehditler, Kişisel Verilerin Korunması Kanun Taslağı'nın yürürlüğe girmesini beklemeden sektörel bir yaklaşım geliştirmeyi gerektirmiş ve Bilgi Teknolojileri ve İletişim Kurumu tarafından ilk olarak 2004 yılında yayınlanan veri koruması yönetmeliği, 2012 yılında revize edilmiştir. 5809 sayılı "Elektronik Haberleşme Kanunu[107]" uyarınca "[Bilgi Teknolojileri ve İletişim Kurumu] elektronik haberleşme sektörüyle ilgili kişisel verilerin işlenmesi ve gizliliğinin korunmasına yönelik usul ve esasları belirlemeye yetkilidir[108]" ve bubyetkisini kullanarak veri koruması yönetmeliğini hazırlamıştır. Veri koruması yönetmeliği veri, trafik verisi, konum verisi, kişisel veri tanımlarını netleştirmekte ve rıza ve kişisel verilerin işlenmesi hususlarındaki tanımları içermektedir. Buna

[103] Türkiye İstatistik Kanunu, m. 15/a
[104] Türkiye İstatistik Kanunu, m. 15/b
[105] Türkiye İstatistik Kanunu, m. 15/c
[106] Resmi Gazete, t. 24.07.2012, s. 28363. Bundan sonra "Veri Koruması Yönetmeliği" olarak anılacaktır.
[107] Resmi Gazete, t. 10.11.2008, s. 27050
[108] Elektronik Haberleşme Kanunu, m. 51

göre, veri "abone ya da kullanıcıyı teşhis etmek için yararlanılan trafik verisi, konum verisi ya da ilgili diğer bilgileri[109]", kişisel veri "belirli veya kimliği belirlenebilir gerçek ve tüzel kişilere ilişkin bütün bilgileri[110]", trafik verisi "bir elektronik haberleşme şebekesinde haberleşmenin iletimi veya faturalama amacıyla işlenen her türlü veri[111]", yer verisi "kamuya açık elektronik haberleşme hizmeti kullanıcısına ait bir cihazın coğrafi konumunu belirleyen ve elektronik haberleşme şebekesinde veya elektronik haberleşme hizmeti aracılığıyla işlenen belirli veri[112]" olarak tanımlanmaktadır. Genel olarak veri koruması yönetmeliği, uygulama esasları, verilerin işlenmesi ve saklanması, sağlanan imkânlar ve yaptırımlar bölümlerinden oluşmaktadır. Uygulama esasları bölümünde kişisel verilerin işlenmesine dair ilkeler, güvenlik ve ihlallerin bildirilmesine dair detaylar bulunur. Buna göre kişisel veriler "hukuka ve dürüstlük kurallarına uygun[113]", "ilgili kişinin rızasına dayalı[114]", "amacıyla bağlantılı, yeterli ve orantılı[115]", "doğru[116]" ve güncellenebilir, "ilgili kişilerin kimliklerini belirtecek biçimde ve kaydedildikleri veya yeniden işlenecekleri amaç için gerekli olan süre kadar muhafaza edilerek[117]" işlenmelidir. Verilerin işlenmesi ve saklanması bölümünde ise haberleşmenin gizliliği, trafik ve konum verilerini nasıl işleneceği, saklanacak veri kategorileri ve saklama süreleri, saklanan verilerin güvenliğinin sağlanması hususlarında detaylar belirlenmektedir. Sağlanan imkanlar bölümü ise, numaranın gizlenmesi, otomatik çağrı yönlendirme, rehber hizmetleri, ayrıntılı faturada gizlilik gibi abone ve kullanıcılara sunulması gereken hizmetleri belirler. Veri koruması yönetmeliği, sadece elektronik haberleşme sektörünü kapsamasına rağmen sektöre temel oluşturacak ve kişisel verierin işlenmesine, saklanmasına, güvenliğine dair esasları belirleyecek niteliktedir.

---

[109] Veri Koruması Yönetmeliği, m. 3/r
[110] Veri Koruması Yönetmeliği, m. 3/h
[111] Veri Koruması Yönetmeliği, m. 3/p
[112] Veri Koruması Yönetmeliği, m. 3/j
[113] Veri Koruması Yönetmeliği, m. 4/a
[114] Veri Koruması Yönetmeliği, m. 4/b
[115] Veri Koruması Yönetmeliği, m. 4/c
[116] Veri Koruması Yönetmeliği, m. 4/d
[117]Veri Koruması Yönetmeliği, m. 4/e

3. Veri İşleme ve Kişisel Verilerin Korunması Bakımından Önemli bir Metot: Anonimleştirme

Büyük veriden fayda sağlanması adımında verinin işlenmesi her ne kadar vazgeçilmez bir durum ise kişisel verilerin ve özel alan gizliliğinin korunmasının gerekliliği de toplum düzeni, tüketici güveni, temel hak ve özgürlükler nezdinde bir o kadar hayatidir.

İşte bu noktada, anonimleştirme bu fayda ve gizlilik dengesini korumak adına üretilen teknik çözümlerden biri olarak karşımıza çıkar.

Bu bölüm bahsi geçen dengeyi sağlamak adına kurgulanan mimari bazlı gizlilik teknolojilerine değinerek, anonimleştirme süreçlerini, yasal mevzuatlardaki dayanaklarını, güvenlik açıkları ve risklerini inceleyecektir.

I. Veri Anonimleştirmesinin Dayanakları ve Amacı

İçinde bulunduğumuz çağın gelişmiş bilgi ve iletişim teknolojilerini hesaba kattığımızda, kişisel verilerin korunması ve özel hayatın gizliliği politikalarının hayata geçirilmesi iki temel düzlemde gerçekleşmeltedir: Politikalara Dayanan Gizlilik ve Mimariye Dayanan Gizlilik. Bilgi teknolojilerinin hızla geliştiği ve birbiriyle yakınsadığı yıllardan itibaren kişisel verilerin ve gizliliğin yalnızca yukarıda değindiğimiz mevzuat ve politikalara dayanan anlayışlarla korunamayacağı anlaşılmış ve gizliliği koruyacak tasarımlar teknolojik altyapıya dahil edilmeye başlanmıştır. 90'lı yıllara kadar gizlilik ve veri koruması sorunlarına karşı halen geçerli olan anlayış "güçlü bir regülatif bakış açısı" edinmek iken bu yıllardan sonra "gizlilik süreçlerini çeşitli teknolojilerin teknik tasarımlarına gömülmesi" olarak bildiğimiz mimariye dayalı gizlilik anlayışı

benimsenmiştir[118]. Böylece özel alan gizliliği ve kişisel veri korunması sorununa sadece politikalara tabi olarak değil aynı zamanda teknik altyapılarla da çözümler üretilmesi beklenmektedir. Lessig, her iki anlayışın farklarını betimlemek adına, bugün Internet'te hemen hemen her web sayfasında karşılaşılan ve sayfayı görüntüleyebilmek veya uygulamayı kullanabilmek için kabul edilmesi şart koşulan "gizlilik politikaları" veya "şartlar ve koşullar" metinlerini örnek olarak vermektedir. Bu metinler, teorik olarak önceki bölümlerde detaylandırdığımız yasal mevzuatlara genel çerçeveleriyle uyumlu olarak hazırlanmış ve kullanıcının rızasını almaya yönelik tasarlanmışlardır. Ancak pratikte metinlerin uzunluğu ve detaylı içeriği bireylerin bu metinleri inceleyip, değerlendirip kabul etmesini engellemekte ve uygulamayı kullanabilmek için bilinçsizce rıza vermelerini sağlamaktadır. Hâlbuki, "bireylerin kendileriyle ilgili veri üzerinde etkili kontrol hakkı veren[119]" gizlilik teknolojileri geliştirilerek, kullanıcıların kendi tercihlerini kontrol etmeleri ve bu tercihlerle uyumlu olmayan sitelerde uyarılar görüntülemeleri sağlanabilir. Böylece, gizlilik gereksinimleri tasarıma dahil edilmiş olur ve gizlilik, mimari altyapı ile sağlanır. Diğer bir deyişle, "hukuk, belli bir çeşit teknolojiyi teşvik ederek, teknolojinin bireylere siber alemde talep ettiklerine daha rahat ulaşmasını sağlar: Hukuk koda yardım ederek gizlilik politikalarını mükemmelleştirir[120]".

Avrupa Veri Koruma Denetçisi, 2010 yılında verdiği görüşte, mimariye dayalı gizlilik korumalarını "teknolojinin ve tasarımın en başından itibaren her adımında veri koruması ve özel alan gizliliğini hesaba katarak tasarlanmış ve geliştirilmiş bilgi ve iletişim teknolojileri" olarak tanımlarken bu anlayışın mevcut veri koruması yönergesine "bağlayıcı prensip" olarak dahil edilmesi gerektiğini savunmuştur[121]. Benzer şekilde, Enformasyon Komiserliği Ofisi'nin 2008 tarihli

---

[118] Ann Couvakian, *Privacy By Design…Take the Challange*, Canada, 2009, s.4
[119] Lawrance Lessig, *Code Version 2.0*, Basic Books, 1996, s. 226
[120] Lawrance Lessig, s. 230
[121] Avrupa Veri Koruma Denetçisi, *EDPS opinion on privacy in the digital age: "Privacy by Design" as a key tool to ensure citizens' trust in ICTs*, Brussels, 2010,
bkz.
https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/PressNews/Press/2010/EDPS-2010-06_Privacy%20in%20digital%20age_EN.pdf

dökümanında, mimariye dayalı gizlilik şu şekilde tanımlanmaktadır: "Gizlilik mimarilerinin amacı gizlilik ihtiyaçlarına tasarımın ilk adımından itibaren öncelik vermek, bir başka deyişle bir sistemin veya sürecin bireylerin gizlilikleri üzerindeki etkisini dikkate almak ve bunu sistemin tüm yaşam döngüsü içinde tüm gerekli kontrollerin uygulandığından ve sürdürüldüğünden emin olarak gerçekleştirmek[122]". Aynı dökümanda gizlilik mimarilerinin risklerin minimize edilmesinde ve tüketicilerin güveninin yeniden kazanılmasındaki rolü de vurgulanmaktadır.

Mimariye dayalı gizlilik korunması süreçleri teknoloji, iş modelleri ve fiziksel tasarım olacak şekilde üç boyutta kendisine uygulama alanı bulmaktadır[123]. Bu anlamda önemli olan adım, tüm tasarım ve modelleme süreçlerinde olduğu gibi, ihtiyacın kesin olarak netleştirilmesi adımıdır. "Talep analizleri tipik olarak mevcut süreçlerin detaylı olarak incelenmesi ve ilgili tarafların ihtiyaçlarının tespit edilmesiyle başlar" ve aynı şekilde "gizlilik mühendisliği süreçlerinde de kullanıcıların gizlilik anlayışlarının ve beklentilerini" ve "gerekli olan gizlilik seviyesini" belirlemek gerekmektedir[124].

Yukarıda değindiğimiz üzere, kullanıcı rızasını sağlamak adına yayınlanan uzun metinler yerine "kullanıcının yükünü azaltacak" nitelikte olan ve "web sayfalarının veri toplama ve işleme metotlarını makine-okunabilir formatta" bir tasarıma sahip olan, kullanıcıya daha çok kontrol hakkı veren gizlilik teknolojileri geliştirme süreci, mimariye dayalı gizlilik telknolojilerine bir örnektir[125]. Hatta bir sonraki adımda "bağımsız denetleyicilerden alınan" ve kurumların "bilgi teknolojileri ürün ve servislerinin arkasındaki yazılım ve geliştirme süreçlerinin incelenerek, gizlilik fonksiyonlarının mevcut olduğunu onaylayan" sertifikalar da mimariye dayalı gizlilik politikalarının artan önemine bir örnek teşkil

---

[122] Enformasyon Komiserliği Ofisi, *Privacy by Design*, 2008, s.3, bkz. http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/documents/pdb_report_html/PRIVACY_BY_DESIGN_REPORT_V2.ashx

[123] Ann Couvakian, s. 3

[124] Sarah Spiekerman, Lorrie Faith Cranor, *Engineering Privacy*, IEEE Transactions on Software Engineering, Vol. 35, Nr. 1, 2009, s. 69

[125] Ian Brown, Christopher T. Marsden, s. 53

etmektedir[126]. Bu bağlamda gizlilik teknolojilerinin pek çok metotla karşımıza çıkması mümkündür. Bu tip teknolojilerin temel amacı, bilgi kullanımına dayalı çalışan servislerin yetkisiz erişimden ya toplanan verinin minimize edilmesiyle, ya toplanan verinin anonimleştirilmesiyle ya da toplanan verinin güvenliğinin sağlanmasıyla korunmasını sağlamaktır[127].

Cranor ve Spiekerman, mimariye dayalı gizlilik tasarımlarının verinin kimlik saptayabilme özelliği dikkate alınarak kurgulandığını savunur. Kişinin kimliğinin saptanabilirliği "hangi verinin direk olarak hangi kişiyle ilişkilendirilebileceğinin derecesi" olarak tanımlarken "anonim verilerin veri sağlayıcısı için en yüksek gizlilik derecesini sağlayacağını" savunmaktadırlar[128]. Bu anlamda verilerin anonimleştirilmesi kişisel veri korumasında ve gizliliğin sağlanmasında kurgulanmış teknik çözümlerden biri olarak karşımıza çıkmaktadır.

Enformasyon Kimserliği Ofisi'nin "Anonimleştirme: Veri Koruması Risklerinin Yönetimi[129]" adıyla yayınladığı kılavuzu, kişisel veri tanımından yola çıkarak şu sonuca varmaktadır: "kişinin kimliğini saptamayan ve kişiyle ilişkilendirilemeyen bilgiler kişisel veri değildir[130]". İlişkilendirme ve kimliğin saptanabilmesi 29. Madde Veri Koruma Grubu'nun kişisel veri kavramına istinaden yayınladığı görüş dökümanında da detaylıca incelenmektedir. Genel olarak, ilişkili veri "kişi hakkındaki veriler" olarak değerlendirilirken, kimlik saptaması "bir kişinin bir grup içinden ayırt edilebilmesi" olarak tanımlanmıştır[131]. Anonimleştirme süreçlerinde belirleyici olan kimlik saptayabilme özelliğidir. Çünkü bu özellik, "belirli bir bireyle yakın ve öncelikli bir bağ kuran ve betimleyici olarak adlandırdığımız bilgi parçalarıyla sağlanır[132]". Betimleyiciler direk ve dolaylı

---

[126] Regulating Code, s. 54
[127] Seda Gürses, George Danezis, A Critical Review of Ten Years of Privacy Technology, UK, 2012, s. 2
[128] Sarah Spiekerman, Lorrie Faith Cranor, s. 74
[129] Enformasyon Kimserliği Ofisi, Anonymization: Managing Data Protection Risk Code of Practice, 2012.
bkz. http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation
[130] Enformasyon Kimserliği Ofisi, s. 11
[131] 29. Madde Veri Koruma Grubu , Opinion 4/2007 on the concept of personal data , 2007, s. 12
bkz. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf
[132] 29. Madde Veri Koruma Grubu, s. 12

olarak kişiyi saptayabildiklerinden ikiye ayrılırlar. Kişinin ismi gibi bilgiler direk betimleyici olarak adlandırılırken, telefon numrası, adresi gibi başka bilgilerle birleşmelerinden oluşan kombinasyonlar ile kişiyi saptayan bilgiler dolaylı betimleyicilerdir. Bazı dolaylı betimleyiciler kişinin hiç çaba göstermeden kimliğini saptayabildiği gibi (İspanya'nın şu anki Cumhurbaşkanı), bazı durumlarda, kategoriler seviyesindeki detayların kombinasyonları da (yaş, etnisite…v.b) gayet kesin sonuçlar vermektedir[133].

Anonimleştirme bir veri kümesindeki betimleyici değişkenlerin çıkartılmasını veya değiştirilmesini içermektedir[134]. Bu durumda kimliğin saptanabilmesi özelliğini kaybederek belli bir kişiye işaret etmeyen verilere anonimleştirilmiş veri demekteyiz. Bu noktada dikkat çekilmesi gereken husus, anonim veri ve anonimleştirilmiş veri arasındaki farktır. Anonim veri belirli bir kişiyle ilişkilendirilmesi mümkün olmayan veriyi ifade ederken, anonimleştirilmiş veri daha öncesinde bir kişiyle ilişkilendirilmiş ancak artık bağlantısı kalmamış veridir[135]. Bu durum ileride tartışacağımız anonimleştirmenin riskleri ve kimliğin geri saptanabilmesi süreçleri açısından önem arz etmektedir.

Çalışmanın kapsamıyla uyumlu olması sebebiyle veri anonimleştirilmesi metotlarını detaylarını inceleyeceğimiz ilerleyen bölümlere dair bazı kavramları netleştirmek gerekmektedir. Bu bağlamda veri kümesi; dikkate aldığımız verilerin tablolara kaydedilmiş ve veri yönetim sistemlerine aktarılmış belli bölümlerini ifade etmektedir. Değişken; bu tablolarda yer alan ve her kayda ait betimleyicileri gösterir. Her bir kaydın her bir değişken için sahip olduğu bilgi ise değişkenlerin aldığı değerler olarak kabul edilmektedir. Örneğin; bir okulun öğrencilerine ait ad, soyad, yaş, okul ortalaması gibi değişkenlerden oluşan öğrenci listesi bir veri

---

[133] 29. Madde Veri Koruma Grubu, s. 13

[134] International Household Survey Network, *Anonymization Principles*, bkz. http://www.ihsn.org/home/node/137

[135] Douwe Korff, *Comperative Study on Different Approaches to New Privacy Challanges, In Particular in the Light of Technological Developments*, Working Paper No. 2: Data Protection Laws in the EU: The Difficulties in meeting the challanges posed by global social and technical developments, London Metropolitan University, 2010, s. 48 bkz. http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf

kümesini oluşturur. Her bir öğrenciye ait bu değişkenlerin karşılıkları olan bilgiler de öğrencilere ait değerlerdir. Ek olarak, anonimleştirme, çalışma alanları farklılığı sebebiyle iki alt kola ayrılmaktadır; veri anonimleştirmesi ve iletişim anonimleştirmesi. "Veri anonimleştirmesi; bir kişi ile o kişi hakkındaki veri arasındaki bağlantıyı koparmak anlamına gelirken, iletişim anonimleştirmesi, kişi ile kişini çevrimiçi aktiviteleri arasındaki bağlantıyı koparmak[136]" anlamına gelmektedir. Bu çalışma kapsamında incelenecek olan veri anonimleştirmesi süreçleridir ve metnin ilerleyen bölümlerinde kullanılan *anonimleştirme* kelimeleri veri anonimleştirmesi süreçlerini ifade etmektedir.

Veri anonimleştirilmesi, pratikte pek çok farklı metot ile sağlanmaktadır. Bu metotlar her yaşanan ihlalde yeni risklerin fark edilmesiyle yeniden tasarlanan ve geliştirilen dinamik yapılara sahiptirler. Bu noktada özellikle dolaylı betimleyicilerin kombinasyonlarıyla sağlanan kimlik saptama işlemleri istatistikçilerin anonimleştirme metolarını geliştirmesinde belirleyici olmakta ve anonimliğin ölçülmesi gerektiği gerçeğini doğurmaktadır. Diyebiliriz ki, ifşa edilen bilginin yarattığı riske göre anonimleştirme metotoları gelişmekte ve başarıları ölçülmektedir.

İki temel ifşa riski bulunmaktadır: kimlik ifşası ve özellik ifşası[137]. Kimlik ifşası, bir kişi ifşa edilen veri içinde belirli bir kayıtla ilişkilendirilebiliyorsa; özellik ifşası ise bir kişi hakkında yeni bir bilgi öğrenilebiliyorsa, örneğin ifşa edilmiş veri, kişiye ait bir karakteristiğin veri ifşa edilmeden önceki duruma göre daha kesin olarak ortaya çıkmasına sebep oluyorsa gerçekleşir[138]. Genel olarak, kimlik ifşasında kişi veri kümesi içindeki bir kayıtla direk ilişkilendirilebilir ve bu kişiye ait pek çok veri ortaya çıkmış olur. Ancak özellik ifşasında kişinin tam olarak kimliğinin ortaya çıkmasına gerek yoktur. Veriyi inceleyen kişinin önceden

[136] Matthijs R. Koot, Measuring and Predicting Anonymity, Gildeprint Drukkerijen, 2012, s. 12
[137] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, *Handbook on Statistical Disclosure Control Version 1.2*, ESSNet, 2010, s.11
[138] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, *t-Closeness: Privacy beyond k-Anonymity and l-Diversity*, Data Enginering (ICDE) IEEE 23rd International Conference, 2007, s. 106

öğrendiği bilgilerine de bağlı olarak yaptığı tahminlere ve algısına dayalı olarak da gerçekleşebilir.

Veri anonimleştirilmesinde ilk adım direk betimleyicilerden kurtulmaktır[139]. Daha sonrasında ise elde edilen kombinasyonlara bakılarak anonimlik ölçümleri yapılır ve ifşa riskini minimize edecek önlemler alınması gerekmektedir.

İlerleyen bölümlerde ilk önce direk belirleyiciler atılarak verinin kimlik saptayabilme özelliğinin ortadan kaldırıldığı yöntemler incelenecektir. Bu metotlar değişkenlere ait değerlerde düzensizlik yaratan ve yaratmayan olmak üzere iki ayrı başlık altında incelenecektir. Sonraki bölüm ise elde edilen anonim veri kümelerinin dolaylı betimleyicilerinin oluşturduğu kombinasyonları sayesinde oluşan ifşa riskini minimize edecek istatistik metotlarını inceleyecektir. Bu istatistik metotları kimlik saptama ihtimalini ve dış bilgilere sahip özellikle art niyetli kullanıcıların isabetli tahminlerde bulunma olasılığını düşürmeyi hedefler.

II. Teknik Altyapı

Bu başlık altında veri anonimleştirmesini sağlamak adına tasarlanmış teknik metotlar ve istatistik hesaplamaları incelenecektir.

A. Değer Düzensizliği Sağlamayan Anonimleştirme Metotları

Bu metotlar veri kümesindeki değerleri değiştirmez, onun yerine orijinal veri kümesindeki detaylarda kısmi gizlemeler ve eksiltmeler sağlarlar[140]. Bu durumda

[139] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, s. 106

değerler değişmediği halde veri kümesinin genelinde değişiklik yaşandığı için bilgi kaybı söz konusudur.

1. Değişkenleri Çıkartmak

Veri içerisinde yer alan değişkenlerden birinden veya bir kaçının tablodan bütünüyle silinerek çıkartılmasıyla sağlanan bir anonimleştirme tekniğidir. Böyle bir durumda tablodaki bütün sütun tamamiyle kaldırılacaktır. Bu çözüm değişkenin "yüksek dereceli bir betimleyici olması", "daha uygun bir çözümün var olmaması","değişkenin kamuya ifşa edilemeyecek kadar hassas bir veri olması" veya "analitik amaçlara hizmet etmiyor olması" gibi sebeplerle kullanılabilir[141]. Aşağıdaki örnekte, Tablo-1 orijinal veri kümesini gösterirken, Tablo-2 değişkenlerden birinin tablodan çıkartılmasıyla anonimleştirilmiş veri kümesini göstermektedir. Bu örnekte çıkartılan sütunda hassas bir veri kategorisi olan etnik köken bilgisi bulunmaktadır.

Yaş	Gender	Posta Kodu	Gelir	Aylık Harcamalar	Etnik Köken
22	K	SO17	20,000	1,100	İngiliz
25	E	SO18	22,000	1,300	İrlandalı
30	E	SO16	32,000	1,800	Afrikalı
35	K	SO17	31,000	2,000	Çinli
40	K	SO15	68,000	3,500	Pakistanlı
50	E	SO14	28,000	1,200	İngiliz

Tablo-1: Gelir, Aylık Harcamalar ve Etnik Köken Bilgisi Dağılımı

[140]Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, s. 32

[141]International Household Survey Network, *Reducing the Disclosure Risk,* bkz. http://www.ihsn.org/home/node/201

Yaş	Gender	Posta Kodu	Gelir	Aylık Harcamalar
22	K	SO17	20,000	1,100
25	E	SO18	22,000	1,300
30	E	SO16	32,000	1,800
35	K	SO17	31,000	2,000
40	K	SO15	68,000	3,500
50	E	SO14	28,000	1,200

Tablo-2 : Etnik Köken Alanı Çıkartılarak Anonimleştirilmiş Veri Kümesi

2. Kayıtları Çıkartmak

Bu yöntemle yukarıdaki yöntemden farklı olarak tüm değişken yerine sadece belli kayıtlar veriden çıkartılmaktadır. İlgili kayıt diğer hiçbir kayıtla ortak değere sahip değilse ve tüm değişkenleri tekillik ihtiva ediyorsa bu durum, tüm veri kümesi için de bu verin öznesinin kimliğinin saptanabilirliğini kolaylaştırmaktadır. Örneğin, anket sonuçlarının yer aldığı bir veri kümesinde, herhangi bir sektörden yalnızca tek bir kurum ankete dahil edilmiş olsun. Böyle bir durumda tüm anket sonuçlarından "sektör" değişkenini çıkartmaktansa sadece ilgili kuruma ait kaydı çıkartmak tercih edilebilir[142].

Aşağıdaki örneklerde Tablo-3 orijinal veri kümesini gösterirken, Tablo-4 kimliği kolaylıkla saptanabilecek kaydın çıkartılmasıyla oluşan anonimleştirilmiş veri kümesini göstermektedir.

[142] International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201

Yaş	Cinsiyet	Doğum Yeri	Üniversite	Derece (GPA)
21	K	İstanbul	İstanbul Bilgi Üniversitesi	3.02
21	E	İstanbul	İstanbul Üniversitesi	3.24
21	E	Ankara	Galatasaray Üniversitesi	2.22
22	K	Ankara	Galatasaray Üniversitesi	2.26
23	E	Muğla	Hacettepe Üniversitesi	2,98
21	K	İstanbul	İstanbul Bilgi Üniversitesi	2.77
22	K	Ankara	Galatasaray Üniversitesi	3.78

Tablo-3: Üniversite ve Derece Dağılımı

Yaş	Cinsiyet	Doğum Yeri	Üniversite	Derece (GPA)
21	K	İstanbul	İstanbul Bilgi Üniversitesi	3.02
21	E	İstanbul	İstanbul Üniversitesi	3.24
21	E	Ankara	Galatasaray Üniversitesi	2.22
22	K	Ankara	Galatasaray Üniversitesi	2.26
21	K	İstanbul	İstanbul Bilgi Üniversitesi	2.77
22	K	Ankara	Galatasaray Üniversitesi	3.78

Tablo-4: Tekillik yaratan kayıt çıkartıldıktan sonra oluşan dağılım

3. Alt ve Üst Sınır Kodlaması

Alt ve üst sınır kodlama yöntemi önceden tanımlanmış kategorilerin yer aldığı değişkenlere ait değerlerin birleştirilmesiyle elde edilen anonimleştirme yöntemidir (örneğin, yaş değişkeinin 5 yıllık yaş gruplarına göre kodlama, veya çalışan sayılarını düşük, orta, yüksek olacak şekilde üç kategoriye göre

kodlama)[143]. Üst sınır kodlaması uygulanırken, sıralı değerlere sahip değişkene ait en yüksek değerler bir araya toplanır, aynı şekilde alt sınır kodlamasında da el düşük değerler bir ara toplanarak yeni kategoriler elde edilir[144]. Elde edilen yeni kategorilere göre tablo yeniden düzenlenir.

Aşağıdaki örneklerde Tablo-5 orijinal veri kümesini, Tablo-6 seçilen değişkenlerin alt ve üst sınır kodlaması yapılarak yeniden tasarlanarak anonimleştirilmiş şeklini göstermektedir.

Yaş	Cinsiyet	Meslek	Gelir	Medeni Durum	Harcamalar (Aylık)
34	K	Avukat	74.000	Bekar	3.000
55	E	Mühendis	54.000	Evli	3.600
45	E	Doktor	63.000	Evli	5.000
61	K	Doktor	36.000	Bekar	1.800
27	E	Doktor	42.000	Evli	2.100
33	E	Avukat	31.000	Bekar	4.300

Tablo-5: Gelir ve Harcamalar Dağılımı

Tablodaki *Gelir* ve *Harcamalar(Aylık)* değişkenlerine ait değerleri global kodlama yöntemi ile aşağıdaki şekilde değiştirelim;

Gelir: Düşük= 40.000'den küçük ve eşit değerler; Orta= 40.000 ve 55.000 arası; Yüksek= 55.000'den büyük ve eşit değerler

Harcamalar(Aylık): Düşük= 2.000'den küçük ve eşit değerler; Orta= 2.000 ve 3.500 arası; Yüksek= 3.500'den yüksek ve eşit değerler

[143]International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201

[144] Anco Hundepool, Aad van de Wetering, Ramya Ramaswamy, Luisa Franconi, Silvia Polettini, Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo, Vicenc Torra, Ruth Brand, Sarah Giessing, µ- *ARGUS version 4.2 User's Manuel*, 2008, ESSNet-Project, s.13

Bu kodlamaya göre tablo aşağıdaki şekli alacaktır.

Yaş	Cinsiyet	Meslek	Gelir	Medeni Durum	Harcamalar (Aylık)
34	K	Avukat	Yüksek	Bekar	Orta
55	E	Mühendis	Orta	Evli	Yüksek
45	E	Doktor	Yüksek	Evli	Yüksek
61	K	Doktor	Düşük	Bekar	Düşük
27	E	Doktor	Orta	Evli	Orta
33	E	Avukat	Düşük	Bekar	Yüksek

Tablo-6: Gelir ve Harcamalar değişkenleri anonimleştirilmiş veri kümesi

4. Global Kodlama

Alt ve üst sınır kodlama yöntemi "sayısal ve sıralı[145]" kategorilere ayrılabilen değişkenlere uygulanmaktadır. Eğer değişken bu özelliklere sahip değilse; "çeşitli kategorilerin birleştirilerek tek bir kategori haline dönüştürülmesi[146]" yöntemine global kodlama adı verilir.

Global kodlama yöntemi alt ve üst sınır kodlaması uygulanan sıralı ve sayısal değerlere sahip olmayan değişkenlere uygulanan bir kodlama yöntemidir. Burada çeşitli kategoriler sadece tek bir kategori teşkil edecek şekilde değiştirilir.

Dikkat edilmesi gereken husus global kodlama yönteminin sadece güvenli olmayan küme değil bütün veri kümesine uygulanıyor olmasıdır.

---

[145] International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201

[146] Anco Hundepool, Aad van de Wetering, Ramya Ramaswamy, Luisa Franconi, Silvia Polettini, Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo, Vicenc Torra, Ruth Brand, Sarah Giessing, s. 12

Aşağıdaki örnekte Tablo-7 orijinal veri kümesini göstermektedir. Tablo-8 ise global kodlama uygulamasından sonraki anonimleştirilmiş verinin durumudur.

Cinsiyet	Meslek	İl	İlçe
K	Avukat	İstanbul	Beylikdüzü
K	Doktor	İstanbul	Beylikdüzü
K	Doktor	İstanbul	Beylikdüzü
K	Doktor	İstanbul	Beylikdüzü
K	Avukat	İstanbul	Beylikdüzü
K	Avukat	İstanbul	Beylikdüzü
K	Doktor	İstanbul	Beylikdüzü

Tablo-7: Meslek ve İlçe Dağılımı

Bu veride tek bir ilçedeki kadınların nüfusuna ait verinin meslek değişkeninde iki kategoride yığılma gürldüğünden bu kategorilerin birleşiminden tek bir kategori elde edilebilir ve bu durum veriyi daha güvenli hale getirecektir.

Cinsiyet	Meslek	İl	İlçe
K	Avukat veya Doktor	İstanbul	Beylikdüzü
K	Avukat veya Doktor	İstanbul	Beylikdüzü
K	Avukat veya Doktor	İstanbul	Beylikdüzü
K	Avukat veya Doktor	İstanbul	Beylikdüzü
K	Avukat veya Doktor	İstanbul	Beylikdüzü
K	Avukat veya Doktor	İstanbul	Beylikdüzü

K	Avukat veya Doktor	İstanbul	Beylikdüzü

Tablo-8: Meslek alanı anonimleştirilmiş veri kümesi

5. Bölgesel Gizleme

Bölgesel gizleme metotu, bir veya birden fazla değişkenin belli bir kayıda ait değerini *bilinmeyen* olarak değiştirmek anlamına gelmektedir. Bu yöntem, birden fazla değişkenin kombinasyonlarından kimlik saptaması açısından risk teşkil eden kayıtlar için uygulanarak kayıtlar daha güvenli hale getirilmektedir. Örneğin, şöyle bir kayda ait iki farklı kombinasyonu ele alırsak; "Medeni Durum=Dul; Yaş=17; Meslek=Öğrenci" ve "Medeni Durum=Dul; Yaş=17; Meslek=Öğrenci; Cinsiyet=Kadın" her iki kombinasyon da azınlıkta kalan bir nüfusa denk geldiğinden risk teşkil etmektedir ve bu sebeple medeni durum hanesi "bilinmez" olarak kaydedilerek her iki kombinasyon da eşzamanlı olarak güvenli hale dönüştürülebilmektedir[147]. Bu yöntem kayıt bazlı uygulanmaktadır. Yukarıdaki örnekte sadece ilgili kayıt içinde bulunduğu kombinasyondan dolayı riskli olarak değerlendirilmektedir. Veri kümesindeki diğer kayıtlar bu tip bir özellik göstermediği durumda o kayıtlara ait değişkenin *bilinmiyor* olarak değiştirilmesine gerek yoktur.

Aşağıdaki tabloda ilgili kaydın içinde bulunduğu bir veri kümesi örneği yer almaktadır ve görüldüğü üzere sadece riskli kayıt için değişiklik yapılmıştır.

Yaş	Cinsiyet	Meslek	Medeni Durum
17	K	Öğrenci	"Bilinmiyor"
28	E	Akademisyen	Evli

[147] International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201

16	E	Öğrenci	Bekar
35	K	Avukat	Evli

Tablo-9: Meslek ve Medeni Dağılımı Anonimleştirilmiş Veri Kümesi

6. Örnekleme

Örnekleme metotuyla bütün veri kümesi yerine, kümeden alınan bir örnek küme ifşa edilir veya paylaşılır. Böylelikle bütün veri kümesinin içinde yer aldığı bilinen bir kişi için bile ifşa edilen örnek alt küme içinde bu kişinin yer alıp almadığı bilinmediği için kişilere dair isabetli tahmin üretme riski düşmüş olur. Örnekleme yapılacak alt kümenin belirlenmesinde basit istatistik metotları kullanılır.

B. Değer Düzensizliği Sağlayan Metotlar

Veri düzensizliği metotlaruyla "orijinal veri kümesindeki tekil kombinasyonlar değiştirelerek yeni tekil kombinasyonlar yaratılır ve böylelikle düzensizleştirilmiş bir küme oluşturulur. Bu yeni oluşan veri kümesinin istatistik değerleri orijinal kümedeki hesaplanan değerlerle aynı olmalıdır[148].

1. Mikro-Birleştirme

[148] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, s. 54

Bu metot ile bütün veri kümesindeki kayıtları öncelikle anlamlı bir sıraya göre dizip sonrasında bütün kümeyi belli bir sayıda alt kümelere ayrılır. Sonrasında her alt kümenin belirlenen değişkene ait değerinin ortalaması alınarak her alt kümenin o değişkenine ait değer ortlama değer ile değiştirilir. Böylece o değişken tüm veri kümesi için geçerli olan ortalama değeri de değişmektedir. Her bir grup belirlenmiş en az *k* gruba ayrılır, *k* bir eşik değerini göstermektedir ve bu işleme k-kümelendirme adı verilir[149]. Mikro birleştirme bağımsız olarak tek bir kümeye uygulanırsa, bu metota bireysel dizilim, eğer her bir grup için bütün değişkenlerin aynı anda ortalaması hesaplanırsa, bu metota çok değişkenli mikro-birleştirme adı verilir[150].

Mikro-birleştirme metotunun yapılabilmesi için n adet kayıt içeren bir mikro veri kümesinin en az *k* adet kayıt içeren *g* adet gruba bölünmesi gerekir. Her bir değişken için, her bir grup için ortalama değeri hesaplanır ve bu değer her bir kaydın o değişken için değeri olarak atanır[151].

Aşağıdaki örnekte Tablo-10 *Gelir* değişkenine göre sıraya dizilmiş bir veri kümesini göstermektedir. Tablo-11 ise mikro-birleştirme hesaplaması yapıldıktan sonraki durumdur.

Yaş	Cinsiyet	Posta Kodu	Gelir
23	K	1556	20.000
37	K	1559	23.000
41	E	1559	32.000
25	K	1557	44.000
34	E	1558	57.000
48	E	1556	72.000

Tablo-10: Gelir Dağılımı

Bu veri kümesini her biri 3 kayıt içerek 2 gruba ayırır ve her bir grubun ortalama gelir değerini hesaplarsak durum aşağıdaki gibi olacaktır.

[149] Enformasyon Komiserliği Ofisi, s. 90
[150] International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201
[151] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, s. 58

Yaş	Cinsiyet	Posta Kodu	Gelir
23	K	1556	25.000
37	K	1559	25.000
41	E	1559	25.000
25	K	1557	57.666
34	E	1558	57.666
48	E	1556	57.666

Tablo-11: Mikro Birleştirme ile anonimleştirilmiş veri kümesi

2. Veri Değiş-Tokuşu

Bu metot, kayıtlar içinden seçilen çiftlerin arasında bir değişken alt kümesine ait değerlerin değiş-tokuş edilmesiyle elde edilen kayıt değişiklikleridir[152]. Bu metot temel olarak kategorize edlebilen değişkenler için tasarlanmıştır ve ana fikir mahrem değişkenlerin değerlerini bireylere ait kayıtlar arasında değiştirerek bir veri tabanının dönüştürülmesidir[153]. Değiş-tokuş yapılacak değişkenlere değiş-tokuş nitelikleri, başlangıçta değiş-tokuş yapılması için seçilen kayıt sayısının veri kümesindeki bütün kayıtlara oranına değiş-tokuş oranı, en uygun değiş-tokuş işleminin hangi çiftler arasında olacağını belirleyen değişkenlere kısıtlayıcı nitelik denir[154]. Bu metotun kolaylıkları ve yararları arasında şu maddeler sıralanabilir: "kişilerle ilgili kesin bilgiyi maskeler, kayıtla gerçek kişi arasındaki ilişkiyi keser, programlanması oldukça kolaydır, hassas olmayan ve kimliği temsil edilemeyen alanlara karşılık gelen değerleri bozmadan bir veya birden fazla değişkenler üzerinde uygulanabilir, hem kategroize edilebilen hem de süreklilik arz eden değişkenler üzerinde uygulanabilir[155]".

---

[152] International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201

[153] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Reiner Lenz, Jane Naylor, Eric Schulte Nordholt, Gionavvi Seri, Peter-Paul De Wolf, s. 58

[154] Enformasyon Komiserliği Ofisi, s. 92

[155] Richard A. Moore, Jr, *Controlled Data-Swapping Techniques for Masking Public Use Microdata Sets*, US Bureau of the Census Washington, 1996, s. 4, bkz. http://www.census.gov/srd/papers/pdf/rr96-4.pdf

Aşağıdaki örneklerde Tablo-12 orijinal veriyi, Tablo-13 ise gelir bilgisinin rasgele değiş-tokuş durumunu göstermektedir.

Yaş	Cinsiyet	İl	Gelir
21	K	İstanbul	20.000
24	K	Ankara	30.000
35	E	İzmir	30.000
36	K	İstanbul	25.000
45	E	İzmir	55.000
50	E	İzmir	15.000

Tablo-12: Gelir ve İl Dağılımı

Yaş	Cinsiyet	İl	Gelir
21	K	İstanbul	**25.000**
24	K	Ankara	*55.000*
35	E	İzmir	15.000
36	K	İstanbul	**20.000**
45	E	İzmir	*30.000*
50	E	İzmir	30.000

Tablo 13: Veri Değiş-Tokuşu ile anonimleştirilmiş Veri Kümesi

3. PRAM Metotu

PRAM metotu kategorize edilebilen değişkenlere uygulanabilen, veri kümesindeki belli değişkenlerin değerlerinin tayin edilmiş bir olasılık mekanizmasına göre değiş-tokuş edildiği bir metottur ve bu değişiklik sonucunda değiştirillmiş her değer orijinal değerden farklı olabilir veya olmayabilir[156]. Bu haliyle PRAM, veri değiş-tokuşu metotunun rasgele veriyonu olarak kabul edilebilir[157]. Yayınlanan veri kümesinde, belli kayıtlar için bazı kategorik değişkenlerin değerleri Markov Matrisi adı verilen bir olasılık mekanizmasına göre farklı bir değere dönüştürülür[158]. Değerlerin belli bir olasılık oranına göre değiş-tokuş edilmesi sonucunda oluşan yeni veri kümesinde o kayıdın kimi temsil ettiğini tahmin etmek güçleşmektedir.

4. Gürültü Ekleme

Bu yöntem, sayısal değerlere uygulanır ve değerlerde belli oranlarda yapıan positif veya negatif bozulalar ile orijinal değerler değiştirilmiş olur. Sağlanacak bozulma her değere belli bir oranda dağıtılır, birbirini karşılayacak şekilde dağıtılır. Bu yüzden toplamda değişiklik yaşanmazken kayıt bazlı değişkenlerde küçük artış veya azalmalar yaşanır. Böylelikle değerlere dair tahminler üretmenin veya gerçek değerin görüntülenmesi engellenmiş olur. Ancak eğer kayıtların değerleri arasında çok büyük farklar veya bazı aykırı örnekler varsa bu yöntem etkili olmaz[159].

5. Tekrar Örnekleme

[156]Bill Gross, Philippe Guiblin, Katherine Merrett, s. 1
[157]International Household Survey Network, *Reducing the Disclosure Risk,* bkz. http://www.ihsn.org/home/node/201
[158] Bilgi Komiserliği Ofisi, s. 94
[159] Bilgi Komiserliği Ofisi, s. 96

Tekrar örnekleme de yalnızca sayısal değerler için geçerli bir metottur. Tekrar örnekleme, orijinal veri kümesindeki n değerin t örneğinin sırlanıp ortalamaları alınarak yer değişikliğine uğraması ile sayısal değerlerde gerçekleştirilen bir koruma yöntemidir[160]. Öncelikle bütün nüfus içindeki belli bir değişken için dağılımı ve ilişkili değişkenlerin değerlerinin dağılımları tahmin edilir, sonrasında yapılan tahminle aynı değişken değerlerine sahip bozulmuş bir örnekleme yapılır, son olarak da bozulmuş örnekler ile orijinal veri kümesindeki değerler değiştirilir[161].

C. Anonimleştirmeyi Kuvvetlendirici İstatistik Metotları

Anonimleştirilmiş veri kümelerinde dahi betimleyicilerin doğru kombinasyonlarla bir araya gelmesi halinde kayıtlardaki kişilerin kimliklerinin saptanabilir olması veya belirli bir kişiye dair bilgilerin rahatlıkla tahmin edilebilir hale gelmesi anonimleştirme süreçlerine dair olan güveni sarsmıştır. Buna istinaden çeşitli istatistik metotlarıyla anonimleştirilmiş veri kümelerinin daha güvenilir duruma getirilmemesi gerekmiştir. Anonimleştirilmiş veri kümelerinin güvenlik açıklarına istinaden en çarpıcı örnek Amerika'da seçmen listelerine dair yapılan araştırmalardır. Bu araştırmayla "A.B.D nüfusunun %87'sinin (248 milyon içinden 216 milyonu) 5-haneli Posta kodu, cinsiyet ve doğum tarihi bilgilerinin birleşmesinden oluşan tekil karakteristik ile kimliğinin saptandığı[162]" ortaya çıkartılmıştır.

Bu konuda tasalanmış temel istatistik metotları sırasıyla k-anonimlik, l-çeşitlilik, t-gizlilik metotlarıdır. Bu üç metot arasında gelişim ve güvenilirlik anlamında hiyerarşik bir sıra bulunmaktadır. İlk tanıtılan k-anonimlik metotunun açıklarına istinaden l-çeşitlilik metotu geliştirilmiş, daha sonrasında l-çeşitliik metotunun

[160]International Household Survey Network, *Reducing the Disclosure Risk*, bkz. http://www.ihsn.org/home/node/201
[161] Bilgi Komiserliği Ofisi, s.98
[162] Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, Carnegie Mellon University, 2002, s. 2

eksiklerine istinaden de t-gizlilik metotu geliştirilmiştir. Böylece her bir versiyonda kimliğin saptanabilmesi veya ifşa edilmiş veriye bakarak isabetli tahminler yapılabilmesi olasılıkları minimize edilmiştir.

1. K-Anonimlik

K-anonimlik, "bir veri kümesindeki belli alanlarla, birden fazla kişinin tanımlanmasını[163]" sağlayarak, belli kombinasyonlarda tekil özellikler gösteren kişilere özgü bilgilerin açığa çıkmasını engellemek için geliştirilmiştir. Bir veri kümesindeki değişenlerden bazılarının bir araya getirerek oluşturduğu kombinasyonlara ait birden fazla kayıt bulunması halinde bu kombinasyona denk gelen kişilerin kimliklerinin saptanabilmesi olasılığı düşmektedir. Diğer bir deyişle k-anonimlik, her eşitlik sınıfının en az k değer içermesini gerektirir[164]. Bu durumda ilgili değişken kombinasyonları veri kümesinin bütününde diğer hassas değişkenlere de sahip olduklarında, her bir hassas verinin kime ait olduğu bilgisi muğlaklaşmaktadır. K değerini aynı kombinasyonlarda farklı hassas verilere sahip kayıt sayısı belirler. Eğer bir tablo belli bir k değeri için k-anonimlik metotuyla düzenlenmişse, bu durumda sadece dolaylı betimleyicilere ait değerleri bilen biri, belli bir kişiye ait kaydı 1/k olasılığından daha doğru bir saptamayla teşhis edemez[165].

Tablo-14'te "1983 tarihinde doğmuş cinsiyeti erkek olan, 3440* posta kodundaki adreste oturan kişi sayısı 5'dir ve tüm veri kümesi üzerinde doğum tarihi, cinsiyet

[163]Ahmet Bacak, *Gizliliği Koruyarak Veri Yayınlamak İçin K-Anonimity ve L-Diversity Metodları*, 2013, bkz. https://www.bilgiguvenligi.gov.tr/siniflandirilmamis/gizliligi-koruyarak-veri-yayinlamak-icin-k-anonimity-ve-l-diversity-metodlari.html

[164] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, s. 107
[165] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, s. 107

ve posta kodu alanlarına göre en az 5 kişi tanımlanabiliyorsa, bu veri kümesi üzerinde 5 anonimlik sağlanmış olur[166].” Bu örnekte k=5’tir.

TCKN	Ad-Soyad	Doğum Tarihi	Cinsiyet	Posta Kodu	Hastalık Adı
*	*	1980	K	3440*	Grip
*	*	1982	E	3440*	Hepatit-B
*	*	1980	K	3440*	Baş Ağrısı
*	*	1982	E	3440*	Beyin Tümörü
*	*	1983	E	3440*	Soğuk Algınlığı
*	*	1983	E	3440*	Yüksek Tansiyon
*	*	1983	E	3440*	Baş Ağrısı
*	*	1983	E	3440*	Astım
*	*	1983	E	3440*	Akciğer Kanseri

Tablo 14: k=5 anonimlik kümesine sahip anonimleştirilmiş veri kümesi

Görüldüğü üzere veri kümesi, öncesinde tckn, ad-soyad gibi direk betimleyicileri çıkartarak anonimleştirilmiştir. Ayrıca dolaylı betimleyici olma özelliğinden dolayı posta kodu alanında da kodlama yapılmış ve sadece ortak haneler görünür olarak bırakılmıştır. Normal şartlarda bu veri kümesini çeşitli anonimleştirilme metotları uygulanmış ve kimlik tespiti yapılabilecek özelliklerini kaybetmiş olarak değerlendirmekteyiz. Ancak doğum tarihi, cinsiyet ve posta kodu gibi dolaylı betimleyicilerin bir araya gelerek oluşturacağı kombinasyonlar özellikle dış veri ile birleşmesi halinde risk teşkil edileceğinden bu üçlü değişken kümesine ait kayıtlarda aynı kombinasyonu tutturanlardan birden fazla değer olması ifşa riskini daha da düşürmektedir. Örneğin bu veriye bakarak, 1983 yılında doğmuş, 3440* posta kodunda oturan bir erkeğin soğuk algınlığı, yüksek tansiyon, başağrısı,

[166] Ahmet Bacak, *Gizliliği Koruyarak Veri Yayınlamak İçin K-Anonimity ve L-Diversity Metodları*, 2013, bkz. https://www.bilgiguvenligi.gov.tr/siniflandirilmamis/gizliligi-koruyarak-veri-yayinlamak-icin-k-anonimity-ve-l-diversity-metodlari.html

astım, akciğer kanseri hastalıklarından hangisine sahip olduğunu bilmek mümkün değildir. Bu üç değişkenin aynı değerlere sahip olduğu 5'li bir alt sınıf yer aldığından bu kombinasyondaki değerlere uyan bir kişinin hastalığına dair yapılacak tahminin gerçek olma olasılığı 1/5 değerindedir. Aynı şekilde 1980 yılında doğmuş, 3440* posta kodunda oturan bir kadının da baş ağrısı ve grip hastalıklarından hangisine sahip olduğu bilinemez. Burada da olasılık ½'dir. Son olarak, 1982 yılında doğmuş, 3440* posta kodunda oturan bir kadının Hepatit-B ve beyin tümörü hastalıklarından hangisine sahip olduğuna dair de bilgi sahibi olunamaz. Bir önceki örnekteki gibi doğru kişiyi sapama olasılığı yine ½'dir. Aynı kombinasyonlarda birden fazla kayıt olması, kişilere dair hassas bilgiye erişmeyi zorlaştırmaktadır. Ek olarak, görüleceği üzere k değeri ne kadar yüksekse olasılık değeri o kadar düşmekte ve bu durum yüksek k değeri olan tabloların daha güvenilir olduğu anlamına gelmektedir.

Diğer taraftan, k-anonimlik pratikte her durumda gerekli güvenilirliği sağlayamamaktadır. Sweeney, k-anonimlik kullanılarak tasarlanmış veri kümelerine yapılacak saldırıların üç temel sebepten ötürü başarı sağlayacağını savunur; "hassas değişkenlere ait değerlerin sıralanışı ve tekrarlaması, tamamlayıcı sadırılar, zamansal saldırılar[167]". İlk durumda hassas değişkenlere ait değerler aynı kombinasyon sınıfları içinde tekrarlanırsa veya sıralanışları bir başka veri kümesi ile aynı şekildeyse bu durum kimlik saptama riskini arttırmaktadır. K-anonimlik uygulanmış gruplardaki hassas verilerin sayısıyla ilgilenmeyişi, k-anonimlik yaklaşımının en büyük eksikliğidir[168]. Yani yukarıdaki örnekten devam edecek olursak, eğer 1980 yılında doğmuş, 3440* no'lu posta kodunda oturan kadınların her ikisi de aynı hastalığa sahip olsalardı (grip ve bağ ağrısı yerine her ikisi de grip olabilirlerdi), bu durumda veri kümesinde olduğunu bildiğiniz ve bu üçlü kombinasyondaki değerlere sahip olduğu (1980; K; 3440*)

[167] Latanya Sweeney, s. 10-11

[168] İlkay Demirci, T-Closeness Metodu Gizliliği Koruyarak Veri Yayınlamak İçin, 2014 bkz. http://www.phphocam.com/t-closeness-metodu-gizliliği-koruyarak-veri-yayinlamak-icin/#sthash.z70qZ2sb.dpuf

bilinen bir kişinin grip olduğunu bu tabloya bakarak tespit etmek mümkün olacaktı. Benzer şekilde, aynı veri kümesine ait anonimleştirilmiş iki alt kümeyi aşağıdaki sıralamada düşündüğümüzde, iki alt kümenin birleştirilmesinden rahatlıkla bir sonuca varılabilir.

Etnik Köken	Posta Kodu
Asya	34789
Asya	34769
Afrika	34455
Afrika	34467
Avrupa	34478
Avrupa	34489
Avrupa	34562

Tablo-15

Etnik Köken	Posta Kodu
*	34789
*	34769
*	34455
*	34467
*	34478
*	34489
*	34562

Tablo-16

Etnik Köken	Posta Kodu
Asya	347**
Asya	347**
Afrika	344**
Afrika	344**
Avrupa	344**
Avrupa	344**
Avrupa	345**

Tablo-17

Yukarıdaki tablolarda Tablo 16 ve 17, Tablo 15'in anonimleştirilmiş versiyonlarıdır. 16 ve 17 no'lu tablolara ayrı ayrı bakıldığında orijinal tablodaki verilere istinaden bir çıkarım yapılamaz. Ancak ikisinin sıralanışındaki benzerlik, birlikte değerlendirilmeleri halinde orijinal veri kümesi olan Tablo-15'e ulaşmamızı sağlamaktadır.

Tamamlayıcı sebepler olarak saydığımız ikinci sebepte de, yukarıdaki sıralama örneğindekine benzer şekilde orijinal veri kümesinden anonimleştirilerek türetilmiş iki alt kümenin birleştirilmesi sonucunda orijinal veri kümesinin elde edilmesidir. Ancak bu durumda alt kümelerin sıralanalışları farklı dahi olabilir. Ancak yine de birleşimleri orijinal kümeyi ortaya getirmektedir.

Zamansal sebepler de ise orijinal veri kümesinin zaman içinde uğradığı değişimlere istinaden oluşan açıklardır. "Veri toplama süreci dinamik bir süreçtir, […], satırlar düzenli olarak eklenir, değişir ve kaldırılır[169]". Kayıt ekleme, çıkarma veya değerlerdeki değişiklikler k değerini etkileyeceğinden, ilk ifşa edilen anonimleştirilmiş alt kümelerde kombinasyonlara ait sınıflar yeteri koruma

[169] Latanya Sweeney, s. 12

sağlayamaz duruma gelebilir. Bu durum ise ilk versiyonda elde edilen güvenilirliğini zaman içinde bozulabileceği anlamına gelir.

Bu tip eksiklikler yüzünden k-anonimlik kimlik saptanmasının engellenmesi açısından başarılı sonuçlar üretirken, özellik saptama boyutunda yeterli bulunmamaktadır[170]. Sonuç olarak, güvenilirliği arttırıp riski azaltmak adına yapılan gelişmeler l-çeşitlilik metotunun kurgulanmasını sağlamıştır.

2. L-Çeşitlilik

K-anonimliğin eksikleri üzerinden yürütülen çalışmalar ile oluşan l-çeşitlilik meototu aynı değişken kombinasyonalrına denk gelen hassas değişkenlerin oluşturduğu çeşitliliği dikkate almaktadır. Yukarıdaki örnekte göreceğimiz gibi, aynı kombinasyonların karşılığı olan hassas değişkenlere ait değerler ne kadar çok çeşitlilik gösterirse tahmin edilebilme ihtimalleri o kadar düşer.

Bu yöntem geliştirilirken ve aynı şekilde k-anonimlik yönteminin açıkları dikkate alınırken her zaman akılda tutulan dış veriye sahip bir kişinin veri kümesine bakarak çıkarımlar yapabileceği gerçeğidir. Bir akrabasının veya komşusunun yaşı, cinsiyeti, posta kodu, adresi gibi bilgilerini kişisel ilişkileri sayesinde bilen bir kişi anonimleştirilmiş dahi olsa bu bilgilere dair kayıtların yer aldığı bir veri kümesinden tanıdığı kişiye dair sonuçlar üretebilir. Dış veriye sahip olan kişinin ifşa edilmiş veri kümesine erişimi olduğu, bu veri kümesinin başka verilerin genelleştirilmiş bir versiyonu olduğu ve bu kümede bildiği bazı kişilerin verilerinin bulunduğunu bildiği varsayılmaktadır[171]. Ancak l-çeşitlilik yöntemi ile

[170] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, s. 107
[171] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, *l-Diversity: Privacy Beyond k-Anonymity*, Cornell University, 2007, s. 2

belli bir kombinasyonun oluşturduğu alt sınıfın hassas değişkenlerine dair değerler "iyi temsil edilmiş[172]" şekilde kurgulanırsa her alt sınıfta l çeşitliliği sağlanacağından verilere bakarak çıkarım yapmak isteyen kişin en az l kadar bilgiye sahip olması gerekir.

Bir örnekle açıklayacak olursak, aşağıdaki orijinal tablo posta kodu, yaş, etnik köken ve hastalık değişkenlerini gösteren bir veri kümesidir. Hiçbir direk betimleyici içermiyor olmasına rağmen posta kodu, yaş ve uyruk gibi dolaylı betimleyicilerin oluşturduğu kombinasyonlar risk teşkil eder. Bunun yanında hastalık değişkeni ise bir hassas değişkendir. Yani, kendi kişisel ilişkileri sayesinde belli bir kişinin posta kodunu, cinsiyetini ve uyruğu bilen bir başka kişi bu veriye eriştiğinde bu üçlü kombinasyon sayesinde kişinin hassas verisine ulaşabilir.

Posta Kodu	**Yaş**	**Etnik Köken**	**Hastalık**
13053	28	Rusya	Kalp
13068	29	Amerikan	Kalp
13068	21	Çin	Viral Enfeksiyon
13053	23	Amerikan	Viral Enfeksiyon
14853	50	Hindistan	Kanser
14853	55	Rusya	Kalp
14850	47	Amerikan	Viral Enfeksiyon
14850	49	Amerikan	Viral Enfeksiyon
13053	31	Amerikan	Kanser
13053	37	Hindistan	Kanser
13068	36	Japonya	Kanser
13068	35	Amerikan	Kanser

Tablo-18: Etnik Köken ve Hastalık Dağılımı

Bu durumda ilk önce, orijinal tablo çeşitli anonimleştirme metotları uygulanır ve k-anonimlik metotuna uygun olarak aşağıdaki gibi yeniden tasarlanır.

Posta Kodu	**Yaş**	**Uyruk**	**Hastalık**
130**	< 30	*	Kalp
130**	< 30	*	Kalp

[172] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer,, s. 6

130**	< 30	*	Viral Enfeksiyon
130**	< 30	*	Viral Enfeksiyon
1485*	≥ 40	*	Kanser
1485*	≥ 40	*	Kalp
1485*	≥ 40	*	Viral Enfeksiyon
1485*	≥ 40	*	Viral Enfeksiyon
130**	3*	*	Kanser
130**	3*	*	Kanser
130**	3*	*	Kanser
130**	3*	*	Kanser

Tablo-19 : k=4 şeklinde annimleştirilmiş veri kümesi

Görüldüğü üzere kodlama ve kayıt çıkartma metotlarıyla anonimleştirilmiş veri kümesi ayrca k=4 olacak şekilde de k-anonimlik metotuyla anonimleştirilmiştir. Normal şartlarda bütün bu uygulamaların kişilere ait özelliklere eirşmeyi engellemesini beklenirken son dörtlü sınıfa baktığımızda belli bir posta kodu, yaş, etnik köken kombinasyonuna ait hassas verilere ait değerlerin hepsinin aynı olduğunu görmekteyiz. Böylelikle posta kodu 130** şeklinde olan, 30'lu yaşlarda bir akrabasının bu istede yer aldığını bilen bir kişi rahatlıkla akrabasının hastalığının kanser olduğunu anlayabilecektir. Çünkü alt sınıfa denk gelen hassas değişkenler yeterli çeşitliliği gösterememektedir.

Ancak farklı bir düzenleme ile aynı tablo aşağıdaki şekilde kurgulanırsa daha farklı sonuçlar doğar.

Posta Kodu	**Yaş**	**Uyruk**	**Hastalık**
1305*	≤ 40	*	Kalp
1305*	≤ 40	*	Viral Enfeksiyon
1305*	≤ 40	*	Kanser
1305*	≤ 40	*	Kanser
1485*	> 40	*	Kanser
1485*	> 40	*	Kalp
1485*	> 40	*	Viral Enfeksiyon
1485*	> 40	*	Viral Enfeksiyon
1306*	≤ 40	*	Kalp
1306*	≤ 40	*	Viral Enfeksiyon

1306*	≤ 40	*	Kanser
1306*	≤ 40	*	Kanser

Tablo-20 : k=4 ve l=3 şeklinde anonimleştirilmiş veri kümesi

Bu haliyle posta kodu gizleme metotunda bazı değişiklikler yapılarak 1. ve 3. sınıfların sıralanışı değiştirilmiştir. Böylelikle hastalık çeşitliliği her bir sınıfta artarak kimlik saptayabilme olasılığı düşürülür. Yukarıdaki örnekte tarif ettiğimiz üzere, 1306* no'lu posta koduna sahip ve 40 yaşın altında olduğu bilinen bir kişinin kalp, kanser veya viral enfeksiyon hastalıklarından hangisine sahip olduğu bu veri kümesine bakılarak saptanamaz çünkü bu at sınıfta l=3 olacak şekilde bir çeşitilik sağlanmıştır.

Ancak l-çeşitliliği metotunun da koruyamadığı durumlar oluşmaktadır. Alt sınıfların her birinde yer alan hassas değişkene ait değerler semantik olarak farklı ancak içerik olarak yakınlık gösteriyorsa yine kişilere dair hassas verilerin yakın değerlerle açığa çıkması riski oluşur. Bu durumu engellemek içinde t-yakınlık metotu geliştirilmiştir.

3. T-Yakınlık

l-çeşitlilik yöntemi hassas verilerde çeşitlilik sağlıyor olmasına rağmen hassas verilerin içeriğiyle ve hassasiyet derecesiyle ilgilenmediği için yeterli korumayı sağlayamadığı durumlar oluşmaktadır. Bu haliyle hassas değerlerin kendi içlerinde birbirlerine yakınlık derecelerinin hesaplanması ve veri kümesinin bu yakınlık derecelerine göre alt sınıflara ayrılarak anonimleştirilmesi sürecine t-yakınlık metotu denmektedir. Eğer bir hassas değişkene ait değerin içinde bulunduğu sınıftaki dağılımı ve bütün tabloda yer alan dağılımı arasındaki mesafe

belli bir t eşik seviyesinden yüksek ise bu durumda bu eşitlik sınıfına t-yakınlık içerir denir[173].

Doğum Tarihi	Cinsiyet	Posta Kodu	Hastalık	Hasta Sayısı
198*	E	3440*	Grip	80
198*	E	3440*	Tansiyon	20
198*	E	3440*	Baş Ağrısı	70
197*	E	3440*	Kanser	10
197*	E	3440*	Beyin Tümörü	10
197*	E	3440*	Hepatit-B	10

Tablo-20: k=3 ve l=3 şeklinde anonimleştirilmiş veri kümesi

ablo 20 ‘de görüldüğü üzere, doğum tarihi, cinsiyet ve posta kodu alanlarına göre k=3 olacak şeklinde k-anonimlik ve l=3 olacak çekilde l-çeşitlilik sağlanmasına rağmen 197* yılında doğmuş, 3440* adresinde oturan ve cinsiyeti erkek olan bir kişinin hastalıkları kanser, beyin tümörü ve hepatit b gibi ciddi hastalıklar olduğu için, bu k anonileştirilmiş grupta %100 oranında bu kişinin hastalığının ciddi olduğu tespit edilebilir[174]. Bu durumda hassas değişken olan hastalık değişkeninin aldığı değerlere göre kendi içinde de hassasiyetinin derecelendirilmesi gerekmektedir. Bu durumda hasta sayıları dikkate alınarak yapılan hesaplamalar sonucunda tablo aşağıdaki gibi yeniden tasarlanır.

Doğum Tarihi	Cinsiyet	Posta Kodu	Hastalık	Hasta Sayısı
≥ 1970	E	3440*	Grip	80
≥ 1970	E	3440*	Kanser	10
≥ 1970	E	3440*	Tansiyon	70
1975 ≤ x	E	3440*	Baş Ağrısı	20

[173]Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, s. 107

[174] İlkay Demirci, *T-Closeness Metodu Gizliliği Koruyarak Veri Yayınlamak İçin*, 2014 bkz. http://www.phphocam.com/t-closeness-metodu-gizliliği-koruyarak-veri-yayinlamak-icin/#sthash.z70qZ2sb.dpuf

≤1985				
1975 ≤ x ≤1985	E	3440*	Beyin Tümörü	10
1975 ≤ x ≤1985	E	3440*	Hepatit-B	10

Tablo-21: t-yakınlık ile anonimleştirilmiş veri kümesi

Böylece bu durumda hangi kombinasyon sınıfında olduğu bilinen bir kişi için bile hem hangi hastalığa sahip olduğu saptanamadığı gibi hem de hastalığının ciddi bir hastalık olup olmadığı da tespit edilemez.

III. Hukuksal Altyapı

Bu bölüm veri anonimleştirmesinin AB ve Türkiye mevzuatlardaki yer aldığı maddelerini ele alacaktır. Buna bağlı olarak hukuksal süreçlerin veri anonimleştirmesine karşı tutumu analiz edilecektir.

Öncelikle Avrupa Birliği mevzuatını inceleyecek olursak 95/46/AT sayılı yönergenin 26. paragrafında yer alan "veri koruması prensibi ilgili kişinin tespit edilmesine imkan vermeyecek şekilde anonimleştirilmiş verilere uygulanmaz[175]" ifadesi ile anonimleştirilmiş veriler tüm veri koruma süreçlerinin istisnası olarak belirlenmiştir. Aynı paragrafa göre "veri koruması prensipleri kimliği saptanmış veya saptanabilir nitelikte olan kişiler için[176]" uygulanmalıdır ve buna bağlı olarak anonim verilerin kişilerle olan ilişkisi koparılmış olduğu için bu veri kümelerine veri koruması prensiplerinin uygulanmasına gerek görülmemiştir.

2002 yönergesinde ise anonim veri ve anonimleştirme süreçlerine atıflar daha fazla yerde geçmektedir. Buna göre, 9. paragrafta yer alan "veri işlemenin mininmize edilmesi veya imkan olan durumlarda anonim veya gizlenmiş veri

[175] 95/46/AT sayılı Yönerge, parag. 26
[176] 95/46/AT sayılı Yönerge, parag. 26

kullanılması[177]" gerekliliği vurgulanmıştır. 26. paragrafta, verilerin hangi amaçlarla ve hangi şartlarda işleneceğine dair esaslar yer almaktadır ve burada yer alan "pazarlama iletişim servisleri veya katma değerli hizmetleri hazırlamak için kullanılan trafik verisi, servisin karşılanmasından sonra anonimleştirilmeli veya silinmelidir[178]" ifadesi ile anonimleştirme süreçleri ile silme işlemi bir tutulmaktadır. Devamında yer alan 28. paragrafa göre, "iletişimin sağlanması amacıyla tutulan trafik verilerine ihtiyaç kalmadığında silinmeleri veya anonimleştirilmeleri, Internet'te yer alan alan adı sistemlerinin IP adreslerinin ön bellekte tutulması veya fiziksel adres bağlayıcıların IP adreslerinin ön bellekte tutulması veya ağa veya servislere erişim haklarını yönetmek için kullanılan oturum bilgilerinin saklanması gibi prosedürlerle çakışmaz[179]". Bu durumda bir kez daha anonimleştirilme süreçleri ve silme süreçleri denk olarak ele alınmıştır. Aynı şekilde yönergede trafik verisine istinaden esasları belirleyen 6. maddenin 1. fıkrasında trafik verilerinin iletişimin transmisyonu için gereken ihtiyaç ortadan kalktıktan sonra "silinmesi veya anonimleştirilmesi" gerektiği belirtilmektedir. 9. maddenin birinci fıkrası ise "trafik verisi dışındaki lokasyon verisinin ancak anonim olmak şartıyla veya kullanıcı veya abonenin rızası ile işlenebileceğini belirtir. Bir önceki maddelerde verinin silinmesi ile eşdeğer tutulan anonimleştirme süreçleri bu maddeyle abone/kullanıcı rızası alınması süreci ile de denk olarak değerlendirilmiştir.

Ülkemizde yürürlükte olan ve kişisel verilerin ve özel alan gizliliğinin korunmasına yönelik olan mevzuatlarda Veri Koruması Yönetmeliği ve Türk İstatistik Kanunu kapsamında anonimleştirme hükümleri bulunmaktadır. Bunun dışında henüz yürürlüğe girmemiş olan "Kişisel Verilerin Korunması Kanunu taslağı[180]" ve "Kişisel Sağlık Verilerinin İşlenmesi ve Veri Mahremiyetinin

[177] 2002 Yönergesi, parag. 9
[178] 2002 Yönergesi, parag. 26
[179] 2002 Yönergesi, parag. 28
[180] Adalet Bakanlığı'nca hazırlanan ve Bakanlar Kurulu'nca 7/4/2008 tarihinde kararlaştırılan *Kişisel Verilerin Korunması Kanun Tasarısı*, bkz. http://www2.tbmm.gov.tr/d23/1/1-0576.pdf

Sağlanması Hakkında Yönetmelik taslağı[181]" anonimleştirme esasları ve süreçlerine dair bilgi vermektedir ancak her iki metin de henüz yasalaşmamıştır.

Yürürlükte olan ve elektronik haberleşme sektöründeki kişisel verilerin işlenmesini ve gizliliğinin korunmasını amaçlayan yönetmelikte anonimleştirme süreçleri "anonim hale getirme" olarak ifade edilmiş ve şu şekilde tanımlanmıştır: "Kişisel verilerin, belirli veya kimliği belirlenebilir bir gerçek ya da tüzel kişiyle ilişkilendirilemeyecek veya kaynağı belirlenemeyecek hale getirilmesi[182]". Avrupa Birliği mevzuatından farklı olarak, bu ifade ile tüzel kişilere ait verilerin de anonimleştirilebileceği sürece dahil edilmiştir. Aynı yönetmelikte 8. maddenin 3. ve 4. fıkralarında sırasıyla "Elektronik haberleşme hizmetlerini pazarlamak veya katma değerli elektronik haberleşme hizmetleri sunmak amacıyla ihtiyaç duyulan trafik verileri anonim hale getirilerek veya ilgili abonelerin/kullanıcıların işlenecek trafik verileri ve işleme süresi hakkında bilgilendirilmelerinden sonra rızalarının alınması kaydıyla, alınan rızaya uygun olarak sadece katma değerli elektronik haberleşme hizmetlerinin, pazarlama faaliyetlerinin ve benzer hizmetlerin gerektirdiği ölçü ve sürede işlenebilir[183]" ve "Abonelere/kullanıcılara ait işlenen ve saklanan trafik verileri, bu verilerin işlenmesini ve saklanmasını gerekli kılan faaliyetin tamamlanmasından sonra silinir veya anonim hale getirilir[184]" ifadeleri yer alır. Aynı yönetmelikte 11. maddede yer alan "Katma değerli elektronik haberleşme hizmetleri sunmak amacıyla ihtiyaç duyulan ve trafik verisi niteliğinde olmayan konum verileri, anonim hale getirilerek veya ilgili abonelerin/kullanıcıların işlenecek konum verileri, işleme amacı ve süresi hakkında bilgilendirilmelerinden sonra rızalarının alınması kaydıyla, alınan rızaya uygun olarak sadece katma değerli elektronik haberleşme hizmetlerinin gerektirdiği ölçü ve sürede işlenebilir[185]" ifadesiyle Avrupa Birliği mevzuatıyla uyumlu olacak şekilde anonimleştirme süreçleri abone/kullanıcı rızasıyla denk

[181] *Kişisel Sağlık Verilerinin İşlenmesi ve Veri Mahremiyetinin Sağlanması Hakkında Yönetmelik taslağı, bkz.* http://www.saglik.gov.tr/TR/belge/1-17634/kisisel-saglik-verilerinin-islenmesi-ve-veri-mahremiyet-.html
[182] Veri Koruması Yönetmeliği, madde. 3/c
[183] Veri Koruması Yönetmeliği, madde. 8/3
[184] Veri Koruması Yönetmeliği, madde. 8/4
[185] Veri Koruması Yönetmeliği, madde. 11

tutulmaktadır. Ek olarak 15. maddede yer alan "işlenen ve saklanan verinin, saklama süresinin bitiminden itibaren en geç bir ay içinde imha edilmesi veya anonim hale getirilmesi ve bu işlemlerin tutanakla veya sistemsel olarak kayıt altına alınmasını sağlamakla yükümlüdür[186]." ifadesi yine AB mevzuatıyla paralel olarak silme ve anonimleştirme süreçlerini denk kabul etmektedir.

Türk İstatistik Kanununda ise gizli verinin tanımına dair önemli bir detay 13. maddenin 2. paragrafında verilmektedir. Buna göre, "bireysel verinin toplulaştırılması ile oluşturulan veri tablosunun herhangi bir hücresindeki istatistikî birim sayısının üçten az olması veya birim sayısı üç ve daha fazla olduğu hâlde bir veya iki istatistikî birimin hakim durumda olması hâlinde ilgili hücredeki veri gizli kabul edilir[187]" ifadesi özünde dağılıma istinaden oluşabilecek güvenlik açıklarına gönderme yapmaktadır. Burada bahsedilen istatistiki birim "hakkında veri toplanacak gerçek ve tüzel kişiler ile kurum ve kuruluşları[188]" gösterdiğinden bu ifadeyle anlatılmak istenen toplu bir veri kümesinde bilgilerin üç veya daha az kişiye veya kuruma ait olması halinde veya üç kişiden/kurumdan fazla olsalar bile birinin küme içinde daha çok kayda sahip olması halinde bu veri kümesindeki veriler gizli veri muamelesi görmelidir. Burada amaç, grup çeşitliliğinin azalması halinde eldeki değişkenlerin kombinasonuyla kişilerin veya kurumların kimliklerinin saptanabilir olmasını engelleyecek önlemler almaktır. Bu madde, önceki bölümlerde detaylandırdığımız istatistik metotları olan k-anonimlik, l-çeşitlilik ve t-yakınlık metotlarının ana fikrini andırmaktadır. Kanunda gizli verinin doğrudan veya dolaylı özelliklerin birleşimiyle kişiyi açığa çıkartacak özelliği kabul edildiğinden 13. maddenin son paragrafındaki "gizli veriler, ancak doğrudan veya dolaylı tanımlamaya yol açmayacak şekilde diğer bilgilerle birleştirilerek yayımlanabilir[189]" ifadesi açıkça anonimleştirme süreçlerine atıf yapmaktadır. Böylelikle birleşirilmesi halinde bile kişi veya kurumu saptayabilme özelliği taşıyan ilişkilerin koparılması halinde gizli verinin ifşası mümkün olacaktır. Yine 13. maddenin 5. ve 6. paragrafları istisnaları

[186] Veri Koruması Yönetmeliği, madde. 15
[187] Türkiye İstatistik Kanunu, m. 13, parag. 2
[188] Türkiye İstatistik Kanunu, m. 2/h
[189] Türkiye İstatistik Kanunu, m. 13, parag. 7

belirler ve "herkese açık kaynaklardan elde edilen veri veya bilgiler[190]" gizli veri sayılmaz ve "istatistikî birimin, kendisine ait gizli verilerin açıklanmasına yazılı onay vermesi hâlinde, veri gizliliği ortadan kalkar[191]" takiben, kanunun 14. maddesi bireysel verilere dair "istatistikî birimlerin doğrudan veya dolaylı olarak tanınmasına yol açacak bölümleri gizlendikten sonra, münferit birimlere atıfta bulunmayan bilimsel amaçlı araştırmalarda kullanılması kaydı ve Başkanlığın yazılı izniyle verilebilir. Bireysel verileri kullanma hakkı elde edenler, bu verileri üçüncü şahıslara veremezler[192]" hükmüyle bireysel verilere dair yine bazı anonimleştirme süreçleri uygulandıktan sonra kullanım amacını ve paylaşım gruplarını kısıtlayarak paylaşılabileceğini veya ifşa edilebileceğini vurgular.

IV. Büyük Veri İçin Veri Anonimleştirmesi

Yukarıda teknik detaylarıyla anlatılan anonimleştirme metotları, örneklerde de görüldüğü üzere ilişkili veri tabanı veya veri ambarı yapıları bünyesinde tutulan yapılandırılmış veri kümelerine uygulanabilir. Büyük veri özellik itibariyle yapılandırılmamış veri türünde olduğundan, büyük verinin ancak yüksek kapasiteki yazılımlar tarafından işlenerek yapılandırılmış bir özellik kazanmasıyla anonimleştirme süreçlerine tabi tutulması mümkün olacaktır. Önceki bölümlerde değindiğimiz gibi çeşitli yazılımlarla (Hadoopi uygulaması v.s) işlenerek, veri madenciliği için uygun hale getirilmiş, sınıflandırılmış ve işlenmiş büyük veri aynı zamanda bu ilişkili yapılar içinde anonimleştirme için de uygun hale getirilmiş olmaktadır. Bu süreçten sonra yukarıdaki anonimleştirme metotları aynen büyük veri içinde geçerli olacaktır.

[190] Türkiye İstatistik Kanunu, m. 13, parag. 5
[191] Türkiye İstatistik Kanunu, m. 13, parag. 6
[192] Türkiye İstatistik Kanunu, m. 14

4. Anonimleştirmenin Güvenilirliği Tartışması

Anonimleştirme süreçleri, kişisel verilerin ve özel alan gizliliğinin korunması hususunda, mevzuatlar nezdinde güvenilir süreçlerden biri olarak değerlendirilmekte ve hem kamu hem de özel sektör kuruluşları bu süreçlerle veri paylaşımı ve ifşasına teşvik edilmektedir. Hatta mevuzatlarda geçen ifadelerde gördüğümüz üzere, ilk bölümde detaylandırdığımız ifşa ve paylaşım adımlarının yanı sıra, verilerin saklanması süreçlerinde de anonimleştirme süreçlerinin uygulanması gerektiği ifade edilmektedir. Diğer taraftan, anonimleştirilmiş verinin, çeşitli durumlarda kişisel verilere ve kişinin sahip olduğu özelliklere istinaden bilgileri açık edebildiği keşfedilmiş ve bu durum anonimleştirilmiş veriye olan güveni sarsmıştır. Kişi ile ilişkisi kesilerek artık kimliğin saptanamadığı bir formata getirilen anonimleştirilmiş veri, yeniden ilgili bağlantılar kurularak geri çevrilebilir ve veriyi üreten kişiye dair bilgiler ifşa edebilir.

Anonim verinin güvenlik açıkları ile günümüzün enformasyonel toplumunu yönlendiren büyük verinin direk etkileşimi bulunmaktadır. İlk bölümde detaylandırdığımız üzere, gelişen veri işleme ve depolama teknikleri daha büyük ölçekte ve daha çeşitli verinin daha hızlı işlenebilmesini sağlamış, işlenen veri de gelişen veri madenciliği teknikleri ile daha anlamlı ve sistematik çıktılar/raporlar üretmiştir. Aynı zamanda bu yaygın ağ yapısının içinde, kişilerin kendi rızaları ile kamuya ifşa ettikleri pek çok kişisel detay da rahatlıkla erişebilir durumdadır. Bütün bu süreç ise anonimleştirilmiş verinin güvenilirliğini tehdit eden bir unsur olarak karşımıza çıkmaktadır. Anonimleştirilmiş verilerin dışarıdan elde edilen veri ile birleştirilmesi halinde anonimlik kolaylıkla bozulabilmekte veya isabetli tahminler yapılabilecek kayıtlar elde edilebilmektedir.

Anonimleştirilmiş verinin güvenlik açıkları sonucunda açığa çıkacak kişisel veriler ve ilgili kişiye dair hassas nitelikteki veriler bireylere maddi ve manevi farklı ölçekte zarar verebilmektedir. Zarar, hem bu hassas verilere yetkisiz ve

izinsiz olarak çeşitli yöntemlerle erişebilmiş art niyetli bir kullanıcıdan gelebileceği gibi, bireylere dair pek çok hassas veriyi kullanarak kontrol ve sansür mekanizmaları kuracak veya temel hak ve özgürlükleri ihlal edecek hükümetler tarafından da sağlanabilir. Özellikle kişilerin düşüncelerini ifade ettikleri iletişim ortamlarından elde edilen verilerin kontrolsüz şekilde ifşa edilmesi veya paylaşılması ciddi güvenlik sorunlarına veya manevi zararlara sebep olabilir. Belli bir düşünceyi ifade eden kişilerin isimlerinin kamuya ifşa edilmesi bu kişilerin hedef gösterilmesine, taciz edilmesine veya aileleri, arkadaşları, komşuları veya iş çevreleri tarafından sosyal baskılara maruz kalmalarına sebep olabilir[193]. Bu bağlamda, günümüzde veri odaklı ilerleyen ve hayatını sürdüren sektörleri, kuruluşları ve araştırmaları düşünürsek, anonimleştirilmiş verinin yaratacağı risklerin topluma ne büyüklükte zarar verebileceğini doğru yorumlayabiliriz.

I. Tehditler

İlerleyen bölümde anonimleştirilmiş verinin kişisel verileri ve özel alan gizliliğini tehdit edebileceği durumlar analiz edilecek ve ihlallerin hangi durumlarda gerçekleştiği incelenecektir.

A. Anonimleştirilmiş Veriden Kişisel Veriye Ulaşma

Gördüğümüz üzere anonimleştirilmiş veri, belli bir kişinin kimliğini veya özelliklerini saptayan ilişkilerin kesilmesi ile elde edilen veridir. Anonimlik bir kayıt ile bir öznenin kimliği arasındaki bağlantının koparılmasıdır[194]. Bu bağlantı direk ve dolaylı betimleyiciler ile kurulmakta ve bu betimleyicilerde yapılan

[193] Seda Gürses, Carmela Troncoso, Claudia Diaz, *Engineering Privacy by Design*, International Conference on Privacy and Data Protection (CPDP) Book, 2011, s. 9

[194] Andreas Pfitzmann, Marit Hansen, *Anonymity, Unobservability, Pseudonymity, and Identity Management:A Proposal for Terminology*, s. 4, bkz. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.18.pdf

düzenleme ve eksiltmelerle kopartılmaktadır. Bu durumda anonimleştirilmiş bir verinin betimleyicilerinin yeniden ilgili kayıt ile birleştirilmesi sonucunda, tersine bir işlem yapılarak anonim veriden yeniden kişisel veri elde edilmiş olur. Anonimleştirilmiş verinin en büyük riski başka bir veri kümesiyle birleşerek yeniden kimlik saptayabilir formata dönüşmesi, kişisel veri ve hassas verilerin kontrolsüzce yetkisiz kişilerin ve mercilerin eline geçmesine sebep olmasıdır.

Anonimleştirilmiş verinin yeniden kişisel veriye dönüştürülmesi için aradaki bağlantıları kuran betimleyicilere ihtiyaç bulunmaktadır. Bu betimleyiciler iki farklı veri kümesindeki ortak noktalar olacağından elimizdeki veri kümesi, anonimleştirilmiş bir veri kümesi dahi olsa kolaylıkla diğer veri kümesiyle birleşerek her bir kayıda dair daha fazla bilgi edinmemizi sağlayacaktır. Bir örnekle açıklayacak olursak, aşağıdaki Tablo-22 bir hastanenin acil servis birimine dair kapasite hesaplamalarını yapabilmek için oluşturduğu bir veri kümesi olsun.

Yaş	Cinsiyet	Posta Kodu	Teşhis	Sigorta Güvencesi	Hastane Giriş Tarihi	Hastane Çıkış Tarihi
23	K	34798	Grip	SSK	01.01.2012	02.01.2012
44	E	34788	Kalp Krizi	SSK	02.01.2012	04.01.2012
53	E	34798	Grip	Özel Sigorta	03.01.2012	03.01.2012
67	K	34796	Depresyon	Güvence Yok	12.01.2012	20.01.2012
14	K	34756	Kol Kırığı	Özel Sigorta	14.01.2012	14.01.2012
89	K	34888	Göğüs	SSK	21.01.2012	21.01.2012

			Kanseri			
71	E	34568	Beyin Tümörü	SSK	27.01.2012	29.01.2012

Tablo-22: Acil Servis Kayıtları

Bu tabloda ad-soyad, T.C kimlik numarası, telefon numarası gibi direk betimleyiciler çıkarılarak anonimleştirilmiş bir veri kümesidir. Böyle bir verinin çeşitli kapasite hesaplamaları yapılabilmesi için planlama birimleriyle paylaşıldığını varsayalım. Böyle bir durumda veri kümesi hastane içinde paylaşılacağı için daha detaylı anonimleştirme tekniklerine başvurulmamış olarak kabul ediyoruz. İlgili planlama birimlerinde çalışan personelden biri, aynı zamanda yaşadığı sitenin yönetici olsun ve bu görevinden dolayı kişisel imkanları ile elde ettiği aşağıdaki gibi bir veri kümesine daha sahip olsun.

Ad-Soyad	Yaş	Cinsiyet	Kapı No	Posta Kodu	Blok No
Aliye Özer	23	K	0012	34798	A
Mehmet Aras	44	E	0023	34788	C
Gülnur Sal	67	K	0098	34796	D
Selin Gürdal	14	K	0087	34756	A
Özlem Temel	89	K	0089	34888	B

Tablo-23: Bir sitede yaşayan kişilerin adres kayıtları

Burada görüleceği üzere her iki veri kümesindeki ortak değişkenler olan yaş,cinsiyet ve posta kodu bilgileri birleştirildiğinde ve iki tablo yan yana eklendiğinde aşağıdaki gibi yeni bir tablo elde edilerek ilk veri kümesinin anonimliği bozulmaktadır.

Ad-Soyad	Yaş	C	Posta Kodu	Kapı	Blok	Teşhis	Sigorta	Hastane Giriş	Hastane Çıkış
Aliye Özer	23	K	34798	0012	A	Grip	SSK	01.01.2012	02.01.2012
Mehmet Aras	44	E	34788	0023	C	Kalp Krizi	SSK	02.01.2012	04.01.2012
Gülnur Sal	67	K	34796	0098	D	Depresyon	Güvence Yok	12.01.2012	20.01.2012
Selin Gürdal	14	K	34756	0087	A	Kol Kırığı	Özel Sigorta	14.01.2012	14.01.2012
Özlem Temel	89	K	34888	0089	B	Göğüs Kanseri	SSK	21.01.2012	21.01.2012

Tablo-24: Birleştirilerek anonimliği bozulmuş veri kümesi

Görüldüğü üzere iki farklı verinin ortak değişkenlerinin bir araya getirilmesiyle elde edilen yeni veri kümesi kişilsere dair oldukça hassas verileri içeren riskli bir küme haline dönüşmektedir. Sitenin yönetici olan kişinin sitede ikamet eden Gülnur Sal'ın depresyon şikayetinin olduğunu, bu sebeple hastanede 8 gün kadar kaldığını ve sosyal sigorta güvencesinin olmadığını öğrenmesi Gülnur Sal'ın özel alan gizliliğinin ihlal edildiğinin ve hassas verilerinin kontrolsüzce yayıldığının en çarpıcı örneğidir.

Yukarıdaki örnek en basit haliyle dış verinin anonimleştirilmiş veri kümeleri için ne kadar büyük risk teşkil edeceğini göstermektedir. Takip eden bölümde ne gibi motivasyonlarla ve hangi şekillerde anonimleştirilmiş verinin anonimliğinin bozulabileceği ve kişisel bilgilerin açığa çıkacağı incelenecektir.

B. Art Niyetli Kullanıcılardan Gelen Saldırılar

Art niyetli kullanıcılar "belli bir kişiye ait bilgileri öğrenmek[195]" veya sadece "veri sahibini veya toplayıcısını utandırmak[196]" gibi motivasyonlarla bile

[195] Anco Hundepool, Aad van de Wetering, Ramya Ramaswamy, Luisa Franconi, Silvia Polettini, Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo, Vicenc Torra, Ruth Brand, Sarah Giessing, s. 11

anonimleştirilmiş verilerden kişisel veri elde etmeye çalışabilirler. Art niyetli kullanıcılara dair en büyük risk ne tip bir dış veriye sahip olduklarının veya olabileceklerinin bilinememesidir. Bu kullanıcılar bir veri kümesi üzerinde çalışarak, tüm veri kümesindeki anonimliğin bozulmasını sağlayabilecekleri gibi, belli bir kişiye dair sahip oldukları bilgiye istinaden veri kümesinde hedef odaklı aramalar yaparak o kişiye dair ek bilgi edinebilirler. Bu noktada tam bir kişisel veri ifşasına bile gerek yoktur, spesifik bir kişiye dair tahmine dayalı ek bilgi edinmek bile ihlale ve kişiye zarara yol açabilirler. Yapılan tahminin doğru olup olmaması bile zarar getirmeyeceği anlamına gelmez. Teknik açıdan iki büyük tehdit dikkate alınmaktadır; erişim yetkisi olmadan bir veri tabanından veri hırsızlığı ve erişim yetkisi bulunan veri tabanlarından yetkiyi kötüye kullanarak elde edilen veriler[197].

İlk durumda anonimleştirildikten sonra yayınlanmış ve hassas veriler içeren bir veri kümesindeki kişisel verileri ifşa edebilmek için erişim yetkilerinin olmadığı farklı başka kaynaklardan çaldıkları verileri kullanarak anonimliği bozabilirler. Bu tip siber saldırılar bankalar, teknoloji şirketleri, operatörler gibi kritik bilgilere sahip şirketlerin veri tabanlarına yapılabileceği gibi kişilerin özel mail adresleri veya bilgisayarlarına doğru da gerçekleşebilir.

Diğer durum ise, yukarıdaki hastane planlama servisinde çalışan ve aynı zamanda ikamet ettiği sitenin yöneticiliğini yürüten bir kişinin erişim hakının olduğu verileri kötüye kullanması ile gerçekleşmektedir. Kişiler erişebildikleri veri kümelerini kişisel bilgisayarları yardımıyla birleştirebilir, anonimliğin bozulmasını sağladıktan sonra elde ettikleri sonuç veri kümesini yayabilir veya satabilierler.

Bütün bunlara ek olarak art niyetli kullanıcının dış veri kümesi elde etmeden sadece kişisel ilişkilerine bağlı olarak öğrendikleri bilgilere dayanarak annimleştirilmiş veri kümelerinde yapacakları tahminler de risk yaratabilmektedir.

---

[196] Anco Hundepool, Aad van de Wetering, Ramya Ramaswamy, Luisa Franconi, Silvia Polettini, Alessandra Capobianchi, Peter-Paul de Wolf, Josep Domingo, Vicenc Torra, Ruth Brand, Sarah Giessing, s. 11

[197] Seda Gürses, Carmela Troncoso, Claudia Diaz, s. 9

Böylesi bir durumda art niyetli kullanıcı, yakın çevresi, akrabaları, komşuları ile ilgili sahip olduğu bilgilere dayanarak veya kişilerin Internet'te özel hayatlarına ve kişisel bilgilerine dair detayları ifşa ettikleri mecralardan edindiği bilgilerle anonimleştirilmiş bir veri kümesinde isabetli tahminler yaparak kişilere dair ek bilgiler edinebilir. Bu durumda edindiği her ek bilgi yeni bir anonimleştirilmiş veri kümesi için risk teşkil edecektir.

C. Araştırmacılar

İlk bölümde açıkladığımız üzere büyük veri sosyal, ekonomik, politik pek çok boyutta araştırma geliştirme çalışmalarının en kritik ham maddesi haline dönüşmüştür. Bu ortamda, araştırmacılar çalıştıkları konuyu ilgilendiren farklı kaynaklardan elde edilmiş veri kümelerini analiz etmek durumundadırlar. Bu durum da araştrmalar esnasında anonimleştirilmiş veriler kullanılıyor olsa bile farklı kaynaklardan veri kümelerine sahip araştırmacıların kolaylıkla kişisel verileri elde edebileceği gerçeğini doğurur.

Bunun yanı sıra anonimleştirme süreçlerine ve anonimliğin ölçeklenmesine istinaden yürütülen çalışmalar ise direk olarak anonimliğin hangi şartlar altında bozulabileceğini tespit etmeye çalışacaklarından burada amaç kişisel verilere ulaşabilmek olacaktır. Risk yönetimi ve anonimleştirmenin açıklarının tespit edilebilmesi veya yeni ve daha güvenli anonimleştirme metotlarının geliştirilebilmesi için büyük önem arz eden bu çalışmaların direk olarak hedefi anonimliği bozabilmektir. Böylesi bir araştırmanın konusu olan anonimleştirilmiş veri ise kasıtlı olarak araştırmacının kendi imkanlarıyla elde edebileceği dış veri sayesinde sınanacaktır.

D. Zaman İçinde Anonimliğin Bozulması

Anonimleştirme süreçlerine istinaden bir büyük yanılgı da veri kümesinin bir kere anonimleştirilerek güvenli hale gelmesiyle ihlallerin engelleneceği algısıdır. Günümüzün enformasyonel toplumunda veri kümeleri dinamik olarak değişmektedir. Diğer taraftan her yeni teknolojik gelişim ve yenilik ile farklı türlerde veriler üreyebilmektedir. Bunun en güzel örneği mobil uygulamalardır. Büyük bir hızla çeşitlenen mobil uygulamalar iş modellerine göre kişilere ait farklı verileri üretebilmektedir. Bu durumda anonimleştirilmiş verinin güvenilirliği bu dinamik ortamda ilk tasarlandığı andaki güvenilirliğini koruyamaz. Zamana bağlı değişiklikleri üç temel gruba ayırabiliriz.

İlk olarak, aynı veri kümesinin zaman içinde kayıtlarındaki eklemeler, çıkarmalar veya değişkenlerin değerlerindeki güncellemeler anonimleştirilmiş veri kümesindeki k-anonimlik, l-çeşitlilik veya t-yakınlık metotlarıyla sağlanan güvenliği bozabilmektedir.

İkinci olarak, anonimlik aynı veri kümesinin ihtiyaca göre farklı zamanlarda iki ayrı alt kümesinin ayrı ayrı anonimleştirilerek ifşa edilmesi sonucunda bozulabilmektedir. Anonimleştirilmiş alt veri kümeleri orijinal veri kümesinin farklı değişkenlerine ait olabilirler ve anonimleştirme metotları yeterli güvenilirliğe sahip olmadığından birleşmeleri halinde orijinal veri kümesini yeniden meydana getirebilirler. Yukarıda çalıştığımız anonimleştirme metotlarında incelediğimiz dizilimlerin uyuşması gibi açıklar iki veri kümesine bakarak anlamlı tahminler yapılabilmesine olanak vermektedir.

Bütün bunların yanında büyük verinin çeşitliliğe ve yeni kurgulara açık yapısı zaman içinde ne tür verilere ne ölçüde erişilebileceğini tahmin etmeyi zorlaştırmaktadır. Yeni iletişim uygulamaları, yeni iş modelleri, Internet'in sınır tanımayan altyapısı ile kurgulanan yeni tasarımlar sonuçta hep yeni veri kümelerine dönüşmekte ve büyük veriye katkı sağlamaktadır. Diğer yönüyle, bütün bu teknolojik gelişmeler kişilerin sisteme daha fazla dahil olmaya ve kendi

rızalarıyla daha fazla kişisel bilgiyi açık erişim ortamlarında yayınlamaya teşvik etmektedir. Bu hızlı gelişim, zaman içinde ne tip verilerin daha rahat erişilebilir olabileceğini ölçmemizi engeller. Böylelikle, belli bir t anında gerekli tüm tedbirleri alınarak anonimleştirilmiş güvenli bir veri kümesi, zaman içindeki gelişmelerin sağladığı yeni veri kümeleri ile ilk baştaki güvenilirliğini kaybedebilir.

II. Güvenilirlik Tartışmaları

Büyük verinin sağladığı fayda ile anonimleştirilmiş verinin güvenlik açıklarının yarattığı risk farklı araştırmacıların olaylara farklı açılardan bakmasına sebep olmuştur. Bazı araştırmacılar büyük veriden sağlanan faydayı öne çıkartırken, diğerleri anonimliğin bozulduğu ihtimallere odaklanmış ve anonimleştirme süreçlerine dair güvenin sorgulanması gerektiğini ifade etmişlerdir. Bu bölümde, anonimleştirmenin riskleri ve faydaları kapsamında yürütülen tartışmalara istinaden mevcut genel durumu yansıtmak için dört temel çalışma incelenecektir. Ayrıca bu çalışmalar, tartışmayı ele aldıkları açıdan iki ana başlık altında sunulacaktır. İlk üç çalışma konuya teknik ve güvenlik açısından yaklaşmakta, son çalışma ise konuyu içerik bağlamında incelemektedir.

Bu çalışmalardan ilki Kolorado Üniversitesi doçenti Paul Ohm'un anonimleştirmeye olan güvene dair "uluslar arası bir tartışmayı alevlendiren[198]" makalesi "Gizliliğin Çiğnenmiş Vaatleri: Anonimleştirmenin Sürpriz Başarısızlığına Cevap[199]" isimli makalesidir. Bu makalede Ohm (2009), anonimleştirilmiş verilerin kolaylıkla anonimliklerinin bozulduğu ve kişisel verilere ulaşılabildiği yaşanmış olaylar üzerinden bir tartışma yürütür ve anonimleştirmenin bu sürpriz başarısızlığına bir cevap verilmesi gerektiğini

[198] Paul Ohm, Associate Profesor of Law and Telecommunications, University of Colorado, Biography, bkz. http://paulohm.com/

[199] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, UCLA Law Review,Vol 57, 2010,

savunur. İkinci çalışma ise, Paul Ohm'un çalışmasından iki yıl sonra Arizona Üniversitesi doçenti Jane Yakowitz, Ohm'un çalışmasına cevap niteliği taşıyan ve veriden sağlanan faydanın sosyal etkilerinin "yanlış anlaşıldığının ve ciddi şekilde hafife alındığının" altını çizerek, Ohm ve benzer görüşte olan araştırmacıların "bilişim literatürünü yanlış yorumladıklarını" vurgular[200]. Üçüncü çalışma ise anonimliği ölçmeye çalışarak, farklı veri kümelerinde farklı risklerin varlığını göstermeyi hedefleyen Matthijs R. Koot'un "Anonimliği Ölçme ve Tahmin Etme" isimli kitabıdır. Bu üç çalışma da, anonimleştirmenin güvenilirliğini, metotların teknik incelemelerine ve risk hesaplamalarına bağlı kalarak değerlendirmiş ve konuyu güvenlik ve tehdit açısından ele almıştır. Son olarak inceleyeceğimiz çalışmada ise, 2013 yılında Felix W. Tu, "Veri Kümelerinde Gizlilik ve Fayda[201]" isimli makalesinde, Yakowitz ve Ohm'un çalışmalarındaki düşünceleri sentezleyerek ve yer yer her ikisinin de varsayımlarını eleştirerek süreci analiz eder. Bu çalışmayla Wu, fayda ve gizlilik içeriğine odaklanır ve önce bu iki kavramın tanımlarının netleştirilmesinin gerektiğini vurgular. Bu haliyle bu çalışma diğerlerinden bakış açısı olarak ayrılmaktadır.

Dört çalışma da anonimleştirmenin fayda ve risk tartışmasında literatüre ve olay bazlı araştırmalara oldukça hakim olduklarından ilerleyen bölümlerde bu üç çalışmanın detayları incelenerek anonimleştirme süreçlerinin risk ve fayda tartışmasına ait genel çerçeve çizilecektir.

A. Teknik ve Güvenlik Tartışmaları

1. Gizliliğin Çiğnenmiş Vaatleri

[200] Jane Yakowitz, *Tragedy of Data Commons*, Harvard Journal of Law and Technology, Vol.25, 2011, s. 4

[201] Felix T. Wu, *Defining Privacy and Utility in Data Sets*, University of Colorado Law Review 1117 (2013)

Ohm, makalesinde teknoloji uzmanlarının veride küçük değişiklikler yaparak kişilerin gizliliklerini koruyabileceklerini düşündükleri yapıya "kuvvetli anonimleştirme varsayımları[202]" adını vermektedir ve makalenin ilerleyen bölümlerinde bu terminoloji etkisinde gelişen hukuksal düzenlemeleri ve çelişen gerçek hayat örneklerini incelemektedir. Anonimleştirmenin gerekçelerini incelediği bölümde, şirket içi paylaşımların gerekliliği, hukuksal yükümlülüklere istinaden resmi mercilerle yapılan paylaşımlar, veriden daha fazla fayda sağlanabilmesi için yapılan akademik amaçlı kamu ifşaları gibi süreçleri kabul etmekte ve hem gizliliğin sağlanabilmesi hem de bu süreçlerin sürekliliği için anonimleştirmenin özel sektör ve kamu tarafından benimsendiğini belirtmektedir. Bu noktada, Lessig'in, dört temel düzenleyicisi olan normların, pazarın, mimarinin ve kanunların anonimleştirmeyi öne sürdüğü iddia edilmektedir. Ancak Ohm çalışmasının çatısını "veri ya faydalıdır ya da kusursuzca anonimleştirilmiştir ama aslı ikisi aynı anda olamaz[203]" tezine dayandırmaktadır ve pek çok hukuk akademisyeninin anonimleştirmeye olan derin inancının yanında son yıllarda anonimleştirilmiş veriye yapılan saldırıların başarılarının bilgisayar bilimcilerini bile hayrete düşürdüğünü vurgulamaktadır.

Çalışmaya göre Ohm, anonimleştirme tekniklerinin yay-ve-unut modelinde çalıştıklarını, sistem yöneticilerinin anonimleştirme tekniklerini uyguladıktan sonra veriyi ister kamuya ifşa etsinler ister özel olarak üçüncü partilerle paylaşsınlar isterlerse de şirket içi birimlerle paylaşsınlar, veriyi yaydıktan sonra unuttuklarını ve kayıtlara ne olduğunu takip edecek bir girişimde bulunmadıklarını iddia etmektedir.

Çalışmada gerçek örnekler içinden anonimleştirilmiş verinin güvenilirliğine dair en ciddi tartışmaları yaratan üç temel saldırı incelenmiştir: AOL Veri İfşası, Massachusetts Grup Sigorta Komisyonu İfşası, Netflix Yarışması. Bu üç temel saldırı mevcut literatürde önemli bir yere sahip olduklarından ve tartışmaların temelini oluşturduklarından çalışmamızda da detaylı olarak incelenecektir.

---

[202] Paul Ohm, s. 1706
[203] Paul Ohm, s. 1704

a) AOL İfşası

AOL, 1998 yılında 12 milyon, 2006 yılında 27 milyon gibi abone sayılarına ulaşabilmiş Amerika'da hizmet veren büyük bir servis sağlayıcıdır[204]. Şirket, 2006 yılında "AOL Research" adıyla yeni bir girişimde bulunarak, AOL arama motorlarındaki 650.000 kullanıcıya ait olan 20 milyon arama sorgu kaydını sitelerinde kamuya ifşa ederek araştırmacıların dikkatine sunmuşlardır. Arama sorguları ifşa edilmeden önce anonimleştirilerek kimlik saptaması yapılabilecek kişisel verilerden arındırılmış ve bunun yerine kullanıcılara numaralar atanmıştır[205]. Ancak kısa zaman içinde araştırmacılar, arama sorguları içindeki ifadeleri takip ederek ve aynı kullanıcı numarasına ait birden fazla sorguyu birleştirdiklerinde birebir kimlik saptaması yapılabildiğini görmüşlerdir. Bu durum önceki bölümlerde çalıştığımız, birden fazla anonimleştirilmiş veri kümesinin birleşiminden ve veri kümelerindeki dolaylı betimleyicilerin kombinasyonlarından orijinal kümenin açığa çıktığı modele güzel bir örnektir.

İfşa edilen sorgu veri kümesinde 4417749 kullanıcı numarası ile yer alan kişi "Lilburn, Ga'daki bahçe düzenleyicileri", "Gwinnet County Georgia'da satılık göl kenarı parsel", ve pek çok "Arnold" soyadlı kişiye ait aramalar yapmıştır. Bu üç veri takip edilip Internet üzerinde arama yapıldığında 62 yaşında Lilburn, Georgia'da yaşayan Thelma Arnold isimli kişinin kimliği kolaylıkla saptanmıştır[206]. Thelma Arnold bunun gibi "hissis parmaklar", "60 bekar adam" , "her yere işeyen köpekler" gibi özel hayatıyla ilgili pek çok hassas detayı açık eden ve toplumda utanç verici bir konuma düşmesine sebep olacak aramalar da yapmıştır. Kimliği ifşa edilmiş kayıtlar içinde teşhis edildikten sonra kişiye dair

---

[204] Lawrance Lessig, s. 88

[205]Michael Barbaro, Tom Zeller, A *Face is Exposed for AOL Searcher No. 4417749*, New York Times, bkz. http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0

[206] Michael Barbaro, Tom Zeller, A *Face is Exposed for AOL Searcher No. 4417749*, New York Times, bkz. http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0

bu ve benzeri pek çok hassas veri açığa çıkmış ve kişinin özel alan gizliliği ve kişisel verileri ihlal edilmiştir.

b) Massachusetts Grup Sigorta Komisyonu İfşası

Massachusetts'de 1990'lı yıllarda Grup Sigorta Komisyonu isimli bir sigorta şirketi bölgedeki kamu personelinin sağlık sigortası süreçlerini üstlenerek, talep eden araştırmacılara ücretsiz olarak işçilerin hastane ziyaretlerine ait olan veriyi anonimleştirerek paylaşabileceğini duyurmuştur. Paylaşımlar öncesinde Grup, isim, adres, sosyal güvenlik numarası gibi direk betimleyicileri veriden çıkartarak güvenli ve anonimleştirilmiş bir veri kümesi yaratmayı hedeflenmiştir. Latanya Sweeney isimli araştırmacı sigorta grubundan bu veriyi talep eder, sonrasında ise Massachusetts eyaletinde yer alan Cambridge şehrinin belediyesinden 20 dolar karşılığında tüm seçmen kayıtlarını satın alır. Bu iki veri kümesinde posta kodu, doğum tarihi ve cinsiyet değişkenleri ortaktır. Bu üç değişken üzerinden iki veri kümesi birbiriyle eşleştirildiğinde kişilerin kimliklerinin kolaylıkla tahmin edilebileceği kayıtlar yaratılmış olur. Örneğin, Massachusetts eyalet valisi William Weld o günlerde Cambridge'de oturmaktadır ve kamu personeli olduğundan kayıtlarının sigorta grubunun sağlık verileri içinde yer aldığı bilinmektedir. Sweeney'in eşleştirmesinden sonra ortaya çıkan veri kümesinde sadece 6 kişi vali ile aynı doğum tarihini paylaşmaktadır, bunlardan sadece 3'ü erkektir ve sadece biri vali gibi 5 rakamlı posta koduna sahiptir[207]. Sweeney araştırmanın önemini vurgulamak adına kayıtlar içinde kimliğini saptayabildiği valinin teşhis ve tedavi detaylarını da içeren sağlık kayıtlarını ofisine postalamıştır.

[207] Latanya Sweeney, s. 3

c) Netflix Yarışması

2006 yılında Netflix isimli film kiralama şirketi Cinematch adını verdikleri film tavsiye algoritmasından daha gelişmişinin tasarlanması amacıyla bir yarışma başlattı ve anonimleştirilmiş puanlama verisi ve hesaplanan puanlama ile gerçek puanlama arasındaki yakınlığı ölçmeye yarayan bir tahmin hassasiyet çizelgesini kamuya ifşa etti[208]. Yarışmaya olan ilgiyi arttırmak için kazanan için 1 milyon dolarlık bir ödül tayin edildi. Filmlere ve puanlamalara ait veriler ifşa edilmeden önce anonimleştirilmişti ve ifşa edilen veri kümesi yalnızca puanlanan filmin adı, puanı ve puanlama zamanı bilgilerini içermekteydi. İfşa edilen veri kümesi 480.189 üye tarafından 1999 ve 2005 yılları arasında oluşturulmuş 100.480.507 adet puan bilgisini içermekteydi[209]. Yarışmanın gerçek amacı en iyi tahminleri üreten algoritmayı kurgulamak iken bazı araştırmacılar bu durumu anonimleştirme üzerine yürütülen araştırmalara bir girdi olarak değerlendirdi. Teksas Üniversitesi'nin iki akademisyeni Narayanan ve Shmatikov, ifşa edilen bu anonimleştirilmiş veri kümesi üzerinden anonimliğin bozulup bozulamayacağını ve kişilerin hangi filmleri izleyebildiğinin bu veri kümesi sayesinde tahmin edilip edilmeyeceğini test etmek için harekete geçtiler. Her iki araştırmacı da, film puanlamalarının sağlık ve finansal kayıtlar kadar hassas kişisel veriler içeridiğini, kişilerin siyasi görüşlerinden cinsel tercihlerine kadar pek çok bilgiyi barındırdığı kanaatindeydi[210]. Bu araştırmada amaç bir önceki eşleştirme örneklerinden farklı olarak, belli bir üyeye ait bazı bilgilere sahip olan bir kullanıcının Netflix veri kümesine bakarak ilgili üye hakkında ne kadar bilgi edinebileceğinin ölçülmesi üzerine kuruluydu. Burada veri kümesini inceleyen kişinin, tanıdığı ve sinema tercihleri hakkında bilgiye sahip olduğu bir Netflix üyesiyle ilgili ne kadar isabetli tahminler yapabileceği ölçülmüştür. Bir kişinin sinema tercihlerini, film puanlamalarını sıradan günlük konuşmalar esnasında bile öğrenebilmek oldukça kolay olacağından, bir Netflix üyesi hakkında puanlanan film ve puanlama zamanı

[208] The Netflix Prize Rules, bkz. http://www.netflixprize.com/rules
[209] Arvind Narayanan, Vitaly Shmatikov, *How to Break Anonymity of the Netflix Prize Dataset*, The Universtiy of Texas, 2008, s. 1
[210] Arvind Narayanan, Vitaly Shmatikov, s. 3

gibi bilgilere erişebilmenin kolaylığı araştırmada hesaba katılmaktaydı. Bu sebeple araştırmacılar, çalışmanın savını şu soru üzerine kurguladılar: "Bir Netflix üyesinin tüm kayıtlarını teşhis etmeyi ve bu üyenin tüm film puanlama tarihçesine ulaşmayı hedefleyen bir saldırganın, bu üyeyle ilgili ne kadar çok şey bilmesi gerekmektedir[211]?" Narayanan ve Shmatikov, yaptıkları istatistiksel hesaplamalar sonucunda tüm veri kümesinin %96'sının 8 adet filme ait puanları kesin olarak bilen (2 tanesinin yanlış olduğu kabul edilerek) ve 3 günlük bir sapmayla puanlama zamanını da öğrenmiş olan bir saldırgan tarafından teşhis edilebileceğini görmüşlerdir. Diğer taraftan, veri kümesinin %64'ü sadece 2 adet puan ve film eşleşmesini bilerek teşhis edilebilmektedir. Eğer sahip olunan dış veri en çok puanlanmış ilk 100 film içinde değilse saptamalar daha isabetli olmaktadır. Örneğin puanlama zamanını 14 günlük sapmayla, 8 adet filme ait puanı bilen (2 tanesinin yanlış olduğu kabul edilerek) kişi ver kümesindeki %80 kişiyi tamamen teşhis edebilmektedir.

Bu araştırmanın başarısı, dış verinin önemini bir kez daha vurgulamaktadır. Anonimleştirilerek kişisel verilerin güvenliğinin sağlandığını varsaymak görüldüğü üzere yeterli olmamaktadır. Ohm, bu araştırma sonucunda ortaya çıkan verilere istinaden 2009 yılında bazı Netflix müşterilerinin şirketi dava ettiğini de belirtmiştir[212].

Her üç örnekten çıkartılacak bir diğer önemli sonuç ise anonimleştirmenin şirketleri yasal yükümlüklerini yerine getirmeleri konusunda akladığını, pratikte sağlanamayan kişisel veri korumasının ve özel alan gizliliğinin teorikte mevzuata uygun işlediğini göstermesidir. Kullanıcıların kimlikleri tespit ediliyor olabilmesine rağmen, ilk adımda her üç şirket anonimleştirmeyi sağlayarak bağlı oldukları mevzuata uyumlu hareket etmişlerdir.

[211] Arvind Narayanan, Vitaly Shmatikov, s. 3
[212] Paul Ohm, s. 1722

d) Yasal Mevzuatın Dönüşümü, Eksikler, Öneriler

Ohm, anonimleştirme süreçlerinin günümüz gizlilik kanunları üzerindeki etkisini eleştirirken, konunun 19. Yüzyılda kişinin uğradığı zarar konsepti üzerinden tanımlanırken, 1970'lerden sonra bilgi teknolojilerindeki gelişmelerle "tamamen veri merkezli ve kişisel veriler odaklı anlayışlara[213]" dönüştüğünü iddia eder. Zarar konseptinde "davacının inziva, yalnızlık veya özel ilişkiler hakları ihlal edilmesi, davacı hakkına utanç verici gizli gerçeklerin kamuya ifşası, davacının kamu nezdinde yanlış tanıtılması[214]" gibi unsurlar dikkate alınır. Diğer taraftan veri merkezli yaklaşımlarda veri kategorileri, ilişkilendirilme ve kimlik saptayabilme özellikleri önem kazanmaktadır. Bu bağlamda Ohm'un da vurguladığı gibi günümüzdeki hemen hemen tüm kişisel veri koruması ve özel alan gizliliği hakkındaki mevzuat, anonimleştirme süreçlerini gizliliği koruyan bir metot olarak kabul etmiştir. Ancak anonimleştirmenin bozulduğu durumlar bu gizlilik mevzualarındaki saklanmış dengesizliği ortaya çıkartmakta, bazı durumlarda süreci gizlilikten yana bazı durumlarda ise verinin serbest dolaşımından yana desteklemektedir[215]. Ek olarak yazar, 95/46/AT sayılı yönergeyi, tüm anonimleştirme süreçlerini hariç bıraktığı için aşırı genel bir kapsamda görmektedir. Yazara göre yönerge başarılı şekilde anonimliği bozulabilen her veri kümesini kapsayacağından sınırsız ve aşırı genel bir kapsama sahiptir.

Ohm, tartışmasının son bölümlerinde düzenleyici ve kanun yapıcıların tek gerekli çözümler olarak değerlendirdiği ama dengeyi sağlama konusunda yeterli olmayan üç temel anlayıştan bahsetmektedir. İlki, son kırk yılın önleyici tedbir anlayışlarını bırakarak yeniden zararın tazmin edilmesine yönelik anlayışların benimsenmesidir. Bu anlayışta eğer süreç yalnızca zarar gerçekleştikten sonra zararın tazmini boyutuyla ele alınırsa bu durum anonimliğin bozulması ile oluşabilecek gizlilik ve kişisel veri tehditlerinin benzersiz şekilde artıracağı

[213] Paul Ohm, s. 1734
[214] Paul Ohm, s. 1733
[215] Paul Ohm, s. 1740

anlamına gelir. Çünkü bozulan her anonimlik ve elde edilen her yeni bilgi bu bilginin başka bilgilerle eşleştirilmesi olasılığını arttırmakta ve bu durum daha geniş kitlelere zarar verecek sonuçlar doğrumaktadır. Bütün bu eşleşebilir veri kümelerinin birleşimi ise toplumun her üyesine dair bilgi içerebilecek büyüklükte en tepedeki büyük verinin hassasiyetlerini arttıracaktır. Günümüz dünyasında hemen hemen her kişi bir saldırgan tarafından şantaj, ayrımcılık, taciz, hırsızlık veya kimlik hırsızlığı amaçlarıyla kullanılabilecek bir veritabanındaki bir veriyle ilişkilendirilebilir[216]. Bu durumda bir veri kümesine yapılan ve anonimliği bozma amacı güden saldırıların sadece o veri kümesindekilere zarar vereceği algısıyla değil, büyük veriye ait tüm anonimliğin etkileneceği algısıyla hareket edilmelidir. Bu durum kimliğin tm olarak sağtanabildiği durumlar dışında kimliğe dair tahmin edilebilirliğin oranının düşürülmesi durumları için de geçerlidir. Ohm, anonimleştirilmiş verinin gerçek kimliklere dair çok yakın tahminler ürettiği veya hassas verilerin ortaya çıkmadığı saldırıların da endişe verici olduğunu vurgular. Sonuç olarak, sadece ardıl düzenlemeler yapılması halinde her anonimliği bozulan veri kümesi toplumun her bireyinin parçası olduğu büyük verinin güvenilirliğini etkileyecek ve zararın ölçeğini ve kitlesini arttıracaktır.

İkinci eksik tedbir olarak Ohm, düzenleyicilerin teknolojinin kurtarıcılığına aşırı güvenini eleştirmektedir. Ohm'a göre, anonimleştirme metotlarındaki yeni gelişmeler basit bir anonimleştirme işleminden daha karmaşık, daha yavaş ve daha maliyetli süreçler doğurmaktadır. Ayrıca, veri kümelerinin fayda ve gizlilik dengesinin, teorisyenlerin imkansızlık sonucu diye adlandırdıkları, anonimleştirilmiş veri ile eşleşmesi sağlanacak her zaman dış veri bulmak mümkündür anlayışı ile tam olarak sağlanamayacağını iddia eder. Ohm'a göre mükemmel şekilde anonimleştirilmiş veri kümeleri faydalı bilgi içermezken, gizlilik ihlallerine daha açık olan veri kümeleri de daha fazla bilgi sağlamaktadırlar. Gelişmiş anonimleştirme tekniklerinin bir diğer büyük engeli ise sadece kendi tasarlandıkları tarihten sonrası için geçerliliklerinin olmasıdır. Yani bundan önce çoktan ifşa edilmiş veri kümeleri için çözüm üretemezler. Ancak

[216] Paul Ohm, s. 1748

daha basit metotlarla anonimleştirilerek kamuya ifşa edilmiş bu veri kümeleri de her zaman daha gelişmiş tekniklerle anonimleştirilmiş veri kümeleri için tehdit oluşturmaktadırlar. Diğer taraftan teknik olarak üretilen çözümlerin yanında verilerin çok sınırlı şartlarla paylaşılması veya ifşa edilmesi gibi daha manuel çözümler de bilişim uzmanları tarafından denenmektedir. Örneğin, bir analistin bütün ham veriye ulaşmasını sağlamak yerine, ihtiyacı olan verileri yetkili bir sistem yöneticisinde öğrenmesi gibi süreçlerin uygulanması söz konusudur. Bu durumda analist ihtiyacı olan veri kümelerinin detay verilerini görmeden toplam değerlerini edinebilir. Ancak bu gibi çözümler de sınırlı bir bakış açısı geliştirme ve zaman kaybı gibi sorunlardan dolayı verimliliğini kaybetmektedir.

Üçüncü olarak anonimleştirilmiş verinin tersine çevirilerek anonimliğinin bozulması işleminin yasaklanması çözümü eleştirilmektedir. Böyle bir yasaklama anlayışının kontrolü ve denetimi sağlanamayacağı için başarısız olacağı baştan bellidir. Anonimliği bozacak işlemleri hangi şartlarda olursa olsun gizli bir şekilde gerçekleştirmek mümkün olacağından anlamlı bir tedbir olamamaktadır.

Bu üç gerçekçi olmayan ama düzenleyiciler tarafından benimsenme ihtimali olan zayıf tedbirlerin yanı sıra Ohm, makalesini daha geçerli çözüm önerileriyle sonlandırmaktadır. Ohm ilk olarak gizlilik yasalarının tasarlanırken şirket bazlı yaklaşımlar tasarlanmasını savunmaktadır. Ancak buradaki şirket bazlı anlayışın sektörel bir bakış açısıyla değil de şirketlerin işledikleri ve yönettikleri veri kümelerinin çeşitliliğine ve bağlantılarına bağlı olarak geliştirilmesi gerektiğini savunmaktadır. Google, Microsoft ve Yahoo gibi servis sağlayıcılar böyle bir sınıflandırmada aynı risk grubunda sayılabilirler. İkinci olarak Ohm, gizliliğe ve kişisel verilerin korunmasına dair düzenleyici prensiplerin teknik yaklaşımlar kadar sosyolojik, psikolojik ve kurumsal anlayışlarla da kurgulanması gerektiğini ileri sürmektedir. Bunlar şu gibi soruların sorgulanmasını gerektirmektedir; "sistem yöneticisi anonimliğin bozulması ihtimaline karşı hangi önlemleri aldı, kimler kişisel veri güvenliğini ve gizliliği ihlal etmeye kalkışabilir ve başarma ihtimalleri nelerdir, sektörün geçmişi, uygulamaları, gelenekleri ve yapısal

özellikleri gizlilik ihtimaline güven mi yoksa şüphe mi aşılamaktadır[217]". Son olarak Ohm, düzenliyicilerin ihlal riskini azaltacak faktörler belirleyerek bu faktörlere bağlı testleri şirket grupları bazlı uygulamaları gerektiğini savunur. Çalışmada riskin azaltılması konusunda beş temel faktörün dikkate alınması gerektiği savunulmaktadır: veri işleme teknikleri, kamu ifşası veya özle gruplara yapılan ifşalar, veri niceliği, motivasyon, güven. Veri işleme tekniklerine göre anonimliğin bozulması ihtimalinin dercelendirmesi gibi yöntemlerle risk hesaplamaları yapılabilir. Aynı şekilde kamuya ifşa yerine, gerekli görülmeyen hallerde sadece güvenli taraflar arasında veri paylaşım ve ifşası bir diğer risk azaltan faktördür. Önceki bölümlerde detayları verilen anonimleştirmenin bozulduğu tüm büyük örnekler kamu ifşaları örnekleridir. Mevcut mevzuatlar hep hassaslık, ilişkilendirilebilirlik gibi veri nicelikleriyle ilgilenmektedir, ancak hiçbir yasal düzenleme bir sistem yöneticinin ne büyüklükte veri toplayabileceğini düzenlememektedir. Bu bağlamda veri niceliği de risk yönetimi anlamında düzenlenebilir bir faktördür. Son olarak verileri kullanan partilerin motivasyon ve güvenilirliği risk hesaplamalarını etkilemektedir. Verinin hangi amaçlarla analiz edileceği analistin motivasyonunu ve analiste duyulan güveni ortaya koymaktadır. Düzenleyicilerin bu iki faktörü hesaba katması özellikle finansal amaçlarla veri analizi yapan kuruluşlara daha ihtiyatlı yaklaşması, bunun yanında akademisyenlere ve araştırmacılara daha çok imkan vermesi risk yönetiminde etkili olacaktır.

Yukarıdaki bölümde özetlediğimiz Paul Ohm'a ait olan çalışma anonimleştirilmiş verinin güvenlik açıklarını kapsamlı şekilde yansıtmakta ve temellerini metotların teknik eksikliklerinden almaktadır. Bu görüşe göre, anonimliğin dış veri unsuru ile kolaylıkla bozulabiliyor olması karşılaşılan yeni bir durumdur ve anonimleştirme süreçlerine güvenen yasal mevzuatları da riskli bir konuma itmektedir. Yaşanan örneklerde olduğu gibi basit yöntemlerle kimlik saptamaları yapılabilmesi ve kişilere ait hassas verilerin açığa çıkması büyük risk teşkil etmektedir. Ohm, çalışmasında düzenleyicilere çözüm önerilere sunarken, bu

[217] Paul Ohm, s.1761

teknolojik değişime hızlıca cevap vermelerini ve risk değerlendirmeleri ile karşılıklı olarak dengeyi sağlamaları gerektiğini hatırlatmaktadır.

Anonimleştirilmiş verilerin yüksek hacimli, çok çeşitli ve hızla üreyen büyük veri karşısında kolaylıkla güvenilirliğinin ortadan kalkması tartışmanın bu tarafındaki düşünürlere göre anonimleştirme süreçlerinin sonunu getirmiştir.

2. Müşterek Veri ve Abartılan Riskler

Jane Yakowitz, Paul Ohm'un tartışma yaratan ünlü çalışmasından iki yıl sonra, bu çalışmayı eleştirir nitelikte olan kendi çalışmasını yayınlamıştır. Buna göre Yakowitz, üç temel kurguyla ilerler; öncelikle araştırma verisinin faydasını yaşanmış örnekler üzerinden analiz ederek bu adımdaki toplumsal faydayı vurgular, daha sonra anonimleştirme süreçlerine dair güveni zedeleyen bilişim literatürünün yanlış yorumlandığını iddia eder, sonuçta ise gerçekçi riskleri ele alarak önerilerini sunar. Bu haliyle Yakowitz'in çalışması, anonimleştirmenin kişisel veri ve özel alan gizliliğini koruma amacıyla bir çözüm olamayacağını savunan görüşe eleştiri niteliğindedir ve anonimliğin bozulmasına istinaden riskleri değerlendirirken daha iyimser bir tablo çizmektedir. Yakowitz, çalışmasını araştırma verisinin önemi üzerine kurmaktadır ve anonimleştirilmiş araştırma verisine "müşterek veri" ifadesiyle tanımlamaktadır. Yakowitz'e göre müşterek veri olmadan literatüre kıymetli katkılar sağlayan bu ve benzeri hiçbir araştırma gerçekleştirilemez. Ancak son yıllardaki anonimliğin kolaylıkla bozulduğunu ispatlayan çalışmalar ve yaşanmış olaylar anonimleştirmenin pratikte mümkün olmadığı inancını geliştirmiş ve kişilerin bireysel olarak kendilerine ait kayıtların veri kümelerinden silinmesini talep ettikleri bir algı yaratmıştır. Ancak Yakowitz, kayıtlarının veri kümelerinde yer almasını engelleyecek taleplerin kişilerin kendi güvenliklerini arttırdığını kabul ederken, müşterek veriden sağlanan kolektif faydanın niteliğini yitireceğini savunmaktadır. Bu nitelik kaybı da büyük bir

risktir ve Yakowitz'e göre "artık araştırma verisinin büyük bir korumaya ihtiyacı vardır[218]".

a) Araştırma Verisi ve Kolektif Fayda

Araştırma geliştirme projeleri ve akademik çalışmaları yönlendiren veri kümeleri ile büyük veri, çalışmamızın ilk bölümünde de detaylı olarak incelenmişti. Aynı şekilde, Yakowitz de kendi çalışmasında bu konuda kapsamlı örneklerle "müşterek veri" olarak nitelendirdiği anonimleştirilmiş araştırma verisinin önemini vurgulamaktadır. Makalede öncelikle araştırma kavramının tanımı yapılır. Araştırma, "beşeri bilgiye doğrulanabilir ve genelleştirilebilir sonuçlar katma amacıyla yapılan metodolojik çalışmadır[219]" ve "örnek havuzda belli kişilerin hareketlerini anlamak amacıyla yapılan analitik çalışmaları hariç bırakır[220]". Bu tanımlamayla Yakowitz, araştırmacıların, kişilerin verileri ile ilgilenmediklerini, bu verilere "kim?" sorusu yerine "kaç kişi" veya "hangi oranda" sorularına yanıt verebilmek için ihtiyaç duyduklarını vurgulamaktadır. Yakowitz'in müşterek veriye dayalı önemli çalışmalara istinaden örnekleri oldukça çarpıcıdır. 1997 yılında Amerika'da yapılan araştırmalara göre uyuşturucu mahkumlarının aldıkları ceza sürelerinin maliyet etkinliğinin düşük olduğunu, tedavi progamlarına harcanan maliyetler ile yasal rejime harcanan maliyetlerden daha etkili olduğunu göstermiş ve düşük dereceli uyuşturucu kullanımına istinaden ceza hükümlerinde değişikliklere gidilmesini sağlamıştır[221]. Bunların yanında Yakowitz, müşterek veri sayesinde nüfus verilerinin iskan alanlarındaki ırksal ayrımlarını yorumlama, doğum kayıtlarının sigara kullanımının cenin üzerindeki etkilerini hesaplama, sabıka verilerinin semtlerin sosyo-ekonomik durumuna bağlı olarak polis kaynaklarını kullanmadaki

[218] Jane Yakowitz, s. 4
[219] Jane Yakowitz, s. 6
[220] Jane Yakowitz, s. 6
[221] Jane Yakowitz, s. 9

eşitsizliğini inceleme gibi kritik toplumsal çalışmalara imkan verdiğini vurgulamaktadır.

Yakowitz, müşterek verinin faydalarını ve kullanılma yöntemini örneklerken, ihtiyaç duyulandan daha fazla verinin saklanmaması ve veri kümelerine açık erişim yerine sınırlı gruplar içinde paylaşımını savunan görüşlere de eleştiriler getirmektedir. Öncelikle, alakasız amaçlarla toplanan verilerin bazen en faydalı veri kümesi haline geldiğini ve hangi veri kaynağının en iyi araştırma sonucunu ve topluma en faydalı katkıyı sağlayacağını kestirmenin mümkün olmadığını savunur. Bu sürece en güzel örnek ise Google'ın Grip Trendleri haritasıdır. Google tamamen farklı amaçlarla sakladığı arama sorguları ve IP adresleri verisinden Grip Trendleri haritası üretebilmiş ve bugün coğrafi olarak dünyadaki grip salgınlarının görüntülenebildiği bir ürün tasarlamıştır. İkinci olarak, veriye kısıtlı gruplarla paylaşmanın araştırmacılara aşırı güvenen bir anlayış yarattığını ve bu anlayışın, masum hataları yakalamayı ve büyük fayda sağlayacak veri kümelerine erişimi engellediğini iddia eder. Örnek olarak, 1970'lerde Isaac Ehrlich isimli araştırmacı, ölüm cezalarının caydırıcılığı üzerine yaptığı çalışmada verilen her ölüm cezasının sonrasında işlenecek 8 cinayeti engellediğini iddia etmiş ve o dönemdeki davaları bile etkilemiştir. Ancak daha sonrasında Ehrlich'in araştırma sonuçlarını inceleyen ekononomistler caydırıcılık etkisinin gözlemlenen döneme ve araştırmacının seçeceği başka keyfi kararlara göre değişebileceğini ispatlamıştır. Yakowitz, Ehrlich'in çalışmasının yaygın bir kitleyle paylaşılmasının, sonrasındaki çalışmaları da tetiklediğini ve ölüm cezasına bugünkü yaklaşımın bu şekilde geliştiğini ifade etmektedir. Diğer taraftan Yakowitz'in sınırlı erişim hakları konusundaki bir diğer eleştirisi ise bazı kuruluşların kişi-odaklı gizlilik yasalarını kendilerine bir kalkan olarak kullanmaları ve veri paylaşımı ve ifşası süreçlerinde bu yasaları referans göstererek keyfi kararlar vermeleridir. Veri işleyen ve yöneten kurumların bazı durumlarda yasalara sığınarak kendilerine gelen veri paylaşım taleplerini reddederken, bazı durumlarda da verileri talep eden kişi veya kurumlarla paylaşmaları, bu hususta takip edilen bir standart olmadığı algısını

oluşturmaktadır. Buna bir örnek olarak da 2008 yılında Kaliforniya Üniversitesi'nin lisans öğrencileri kabul komitesinden öğrenci kabul verilerini talep eden iki profesörün durumunu göstermektedir. Bu profesörlerden biri, öğrenci kabul süreçlerine istinaden eleştirel bir tutum içindedir ve kabul komitesinin süreçlerinde ırk ayrımını dikkate aldığını savunmaktadır. Bu profesörün talebi "ciddi gizlilik endişesi" gerekçesiyle reddedilirken, diğer profesör verilere ksııtlanmış bir lisans ile sahip olabilmiştir[222]. Bu bölümde Yakowitz, müşterek verinin gücünü vurgularken, ayrıca özünde bilgi istismarına açık olduğunu belirtmiştir. Ayrıca veri kümelerinin değerlerinin önceden hesaplanamayacağını ifade ederken, araştırma verisinin yayılmasını engelleyen her teşebbüsün belirlenemeyen ölçekte sosyal maliyeti olacağını savunur.

b) Teknik Yaklaşımlar ve Yorum Farkları

Yakowitz, anonimliğin bozulmasını sağlayan istatistik metotlarının ve bu metotların kişisel verileri açığa çıkarmadaki başarısının, kanun koyucuların ve düzenleyicilerin anonimleştirme hakkındaki görüşünü daralttığına inanmaktadır. Buna göre, bilişim literatürü beş temel kabulle ilerlemetedir; veri kümesindeki her değişken bir dolaylı betimleyicidir, ilgili kişilerden oluşan bir topluluğa ait veri destekli çıkarımlar gizliliği ihlal eder, faydalı veri ister istemez gizliliği ihlal eder, anonimliği bozan metotlar kolayca uygulanabilir, kamuya açık veri kümeleri bir saldırganın sahip olduğu verinin üzerinde değer içerir.

İlk olarak Yakowitz, Netflix ifşası örneğini temel alarak bu konuda yapılan çalışmanın yarattığı yanlış algıyı tartışır. Internet gibi büyük bir kaynağın anonimleştirme süreçlerini geniş arşiv ve büyük ölçekte verinin birleştirilebilir olması gibi özelikleriyle etkilediğini kabul etmekle beraber iş arkadaşları veya meraklı tanıdıklarımızın öğrenebileceği her bilginin gizlilik ve kişisel verilerin korunması mevzuatlarının kapsamında değerlendirilemeyeceğini ileri sürer. Bu

[222] Jane Yakowitz, s. 19

bağlamda veri kümelerindeki her değişkene dolaylı betimleyici muamelesi yapmak ve yasal çerçeveyi böyle geniş bir gizlilik tanımı içerisinde kurgulamak veri paylaşımını imkansız hale getirir. Yakowitz, Ohm'un çalışmasındaki "bir dahaki sefere yemeğe davet edildiğiniz bir yerde en sevdiğiniz 6 filmin ne olduğunu soran bir kişiye, eğer tüm Netflix oylamalarınızı öğrenmesini istemiyorsanız, sakın söylemeyin" yorumunu da açıkça eleştirmektedir. Yakowitz, Netflix ifşasının literatüre büyük katkısı olduğunu kabul eder ancak bu katkıyı teorik olarak değerlendirir. Gerekçe olarak ise, bu çalışmada kullanılan algoritmaların büyük ve çeşitli veri kümelerinde uygulanmaları gerektiği, veri kümelerinde ilgili kişi kayıtlarının aralıklı olarak sıralanması gerektiği, örnek veri kümesindeki kayıtları tam ve doğru bilgiler içermesi gerektiği ve ek olarak saldırıyı gerçekleştiren kişinin anonimliği bozan entropik metotları biliyor olması gerektiği gibi kısıtların varlığını hatırlatır. Halbuki bir başka çalışmada, 15.000 hasta kaydının dış veri ile birleştirilerek kimliklerinin tespit edilip edilmeyeceğini ölçmeye çalışan bir grup istatistikçi, bütün grup içinde sadece 2 kişinin %0.013 oranla tespit edilebileceğini ölçmüştür[223].

İkinci olarak Yakowitz, veri kümesinin bir alt grubuna ait çıkarımların ifşa olmadığını düşünmektedir. Bir alt gruba istinaden yapılan çıkarımları tüm veri kümesi nezdinde genelleştirmek basmakalıp anlayışların benimsenmesini sağlar. Ayrıca grup bazlı varsayımlar gruplar arasındaki farklılıkları anlamak ve kişinin hareketleri yerine toplu davranış şekillerine odaklanmayı hedefler.

Üçüncü olarak, Ohm'un anlayışının tersine Yakowitz, ifşa edilmiş veri kümelerinin hem faydalı hem de güvenli olabileceğini iddia etmektedir. Buna örnek olarak, cinsiyet gibi sadece tek bir dolaylı betimleyici ve sağlık malzemesi alımları gibi betimleyici özellikte olmayan bir değişken içeren iki değişkenli bir tablonun, kadınların ilaç satın alma oranlarını inceleyen bir çalışma için çok faydalı olacağını ve böyle bir veri kümesindeki anonimliğin bozulma riskinin bulunmadığını ortaya koyar.

[223] Jane Yakowitz, s. 28

Dördüncü olarak Yakowitz, anonimliği bozan algoritmaları kurgulamanın düşünüldüğü adar kolay olmadığını belirtir. Anonimliği bozacak birleştirme işlemlerinin basit uygulamalarla yürütülebileceğine inanan yaygın kanıya rağmen Yakowitz, eşleştirmeler esnasında uygulanacak algoritmaların arka planındaki hesaplamaların ve farklı kaynakların birleşmesinden oluşan yanlış eşleşmelerin değerlendirilmesi için eğitim ve tecrübe gerektiğini savunur.

Son olarak yazar, ifşa edilen bir kişi hakkında bilgi edinme motivsyonuyla hareket eden kişiler için anonimliği bozulmuş bir veri kümesinin düşünüldüğü kadar büyük öneme sahip olmadığını düşünmektedir. Netflix ifşasında olduğu gibi belli bir kişinin 5 veya 6 tane filme verdiği puanlamayı bilen saldırgan zaten bilgi sahibi olmak istediği kişi hakkında pek çok çıkarım yapabilmektedir. Böyle bir durumda entropi formüllerini anlamak ve uygulamak saldırganın zaman kaybından başka bir şey değildir.

c) Gerçekçi Riskler ve Öneriler

Yazar, kusurlu anonimleştirme metotlarını, art niyetli saldırganların varlığını, gizliliği müşterek veri dışındaki yöntemlerle ihlal edecek eylemleri gerçekçi riskler olarak değerlendirir. Anonimleştirme metotlarının doğru şekilde uygulanmaması, tüm riskleri hesaba katamıyor olması Yakowitz nezdinde de büyük risk teşkil eder. Bu konuda büyük şirketlerin bile yeterli özeni göstermediğine dikkat çeken Ohm ile aynı görüştedir. Ayrıca, yazar art niyetli kullanıcılarının varlığının bir risk olduğunu kabul eder ancak Internet'te kişilerin kendi yayınladıkları veriler ve ticari olarak erişilebilir müşteri bilgileri gibi veri kümeleri varken bu kullanıcıların kamuya ifşa edilen veri kümelerine odaklanmayacağını savunur. Bu noktada arkadaşlar tarafından yürütülen dolandırıcılık faaliyetleri veya para karşılığı satın alınabilen müşteri veri tabanları karmaşık anonimleştirme algoritmalarından daha yaygındır. Benzer şekilde kapsamlı veri kümelerini hedef alan art niyetli ve donanımlı kullanıcılar

anonimliği bozacak gelişmelere zaman harcama yerine, direk güvenlik açıklarına korsan saldırılar düzenleyerek veri ve kimlik hırsızlığı yapacaklardır. Sistemlerin güvenlik açıklarını kullanarak güvenlik duvarlarını aşabilen kötücül yazılımlara kara borsada bile erişmek mümkün iken, bu tip korsanlık aktiviteleri anonimliği bozacak algoritmaları kullanmaktan daha kolaydır.

Yakowitz'in sürece dair sunduğu öneriler şu şekildedir; verinin uygun metotlar kullanılarak anonimleştirilmesi ve risk değerlendirilmelerini yapılması, veri paylaşımı yapılacak partiler arasında özel anlaşmalar yapılması, veri istismarı halinde cezai yaptırımlar uygulanması. Yakowitz anonimleştirme tekniklerine önceki bölümlerde detaylandırdığımız direk betimleyicilerin çıkartılması, örnekleme ve k-anonimlik ve benzeri istatistiksel metotları saymaktadır. İkinci olarak, partiler arasında yapılacak özel anlaşmalarla anonimleştirilmiş verinin paylaşılmasının ticari amaçlı paylaşımları destekleyip, araştırmacıların veri taleplerini gizlilik yasalarını öne sürerek reddeden kurumların bu eğilimlerini engelleyecektir. Son olarak, herhangi bir araştırma amacı gütmeden veya bir araştırmanın sonucunda istemsizce gerçekleşmeyen, bir kişinin kimliğini ve bu kişiye air bir miktar kamuya ifşa edilmemiş veriyi açığa çıkaran kişilerle ilgili sıkı cezai yaptırımlar uygulanması gerektiğini öne sürmektedir. Bu cezai yaptırımlar özellikle belli kişileri hedef aldıkları durumda daha anlamlı olmaktadırlar.

Bunların yanında Yakowitz, özellikle Ohm'un ileri sürdüğü ve gizlilik savunucuları tarafından yaygın şekilde kabul gören bazı düşüncelerini de, gerekeçelerini sunarak eleştirilermektedir. Öncelikle Ohm'un giderek büyüyen veri yığınları içinde hemen hemen herkesin kişisel verilerinin açığa çıkartılabileceği tezine karşılık Yakowitz, Internet'in çok büyük bir veri kümesi barındırdığını ancak bu verilerin sistematik ve kapsamlı olmadıklarını vurgulamaktadır. Internet'teki verilere dayanarak yapılacak saldırılarda mutlaka eşleşmenin doğru olup olmadığına dair başka veri kümelerine ihtiyaç duyulacaktır. Ohm'a göre, aynı şekilde her yeni açığa çıkan kişisel veri kümesi yeni eşleşmelere imkan sağlayacak ve yeni anonimleştirilmiş veri kümelerinin de anonimliğinin bozulmasına sebep olacaktır. Ancak Yakowitz, her veri kümesi

birleşmelerinde mutlaka yanlış eşleşen sonuçların varlığına dikkat çeker. Eşleşmeler sonucunda ortaya çıkan yanlış kayıtlar yeni oluşan ve anonimliği bozulmuş veri kümesindeki bilgilerin doğruluğunu etkileyecektir. Bu durumda her yeni birleşmede hata payı da aynı oranda artacaktır. Son olarak Ohm, yeni anonimleştirme metotlarının geçmişte ifşa edilmiş veri kümeleri için geçerli olmayacağını, aynı şekilde Internet'teki veri birikimi ile anonimliğin zaman içinde riske gireceğini savunur. Yakowitz ise burada eski verinin kaybettiği değerine dikkat çeker. Çünkü kişilerin özellikleri değişkendir ve bu durum veri kümelerindeki kayıtların değerlerini etkiler. Zaman içinde değişen veri kümesinin ilk ifşa edildiği haliyle yapılacak birleşmeler yanlış ve tutarsız kayıtlar üretecektir. Diğer taraftan birleşmeler sonucunda doğru eşleşmeler elde edilse bile eski bilgiye sahip olmanın anlamlı bir değeri yoktur.

Yakowitz sonuç bölümünde, bu konuda çalışmalar üreten diğer tüm araştırmacılar gibi gizlilik politikaları ile veri ifşasının dengelenmesi gerektiğini, ancak dikkate alınan risklerin araştırma verisi için geçerli olamayacağını, bu sebeple kanunların araştırma verilerinin paylaşımını ve ifşasını desteklemesi gerektiğini belirtir. Bunların yanında kişilerin araştırma verilerinin içindeki kendi verileri üzerinde hak iddia etmeleri ve mülkiyet kavramına uygun olacak şekilde kendi verilerini araştırma verilerinin içinden çıkartmaya yeltenmeleri araştırma-geliştirme süreçlerinin sonunu getirecek aşırı korumacı sonuçlar doğuracaktır. Yakowitz'e göre paranoyayı beslemek yerine, hiçbirimizi açığa çıkarmayan ama hepimizi içeren araştırma verisine katkıda bulunmak yurttaşlık görevidir.

3. Anonimlik Dereceleri

Anonimleştirmenin güvenilirliğine dair tartışmayı başlatan teknik açıklar, anonimliğin derecelendirilmesinin gerekliliğini ortaya çıkmıştır. Yaşanan ihlaller anonimleştirmenin standart bir çözüm olmadığını, veri kümesine, değişkenlerin

dizilimine ve zamana göre farklı sonuçlar üretiğini göstermiştir[224]. Bu durumda araştırmacılar anonimliğin derecelendirilmesine odaklanmıştır. Bu konuda yapılan kapsamlı çalışmalardan biri Hollanda'lı bilişimci Matthiaj R. Koot tarafından gerçekleştirilmiştir. Koot, çalışmasında belediyedeki sicil bilgilerini kullanarak, Hollanda ulusal sağlık kayıtlarının ve Hollanda dolandırıcılık istatistik verilerinin dış veri kullanılmadan sadece belli dolaylı betimleyicilerin kombinasyonları sayesinde kişisel verileri açığa çıkarıp çıkarmadığını incelemiştir[225]. 2009 yılında Hollanda'da 12 eyalet ve 441 belediye bulunmaktadır. Belediyelerdeki sicil kayıt dairesi, vatandaşların resmi kayıtlarının tutulduğu mercilerdir. Bu çalışma kapsamında 15 belediyeden kayıtlar talep edilmiş ve toplamda 2.774.476 kişiye ait kayıt üzerinde yürütülmüştür. Ülkedeki ulusal sağlık kayıtları ise hastanelerin düzenli olarak gönderdikleri idari ve tıbbi kayıtlarla oluşmuş bir veri tabanıdır. Bu kayıtlar tedavi etkisi, hastaneler arası rekabet performansı ve epidomolojik çalışmalar için kullanılmaktadır. Kayıtlar sadece Hollanda vatandaşı olan kişilere aittir ve her bir kayıt bir kişinin hastane kabulu veya günlük bakım bilgilerini içermektedir. Araştırmacılar kayıtların yalnızca 2005 ve 2007 yılları arasındaki bölümüne ulaşabilmektedir. Dolandırıcılık istatistik kayıtları ise Hollanda vatandaşlarının maruz kaldığı dolandırıcılık faaliyetlerine ait soruşturmaların kayıtlarını içerir. Her bir kayıt bir kişi için tamamlanmış tek bir soruturmayı içerir, bir kişi için birden fazla kayıt bulunabilmektedir. Çalışmanın temel mantığı, k-anonimlik hesaplamasına uygun olacak şekilde belli dolaylı betimleyicilerden kombinasyonlar oluşturarak, bu kombinasyonları paylaşan kişi sayısına göre çıkarımlar yapmaktır. Temel olarak dikkate alınan dolaylı betimleyiciler posta kodu, doğum tarihi ve cinsiyettir. Hollanda'da posta kodları 6 haneli olduğundan bazı durumlarda son iki hane maskelenmiş ve 4 haneli posta kodu dikkate alınmıştır. Hesaplama sonuçlarından bazı çarpıcı sonuçlar şöyledir; 6 haneli posta

[224] Martin R. Koot'un aktardığına göre, Latanya Sweeniey'in 1990 yılında ABD nüfus kayıtları üzerinde cinsiyet, posta kodu ve doğum tarihi değişkenlerini kullanarak nüfusun %87'sinin kimliğinin saptanabildiği çalışmasını, 2000 yılında Philippe Gool aynı şekilde tekrarlamış ve nüfusun sadece %67'sinin kimliğini tam olarak saptanabildiğini hesaplamıştır. Bu durumu Golle'un, Sweeney'in veri toplama ve analiz tekniklerine dair detayları bilemediği için açıklayamadığı aktarılmıştır.

[225] Matthijs R. Koot, s. 29

kodu+cinsiyet+doğum tarihi kombinasyonunda 2.766.475 kişinin farklı kombinasyonlar oluşturmasından ötürü kimliğinin tespit edilebildiği ortaya çıkmıştır. Bu rakam toplam veri kümesinin %99.4'üne denk gelmektedir. Diğer taraftan 4 haneli posta kodu+cinsiyet kombinasyonunda ise toplam veri kümesinin %67'sine denk gelen 1.861.081 kişi teşhis edilebilmektedir. Ulusal sağlık kayıtlarında yer alan 4 haneli posta kodu+cinsiyet+doğum yılı+doğum ayı kombinasyonunda tomplam veri kümesinin %4.8'i tamamen teşhis edilebilirken, %79.1'i 10 kişi veya daha az sayıdaki gruplar içinde kümelenmektedir. Dolandırıcılık istatistiklerinde yer alan belediye+cinsiyet+doğum yılı+doğum ayı kombinasyonunda ise toplam veri kümesinin %0.07'si tamamen teşhis edilebilmekte, %2.14'ü 10 kişi veya daha az sayıdaki gruplar içinde kümelenmektedir. Sonuçlara göre belediye kayıtlarına ve ulusal sağlık kayıtlarına erişebilen bir kişinin kimlik teşhis edebilme oranları, aynı şekilde hem belediye kayıtları hem de dolandırıcılık kayıtlarına erişebilen bir kişiye göre daha fazladır. Bütün bu sayısal sonuçlara rağmen, Koot böyle bir çalışmanın gerçekleşmesi için özel izinlerle, belli bir amaca istinaden ve belli kişilere yetki verilerek erişilebilen bu veri kümelerine sahip olunması gerektiğini vurgular[226]. Bu yüzden, bu çalışma için gereken yatırım, anonimliği bozulan kayıtlardan beklenen katkıyla orantısızdır. Aynı şekilde Yakowitz de benzer durumlarda saldırganların anonimliği bozmanın maliyeti yerine daha kolay olan korsanlık faaliyetlerini tercih edeceklerini ileri sürmüştür. Ancak yine de Koot, böyle bir çalışmanın benzer dolaylı betimleyiciler içeren herhangi başka veri kümelerinde de uygulanabileceğini hatırlatır. Bu durumda bu tip betimleyicilere sahip her veri kümesi aynı riski barındırmaktadır. Bu sebeple Koot, ifşa edilmesi halinde bilginin istismar edilebilme oranı, yaratacağı duygusal, sosyal veya başka oluşabilecek zararları hesaba katan risk hesaplamaları ile ortaya çıkabileceği sonucuna varır.

[226] Martin R. Koot, s. 39

B. İçerik Tartışması

1. Hem Fayda Hem Gizlilik

Ohm ve Yakowitz'in çalışmaları, anonimliğin bozulmasıyla oluşan gizlilik tehditleri ve veri ifşasıyla sağlanan sosyal fayda arasındaki tartışmayı en iyi şekilde özetlemektedir. Bu çalışmaları dikkate alarak yapılmış bir başka çalışma da Felix T. Wu tarafından hazırlanmıştır. Wu, kendi çalışmasında kendisinden önceki tartışmalara eleştiriler getirir ve bu tartışmaların kavramların içeriklerini atladığını düşünür. Buna istinaden Wu, çalışmasında iki karşıt görüşü sentezlemeyi hedeflemiştir diyebiliriz.

Wu'ya göre anonimleştirmeyi savunan ya da eleştiren tüm çalışmalar bilişim literatürünün örnekleri üzerinden sonuçlar üretmektedir ve atladıkları nokta gizlilik ve fayda olgularının kavramsal olarak içerikleridir. Wu, bu çalışmalardaki gizlilik ve fayda ifadelerinin farklı algılarla oluşturulduğuna dikkat çeker. Wu'ya göre bu kavramların içeriğine yeterli dikkati göstermeden teknik sonuçlara bakarak aşırı genellemeler yapmak oldukça kolaydır.

Wu, Netflix ifşası ve benzer şekilde anonimliğin kolaylıkla bozulduğu örneklerin, anonimleştirme süreçleriyle ilgili farklı kutuplar yarattığını kabul eder. Buna göre Ohm ile benzer düşünceye sahip kesim, anonimleştirme süreçlerine bütün olarak şüpheci yaklaşmaktadır. Yakowitz ile benzer düşünceye sahip kesim ise bu olayların aykırı örnekler olduklarını ve gerçekçi risklerin araştırma verisi için geçerli olmadığını savunur. Ancak Wu'ya göre tüm kutuplaşma, bilişim ve istatistik sonuçlarının yanlış yorumlanmasının sebebidir ve ne Ohm kadar kötümser ne de Yakowitz kadar iyimser olunmaması gerekir. Wu, Ohm'un "fayda artarken gizlilik azalır[227]" önermesindekiyle, Yakowitz'in "çağdaş gizlilik risklerinin anonimleştirilmiş araştırma verisini etkilemediği[228]" ifadesindeki

[227] Felix T. Wu, s. 6
[228] Felix T. Wu, s. 6

gizlilik kavramlarının farklı anlamlarla kullanıldığını savunur. Eğer gizililik, komşularımızın, hakkımızda yapacağı basit tahminleri de kapsayacaksa ve eğer fayda, tamamen beklenmedik örnekler için veri madenciliği yapılmasına imkan verecekse, bu durumda bu iki kavram birbiriyle çakışacaktır. Wu'ya göre, gizlilik kimlik hırsızlığı gibi süreçleri kapsarken, fayda belirli istatistik araştırmalarını içermelidir. Burada Wu'nun dikkat çektiği husus, istatistiksel sonuçlardan çok, kavramların içeriklerinin doğru yorumlanmasının gizlilik ve fayda arasındaki dengeyi sağlayacağıdır. Bundan sonraki bölümlerde Wu, neden kötümser ve neden iyimser olunmaması gerektiğini açıklarken, gizlilik ve fayda kavramlarının içeriklerine dair yol gösterir.

İlk olarak, kötümser yaklaşımları destekleyen tehdit senaryoları incelendiğinde varsayımlarda şu detay dikkat çekmektedir; saldırganın ihtiyacının olduğu veri değeri ile veri kümesinde fayda sağlaması beklenen değer aynı olarak kabul edilmektedir. Örneğin bir veri kümesinde fayda sağlayacak değer, belli bir kayda ait medeni durum değeri ise, o veri kümesine bakarak tahminler yürütecek saldırganın da ihtiyaç duyduğu verinin medeni durum değeri olduğu varsayılır. Bu tip risk hesaplaması yapılan araştırmalar başlangıçta fayda ve gizliliği zıt kavramlar olarak kabul etmektedirler. Halbuki, verilerin değerlerinin genelleştirilerek ifşa edilmesi, araştırmacıların ihtiyacı olan kümülatif ve birleşik sonuçlar sunarken, belli bir kişinin kimliğinin açığa çıkmasını da engellemektedir. Verilerdeki genellemeler ve bozulmalar belli bir kişiye ait veri değerinin doğruluğunu etkileyerek, değerin belli bir kişiyle ilişkilendirilmesi ihtimalini azaltacaktır[229].

İkinci olarak, iyimser yaklaşımlar genel olarak k-anonimliğe aşırı güven geliştirmektedirler. K-anonimliğin önceki bölümlerde de incelenen eksikleri Wu'nun alışmasına da konu olmuştur. K-anonimlik vasıtasıyla ayrıştırılan alt gruplarda yeterli çeşitliliğin olmaması halinde, veri kümesindeki kayıtlara dair isabetli çıkarımlar yapılabileceğini önceki bölümde görmüştük. Aynı şekilde

[229] Felix T. Wu, s. 13

saldırganın kişisel ilişkileri veya imkanlarıyla edindiği dış bilginin büyüklüğü ve çeşitliliği de k-anonimlik kullanılarak anonimleştirilmiş veri kümelerine tehdit yaratabilir. Wu, bu noktada Yakowitz'in referans gösterdiği ve 15.000 kayıt içinde yapılan istatistik hesaplamalarına istinaden sadece %0.013 oranla iki kişinin tespit edilebildiğini gösteren çalışmaya farklı bir yorum getirmektedir. Çalışmanın detayları incelendiğinde araştırmacıların yalnızca tam bir kimlik saptamasının yapılabildiği durumları incelediklerini, başka hassas verilerle ilgili isabetli tahminler üretilebilecek durumları dikkate almadıklarını vurgular. Burada araştırmacılar sadece bir kişinin kayıtlar içinde isim soyadıyla beraber açıkça teşhis edilmesi olasılığını başarılı olarak değerlendirmektedir. Bu detay, Wu'nun bakış açısına göre anonimliğin güvenilirliğinin test edilmesinde dar bir yorumlamadır.

Wu'nun çalışması iki önemli tartışma etrafında şekillenmiştir. Gizliliğin içeriği ve faydanın içeriği nasıl belirlenecektir? Tehdit senaryolarının dışarıdan gelecek tehditlere odaklandığını, arkadaşlar, aile fertleri veya hizmet alınan şirketin veya kurumun çalışanları gibi içeridekileri kapsamadığını vurgular[230]. Bu bağlamda, farklı gizlilik tanımları farklı tehdit algıları yaratır ve Wu'ya göre tehditler statik değildir, zaman içinde ve etraflarındaki değişen dünya ile birlikte değişmektedirler. Ek olarak çalışmada, kimliğin tespit edilebilmesine dair hesaplanan olasılık bilgisinin her durumda tehdit yaratmayacağına dikkat çekilmiştir[231]. Bazı durumlarda, kimlik hırsızlığı gibi kritik sonuçlar doğmasına sebep olabilecekken, bazı durumlarda olasılıklar sınırlı bilgilere dayalı hükümler üretir ve bunlar etkili olumsuzluklar yaratmaz. Ancak mevcut yaklaşımlar, her olayda hesaplanan olasılık bilgisinin aynı muameleyi gördüğünü, bir kesim tarafından tamamen tehdit unsuru olarak değerlendirildiğini bir kesim tarafından da tamamen ihmal edildiğini ifade etmektedir. Wu'ya göreyse belli bir oranın içinde saklanmak bazı durumlarda işe yarar, bazılarında ise yaramaz.

[230] Felix T. Wu, s. 26
[231] Felix T. Wu, s. 30

Gizliliğin içeriği kadar faydanın içeriği de tartışma konusudur. Faydalı verinin ayırt edilmesindeki en büyük özellik verinin genelleştirilebilir olma özelliğidir. Bu özellik sayesinde kamu yararı sağlayacak bilgi ile kendi kişisel merakının gidermek amacıyla başkalarının gizliliklerini ihlal etme arasındaki fark belirginleşir[232]. Örneğin, bir veri kümesindeki belli bir yaş grubu, cinsiyet ve kilo değerine sahip kayıtların %50'si diyabet hastasıysa, bu veri o özelliklere sahip ancak kayıtları o veri kümesinde yer almayan bir kişi için de genelleme yapılabilmesini sağlar. Bu verilere göre ilgili kişinin diyabet olma ihtimali %50'dir. Bu kapsamda, genellemeler kesinlikle faydalı verilerdir. Ancak ilgili kişinin kayıtlarının bu veri kümesinde yer aldığını ve başka bir kişinin de bu kişiye dair dış veriye sahip olduğunu varsayalım. Bu durumda genellemeden bahsedilemez. Dış veriye sahip kişi, veri kümesinde yer aldığını bildiği ilgili kişi ile ilgili tahminler yaparak hastalığını tahmin etmeye çalışacaktır. Bu durumda da ilgili kişiye dair edinilen bilgi, %50 oranla diyabet hastası olduğudur ancak oranlar her iki durumda da eşit olmasına rağmen, ilk durum genellemelerle elde edilen faydalı veri olarak nitelendirilirken, ikinci durum art niyetli tahminler içermektedir. Sonuç olarak Wu, genellemelerin matematiksel değil, sosyal ve kavramsal olgular olduğuna dikkat çekmektedir[233].

Wu, fayda algısının yüksek oranla içeriğe dayalı olmasından ötürü, bilişim ve istatistik sonuçlarının faydayı doğru ölçemeyeceğini savunmaktadır. Bu konuda verdiği yaşanmış iki dava sonucu fayda algısının arasındaki farkların ceza hukukunda da ayrı tutulduğunu göstermektedir. İlk davada, az rastlanan bir hastalıkla ilgili yazılan bir makalede, bir hastanın kişisel verilerinin paylaşılması üzerine, hasta hukuksal süreç başlatmıştır. Bu durumda mahkeme, hastanın kişisel verilerinin hastalığa istinaden gerçekleri ve bulguları değiştirmeyeceğinden bu bilgilerin makaleye bir katkısının olmadığı sonucuna varmış ve davacıyı haklı bulmuştur[234]. Bir başka örnekte ise, Afrikalı-Amerika'lıların 20. Yüzyılda Amerika kıtasına göçlerini anlatan bir tarih kitabında ismi dahil kişisel verileri yer

[232] Felix T. Wu, s. 35
[233] Felix T. Wu, s.35
[234] Felix T. Wu, s. 38

alan bir kişinin, kitabın yazarına açtığı davada mahkeme davalıyı haklı bulmuştur. Gerekçe ise, tarihsel olayların gerçek kişilerin hikâyeleri ile desteklenmesi gerektiği, ancak bu şekilde tarih kitabı sayılabileceklerini, gerçek örnekler kullanılmaması halinde kitabın bir sosyolojik roman olacağı ileri sürülmüştür[235]. Faydanın ayırt edilmesindeki bir diğer sorun ise zaman içinde verinin hangi amaçlara hizmet edeceğinin kestirilememesidir. Fayda verinin soyut bir özelliği değil, içeriksel bir özelliğidir[236]. Bu noktada da sorun, içeriğin hangi yöne gideceğini kestirememektir. Bu sebeple Wu, gizlilikten tamamen vazgeçmeden, verinin olası tüm gelecek kullanımlarını desteklemenin mümkün olmadığını kabul eder.

Sonuç olarak, gizlilik tehditleri içeriğe bağlı olarak her durumda farklı yorulanmalıdır, tekdüze bir içerikle durum değerlendirilmesi yapılamaz. İçerik tartışması, teknik sonuçlara dayalı tartışmaların yanında daha temel bir özelliğe sahiptir. Anonimliğe güvensizlikle geliştirilen geniş gizlilik içeriği, ailemiz ve arkadaşlarımızın hakkımızda basit tahminler yapmasını da engellemek ister. Anonimliğe inanarak geliştirilen daha dar gizlilik içeriği ise hukuksal varsayımların geniş bir bölümünü dışlar. Ancak Wu temel hususların, içerikte çözülebileceğini savunur.

5. Sonuç

Günümüz enformasyonel toplumunun hammaddesi kişilerin ürettiği veridir. Her yeni gelişen uygulama, altyapı, arayüz, donanım sağladığı yeni iletişim, haberleşme, otomasyon imkânları ile toplumları sisteme bağlı hale getirerek, daha çok ve daha çeşitli veriler üremesine sebep olur. Daha çok veri daha çok bilgi anlamına gelir ve ilerleme döngüsel bir sürece girer. Gelişen teknolojinin ürettiği bilgi, yeni gelişimlerin önünü açmak için kullanılır. Bu ortamda verinin

[235] Felix T. Wu, s. 39
[236] Felix T. Wu, s. 37

paylaşılması ve ifşası kaçınılmaz bir hal almıştır çünkü gelişim, bilgiden maksimum faydayı sağlamak üzerine kurgulanmıştır. Ancak verinin yoğun paylaşımları ve geniş kitlelere ifşası, olası gizlilik ihlalleri ve kişisel verilerin yeterince korunamaması endişelerini şiddetlendirmiştir. Kişisel verilerin korunması ve özel alan gizliliği tartışması yeniçağın iletişim araçları ile daha hassas bir konu haline gelerek, veriden sağlanacak faydayı gölgeleyecek sonuçlar doğurmuştur. Bu noktada üretilen çözüm ise verinin anonimleştirilerek, kişisel bilgilerle olan ilişkilerinden koparılarak, paylaşımı ve ifşasının gerçekleşmesi olmuştur. Teoride anonimleştirme büyük veriden sağlanan fayda ile kişilerin özel alan gizliliklerinin sağlanması arasındaki dengeyi sağlayacak bir çözüm olarak kurgulanmıştır. Bu çalışmada da vurgulandığı üzere anonimleştirme hukuksal yaptırımlarla da desteklenen ve yasal mevzuatlara taşınan bir süreç haline gelmiştir. Ancak bu kadar güven duyulan anonimleştirme beklenen dengeyi sağlayabilmekte midir? Yaşanmış olaylar ve bu konu üzerine yapılan araştırmalar anonimleştirilmiş veri kümelerindeki anonimliğin kolaylıkla bozulabileceği ihtimalleri ortaya çıkardıktan sonra konuyla ilgili yukarıda detaylandırdığımız kapsamda tartışmalar doğmuş ve farklı görüşler kabullenilmiştir. Konu itibariyle anonimleştirme süreçleri ve kişisel veri korunması multidisipliner yaklaşımlar geliştirmeyi gerektirmektedir. Yalnızca teknik metotlar geliştirmek veya yalnızca hukuksal tedbirler almak dengenin ya aşırı korumacı ve engelleyici ya da riskleri fazlasıyla ihmal eden bir yöne doğru kaymasına sebep olur.

Bu bağlamda, öncelikle mevcut anonimleştirme ve ifşa/paylaşım süreçlerine dair kontrol edilmesi veya ölçülebilmesi mümkün olamayacak kadar zor olan parametreleri belirlemek çalışmanın anlamlı öneriler sunabilmesi için gereklidir. Bu parametrelere bağlı geliştireceğimiz kabullerle hareket etmemiz, sonuç üretmeyecek tartışmalardan kaçınmamız için anlamlıdır.

I. Kontrol Edilemez Parametreler

İlk olarak, belli bir anonimleştirilmiş veri kümesine bakarak art niyetle hareket edecek kullanıcıların sahip olduğu veya erişebildiği dış verinin kapsamı ve ölçeği hakkında öncül tedbirler alabilmek mümkün değildir. Dış veri, basit Internet aramaları, tanıdıkların kişisel cihazlarından bilgi kaçırılması, yakın ilişkilere istinaden edinilen gizli bilgiler, ortak yaşam ve iş ortamlarını paylaşma sonucunda edinilen bilgiler, bilgisayar korsanlığı yoluyla çalınan bilgiler gibi sayılabilecek pek çok ihtimalle sağlanabilir. Dış verinin hangi yollarla sağlandığı ve ne ölçüde bilginin ele geçebileceği gibi problemler, sınırlandırılamayacak kadar çok ihtimali içinde barındırır. Bu sebeple, anonimleştirilmiş veri kümelerindeki anonimliği ölçen araştırmalar veya riski minimize etmeyi hedefleyen istatistik metotları her zaman "en kötü senaryoyu[237]" dikkate almaya çalışmaktadır. Ancak art niyetli kişinin erişim yetkileri ve sahip olduğu bilgi çeşitliliği kontrol edilemez bir parametredir. Ayrıca kişisel verinin veya kimliğin tam olarak saptanmasına gerek kalmadan, edinilecek varsayımlarla bile ihlal gerçekleşeceğinden ötürü, anonimleştirme metotları her daim tam bir koruma sağlayamayacaktır. Bu yüzdendir ki art niyetli kullanıcılar yerine ifşa edilen veya paylaşılan veri kümelerinin içeriğine ve ölçeğine bağlı olarak önlemler almak veya metotlar geliştirmek daha etkilidir. Aksi takdirde dış verinin zenginliği gibi kontrol edilemez bir parametre açısından, her anonimleştirme metotu bir açık içerecektir.

İkinci olarak, Yakowitz'in vurguladığı anonimliği bozmanın maliyeti, özellikle de belli tedbirler alınmış veri kümeleri için caydırıcılık unsuru yaratmaktadır. Özellikle de kişilerin kendileri hakkında pek çok kişisel veriyi kendi rızaları ile yayınladıkları sosyal medya ortamlarına erişim kolaylığı dururken, art niyetli kullanıcıların anonimleştirilmiş veri kümelerinde, özellikle de yukarıda incelediğimiz araştırmacıların ilgilendiği detayda, anonimliği bozacak çalışmalar yapması kesin olmamakla beraber olası gözükmemektedir. Ancak, bu parametre için de kesin yargılar üretmek mümkün değildir. Anonimliği bozmak her veri kümesi için aynı maliyeti gerektirmeyebilir. Koot'un çalışmasında dikkate aldığı

[237] K-anonimlik sonrasında l-çeşitlilik ve t-yakınlık metotlarının geliştirilmesi bu en kötü senaryo ihtimallerinin bir sonucudur. Aynı şekilde Koot'un yürüttüğü çalışmalar veya Netflix ifşası hakkında yapılan çalışmalarda da art niyetli kullanıcının, hedef kişiler hakkında dış veriye sahip olduğu ve hedef kişinin ifşa edilmiş veri kümesi içinde yer aldığının bilindiği varsayılır.

belediye kayıtlarında herhangi bir dış veri kullanılmadan basit kombinasyonlarla veri kümesindeki kayıtların %99'unu tespit edebilmek mümkün olabilmiştir. Ancak aynı durum Netflix çalışması için söylenemez. Burada belli bir dış verinin varlığı söz konusudur ve entropik hesaplamalar yapılmıştır. Özellikle iki veri kümesinin birleştirilmesiyle sağlanan bozulmalarda veri kümelerinin kayıt sayılarına bağlı olarak doğru eşleşmeler yaratacak bir veri kümesi oluşturabilmek, yatırım ve zaman gerektirir. Bu durumda art niyetli kullanıcının hangi veri kümelerini hedef alacağı, korsan yazılımlar kullanarak mı yoksa anonimliği bozacak algoritmalar geliştirerek mi saldırılar düzenleyeceği kontrol edilemez parametrelerden biri olarak kalmaktadır.

Üçünü olarak, Wu'nun kapsamlı şekilde incelediği içerik kavramı bir diğer kontrol edilemez parametre olarak karşımıza çıkar. Gizlilik ve faydanın içeriğinin yaratacağı farklı sonuçlar her durumda farklı bir anlayış geliştirmeyi gerektirir. Ayrıca hem gizliliğin hem de faydanın içeriği direk olarak verinin içeriğiyle ilişkilidir. Bu sebepledir ki yasal mevzuatlarda da kişisel veri yanı sıra "hassas veri[238]" kavramına da raslamaktayız. Wu'nun ifade ettiği üzere arkadaş ve akrabalarımızın hakkımızda yapacağı basit tahminler de tahmin edilen verinin içeriğine bağlı olarak kişinin güvenliğini veya haysiyetini zedeleyecek ciddi tehditler yaratabilir. İçinde yaşadığı kültürel yapıya bağlı olarak cinsel tercihlerini açıklamayan bir kişinin ailsinin veya yakınlarının bu durumu açığa çıkarak varsayımlar yapabiliyor olması, kişinin hayatında kayda değer zarar yaratacaktır. Diğer taraftan da ilişki içinde bulunduğumuz tüm çevremizin bildiklerinden sakınabileceğimiz gizlilik politikaları geliştirmek de bilginin serbest dolaşımını imkansız hale getirecektir. Olasılıklar, veri kümelerindeki çeşitliliğe ve kombinasyonlara bağlı olarak farklı içeriklerle karşımıza çıkacaktır. Örneğin, Yakowitz tezini yalnızca araştırma verisi üzerine geliştirmekte ve saptanan gizlilik risklerini araştırma verisini etkilemeyeceğini savunmaktadır. Ancak büyük veri yalnızca araştırma verisinden oluşmamaktadır veya yalnızca araştırma amacıyla kullanılmamaktadır. Pek çok ticari ve politik amaçla başlatılan çalışma da büyük

[238] 95/46/AT sayılı Yönerge, m. 8

veriden faydalanmakta ve gerek bireyin gerek toplumun hareketlerini büyük veri sayesinde modellemeye çalışmaktadır. Daha da önemlisi, ticari ve politik amaçlar güden kurumlar da, yapılan araştırmaların sonuçlarında ifşa edilen veya paylaşılan veri kümelerinden faydalanabilmektedir. Bir veri kümesi, belli bir araştırmaya konu edildikten sonra sadece o araştırmayla ilişkilendirilecek amaçlar çevresinde kullanılacaktır gibi bir sınırlama getirilemez. Örneğin, hükümetlerin sırf kendi amaçlarına hizmet edecek araştırma enstitüleri kurması veya bu tip merkezleri desteklemesi araştırmacılara sağlanacak ek imkanlar, kişisel verilere ve özel alan gizliliğine tehdit oluşturabilecektir.

Son olarak, zamanla değişen veri kümeleri kontrol edilemez birer parametredir. Bu durumda da verinin içeriği beirleyicidir. Ek olarak, verinin hangi yöntemlerle, ne periyotta güncellendiği, tarihsel olarak ifşa edilip edilmediği gibi detaylar da pek çok farklı sonuçlar doğurabilir. Zaman verinin değerini kaybetmesine sebep olabildiği gibi daha anlamlı hale gelmesine de sebep olabilir. Adres bilgilerine istinaden tıbbi kayıtların yer aldığı bir veri kümesinde eğer kişi hala aynı adreste ikamet etmiyorsa olası bir anonimliğin bozulması durumunda tehdit altında olmayacaktır. Ama diğer taraftan, arama motorlarında yapılan sorgu cümlecikleri gibi bilgiler zaman içinde anlamlı bir özellik gösterir. Ya da lokasyon bazlı uygulamaların gelişmesi ile önemli bir hale gelen ve ticari bir algı kazanan lokasyon bilgisi de zaman içinde değeri artan veri türüdür. Veri türünün zaman içinde kazandığı veya kaybettiği değere benzer şekilde, anonimleştirilmiş veri kümesindeki anonimliğin derecesinin değişimi de zaman içinde farklı yönlere kayabilir. Belediyelerdeki kayıtlara ait veri kümelerinin ihlal riski yalnızca kayıtların toplandığı yıl için geçerli olabilir. Kişilerin hayatlarındaki doğum, ölüm, taşınma, boşanma gibi değişiklikler belediyedeki kayıtların sürekli güncellenmesine sebep olacağından zaman içinde anonimliğin bozulmasından doğacak riskin, kayıtlardaki kişilere etkisi azalacaktır. Ancak etnik köken ve din gibi zamanla ilişkilendirilemeyen değişkenlere dair yapılabilecek ihlaller zaman içinde bile değerini kaybetmez. Sonuç olarak, zaman, veri içeriğinde ve anonimlik derecesinde kontrol edilemez bir parametre durumundadır.

II. Öneriler ve İlkeler

Yukarıdaki tartışmaları dikkate aldığımızda konunun yalnızca teknik ve içeriksel boyutuyla ele alındığını ancak bu tartışmanın hukuki boyutunun geri kaldığını görmekteyiz. Anonimleştirme süreçlerinin teknik ve istatistiksel çözümler üretiyor olması, konunun yalnızca teknik çerçevesine odaklanılmasına sebep olmuş ve süreç istatistiksel metotların başarı oranlarına odaklanmış bulunmaktadır. Farklı olarak, Wu gizlilik ve fayda kavramının içeriklerine odaklanmış ancak bu içeriğin hukuki tanımı yerine sosyal içeriğini vurgulamıştır. Buna istinaden, bu çalışmada elde edilen en önemli sonuçlardan biri anonimleştirmenin güvenilirliği tartışmasının hukuki olarak ele alınmamış ve anonimleştirmenin genel esaslarının hukuksal bir yaklaşımla belirlenmemiş olmasıdır.

Ayrıca, yasal mevzuatlar incelendiğinde görülmektedir ki, anonimleştirme silme ve rıza kavramlarıyla ikame olarak ele alınmış ve birbirinin yerine geçebilen süreçler olarak değerlendirilmiştir. Veri yönetimi süreçlerinde verilerin silinmesi kayıtların tüm arşiv ve yedekleme ortamlarından geri dönüşsüz olarak yok edilmesi anlamına gelmektedir. Ancak böyle bir yok etme işlemi, ilişkisel veri tabanlarındaki mimariyi bozacağından sistem yöneticileri verilerin tamamen uçurulması yerine pasif olarak sistemde varlıklarını sürdürmesini tercih etmektedir. Yani veriler zaman içinde ilişkiler kurdukları tablolara, raporlara, veri ambarlarına zarar gelmemesi için tamamaen yok edilmez, sistemde pasif olarak tanımlanır. Örneğin mobil operatörüyle aboneliğini sonlandırmış bir müşterinin veriler, operatörün veri tabanlarından hemen silinemez. Bu durum o müşterinin verilerinin yer aldığı tüm strateji, pazarlama, trafik yönetimi vs. raporlarını bozacak bir eylemdir. Diğer taraftan, kayıtların çok eskimesi halinde verinin tamamen yok edilmesi halinde de o veriye yeniden ulaşmak mümkün olmayacaktır. Her iki durumda da silme işlemi anonimleştirilmiş veri ile denk değildir. Anonimleştirilmiş veri her daim belli kimlik saptama risklerini

barındırmaktadır. Ancak silme işleminin uygulanış şekline göre riskleri değişkendir ve anonimleştirilmiş veri ile bir tutulması yerine silme işleminin de süreçlerinin net şekilde çalışılması gerekmektedir. Benzer şekilde rıza kavramı da farklı dinamiklere sahiptir. Rızası alınan müşteri veya kullanıcının verileri genel bilgi güvenliği kuralları çerçevesinde işlenebilir hale gelmektedir. Ancak burada veri öznesiyle olan ilişkinin kopartılmasına dair bir şart koşulmamıştır. Hâlbuki, veri anonimleştirme süreci pek çok şartı ve hesaplamayı içerir. Rıza alındıktan sonraki süreç açıkça belirlenmediğinden, veri işlem sorumlusu olan işletmeciler veya kurumlar, rızası alınan veri öznesinin verileri üzerinde daha fazla hak iddia edebilmektedirler. Bu anlamda anonimleştirilmiş veri, rızası alınmış veriden daha güvenli hale gelmektedir.

Anonimleştirmenin pek çok farklı metotu ve uygulama alanı olması ve farklı veri kümelerinde farklı sonuçlar üretiyor olması, anonimleştirmeye karşı ya detaylara aşırı odaklanan ya da çok geniş bir açıdan değerlendiren anlayışlar benimsenmesini sağlamıştır. Bu durumda da, yapılan araştırmalar araştırmacının tercihlerine göre farklı boyutlarıyla yorumlanmış ve anonimleştirmenin esaslarının belirlenmesini engelleyecek, olayları sadece tek yönüyle ele alan çıktılar üretmiştir. Burada önemli olan, anonimleştirmenin çerçevesinin ve ilkelerinin belirlenmesi ve anonimleştirmeye hukuki bir yaklaşım kazandırılmasıdır.

Buna istinaden, anonimleştirmenin ilkelerini şöyle tanımlayabiliriz;

- **Anonimleştirme tekil bir çözüm olarak ele alınmalıdır:** Anonimleştirme ikame bir çözüm olarak değil, uygulama alanı ve sınırları belli bir tekil çözüm olarak ele alınmalıdır. Silme ve rıza gibi farklı dinamikleri olan süreçlerin bir ikamesi olarak ele alınması anonimleştirme süreçlerine karşı hukuki yaklaşımda yanlış algıların oluşmasına sebep olmaktadır. Böylesi bir durum, hukukçuların ve kanun koyucuların süreçlerin işleyişlerine dair fazla genel bakış açıları geliştirmesini sağlarken, süreci hayata geçirecek olan sistem yöneticilerinin keyfiyetle hareket etmesini sağlayabilmektedir.

- **Anonimleştirme veri kümesinin niceliğine ve niteliğine bağlı gerçekleşmelidir:** Anonimleştirme süreçleri anonimleştirmenin uygulandığı veri kümesinden bağımsız olarak ele alınamaz. Burada önemli olan verinin niteliği, hassas ve özel kategorilerde veriler içerip içermediği, veri öznelerinin koruma dereceleri (çocuklar v.s), verinin çeşitliliği ve büyüklüğü, dış veriye olan hassasiyeti konularında değerlendirmelere tabi tutulduktan sonra anonimleştirmenin uygulanıp uygulanmaması gerektiğine ve hangi metotun daha uygun olacağına karar verilmelidir.
- **Anonimleştirme iş ve çalışma modellerini dikkate almalıdır:** Veri sorumlularının ve veri odaklı çalışan tüm ticari ve idari kuruluşların çalışma yöntemleri birbirinden farklıdır. Bu durum kuruluşların veri yönetim süreçlerine, veri politikalarına ve yapılan yatırımlara yansımaktadır. Bu çeşitliğin içinde anonimleştirme süreçleri kuruluşların iş ve çalışma şekillerini dikate alarak uygulanmalıdır. Bir üniversitenin verilerine istinaden uyguladığı anonimleştirme süreci ile dünya çapında tanınan bir arama motorunun uygulaması beklenen anonimleştirme süreci aynı olamaz. Kuruluşun sahip olduğu veri hacmi, veri yönetimi yatırımları, tabi olduğu güvenlik politikaları, bilinirliği, dış kaynak ilişkileri, yurt dışı bağlantıları gibi iş modelini etkileyen kriterler dikkate alınarak anonimleştirme çözümleri değerlendirilmelidir.
- **Anonimleştirme seviyelendirilmelidir:** Anonimleştirme kuruluşların bilgi güvenliği politikaları nezdinde seviyelendirilmeli ve hangi şartlar altında başvurulacak bir çözüm olduğu netleştirilmelidir. Özellikle şirket içi paylaşımlar söz konusu olduğunda bilgi güvenliği politikaları, yetki profilleri, erişim kısıtları, fiziksel önlemler gibi süreçler dikkate alınarak anonimleştirmenin konumu diğer tüm tedbirler içinde netleştirilmelidir.
- **Anonimleştirmeye bağlı ihlaller öncül ve ardıl yaptırımlarla denetlenmelidir:** Hukuksal yaptırımların sadece öncül ya da sadece ardıl olarak ele alınması anonimleştirme riskleri hususunda eksik yaklaşımlar gelişmesine sebep olacaktır. Yalnızca öncül yaklaşımlar geliştirilmesi, yukarıda incelediğimiz üzere kurumların anonimleştirme metotlarını

uyguladıktan sonra yasal yükümlülüklerinden kurtuldukları imajını yaratarak olası bir ihlal durumunda sorumluluk almalarını engelleyecektir. Aynı şekilde fazla korumacı gizlilik politikaları, kurumlar tarafından art niyetli veya dar yorumlanarak özellikle araştırma ve geliştirme süreçlerinin devamlılığı için gerekli olan veri kümelerini paylaşmaktan veya ifşa etmekten kaçınmalarına yol açabilir. Diğer taraftan yalnızca ardıl yaptırımlar uygulanması, zararın oluşmasından sonra sürece müdahale edilmesini gerektirir. Ancak günümüz teknolojik gelişmeleri ve enformasyonel toplumu yalnızca zararın tazmin edilmesine odaklanan yaklaşımlar geliştiremeyecek kadar karmaşık sorunlar üretmektedir. Literatüre "Streisand Etkisi[239]" olarak geçen ve 2010 yılında yaşanan olaylar bu duruma en güzel örnektir. Internet'te yayınlanan kişisel bilgilere istinaden başlatılan hukuksal süreçler daha çok ilgi yaratmakta ve verilerin daha hızlı yayılmasını ve daha geniş kitlelere ulaşmasını sağlayan ters bir etki oluşturmaktadır. Bu durumda kişisel verilerin ve özel alan gizliliğinin korunmasına istinaden geliştirilecek anonimleştirme süreçlerindeki olası bozulma ve ihlaller için ardıl ve öncül yükümlülükler multidsipliner bir anlayışla tasarlanmalıdır.

- **Anonimleştirme muafiyet getirmemelidir:** Anonimleştirilmiş verinin "tüm veri koruması ilkelerinden muaf tutulması[240]", anonimleştirme sürecini yerine getiren işletme veya kurum için veri güvenliğini sağlanmıştır algısını oluşturmaktadır. Halbuki anonimleştirilmiş veri de hassas veriler gibi ayrı bir veri sınıfı olarak algılanmalı ve anonimleştirmenin olası risklerine istinaden de güvenlik önlemleri önemini korumalıdır.

[239] 2003 yılında Barbara Streisand Kaliforniya kıyılarında erozyona dikkat çekmek adına bölgenin tepeden pek çok fotoğrafını çeken fotoğrafçıya, kendi malikanesinin de fotoğraflarını çekip yayınladığı için 50 milyon dolarlık bir tazminat davası açar. Davanın basına yansımasından sonra site yaklaşık 420.000 kişi tarafından ziyaret edilir ve Streisand'ın malikanesinin fotoğrafları Internet'te hızla paylaşılmaya başlanır. Bkz. http://www.californiacoastline.org/news/sjmerc5.html

[240] 95/46/AT sayılı Yönerge, parag. 26