How to Deal with Security and Privacy Threats in the Context of Data Science

Sandro Schweiss

IU International College

Ethical Considerations in Data Science DLBDSSECDS01

Tutor Claudia Heß

Abstract

The privacy and security of data have become increasingly important in recent years. The industry, the everyday person, and cybercriminals have become progressively invested in the topic of data. Furthermore, since the subject of data is at the core of the data science profession, it is the obligation of a data scientist to handle data appropriately. Therefore, this research essay addresses the question of “How to deal with security and privacy threats in the context of data science?”. The topics of data privacy and security are elaborated and then translated to the context of data science. It is then discussed how data privacy and security impact data science. The following segment elaborates on what precisely the problems of privacy and security in data science are and how to best deal with these issues.

Table of Contents

I. **Table of Abbreviations** 3

2. **Introduction** 4

2.1 Overview 4

2.2 Rationale for the Topic 5

2.3 Aim of the Research Essay 5

2.4 Delimitation and Research Methodologies 5

2.5 Organization and Structure 5

3. **How to Deal with Security and Privacy Threats in the Context of Data Science** 6

3.1 Difference between Data Privacy and Data Security 6

3.2 What is Data Privacy? 6

3.3 Why is Data Privacy important? 7

3.4 What is Data Security? 7

3.5 Why is Data Security Important? 8

3.6 What is Data Science and its Role in the Industry? 8

3.7 What is Data Privacy and Security in the Context of Data Science? 8

3.8 The Impact of Data Privacy and Security on Data Science 9

3.9 What are the Security Threats in Data Science and How to Deal with Them? 9

3.9.1 Infrastructure Security and General Security 9

3.9.2 Hardware Security Threats and How to Deal with Them 10

3.9.3 Software Security Threats and How to Deal with Them 11

3.9.4 Network Security Threats and How to Deal with Them 11

3.9.5 Data Security Threats and How to Deal with Them 12

3.10 What are the Privacy Challenges in Data Science? 12

3.10.1 Transparency and Consent 12

3.10.2 Right of Access and Deletion 12

3.10.3 Data Anonymization and Minimization 13

4. **Conclusion** 13

V. **Bibliography** 14

I. Table of Abbreviations

Abbreviation	Meaning
IDS	Intrusion detection system
IT	Information technology
SDLC	Software development life cycle
NIDS	Network intrusion detection systems
GDPR	General Data Protection Regulation

2. Introduction

2.1 Overview

The usage of data in the industry is not a new phenomenon. The first information systems, for example, can be traced back to even the 1960s. Over these decades, data has become a crucial aspect of the business world (“IU Learn,” 2021b). However, not only the use of these information systems but also the data itself increased and is growing exponentially. It is estimated that the total volume of data will increase to 180 zettabytes in 2025, which is a three-fold increase in just four years (Statista, 2021).

Large Organizations like Yahoo, Sony, Apple, Google, et cetera have been confronted with this problem for many years now. They rely heavily on data and machine learning to efficiently store and process this vast amount of data. However, data is not just for large corporations anymore. Progressively more startups and small to medium businesses need more data for their operations. Even the everyday person is now confronted with data on a daily basis. (Tellenbach, Rennhard, & Schweizer, 2019). This is where data scientists become increasingly important. Large amounts of data can be difficult or even impossible to manage without the help of data scientists. Data science has led to incredible advancements in various fields and is becoming an essential profession in the business world. However, this progress is threatened by the increasing privacy and security challenges (Altman, Wood, O’Brien, & Gasser, 2018). The immense prevalence of data in our current society has unfortunately also led to a rise in cybercrime. Wherever there is data, hackers will want to get access and manipulate it, either for fun or for personal gain (Tellenbach et al., 2019).

Various security threats like unauthorized access, malware attack, zero-day attack, or phishing have grown exponentially in the last decade, and the financial losses of the data breaches lie in the millions (Sarker et al., 2020). But it is not only the standard cyber-attacks data scientists need to be aware of. Many of the common methods employed in data science are not inherently secure. It is, for example, enough to have access to the public API of a classification service to potentially steal confidential data (Tramèr, Zhang, Juels, Reiter, & Ristenpart, 2016).

Data security and privacy issues are a massive roadblock in the evolution of data science methods. Increasingly more focus needs to be shifted to design products with security in mind and adhere to the strict laws and regularizations of data privacy. On the other hand, however, data science methods have led to immense improvements in existing security products. The detection of anomalies in systems, for instance, is increasingly improving and helps in various fields to protect data. (Tellenbach et al., 2019). Even completely new solutions are being created, such as next-generation anti-virus products (“Cylance,” 2021a). Therefore, the following research essay discusses various strategies to handle these security and privacy threats in the context of data science.

2.2 Rationale for the Topic

The motivation and rationale behind the research topic “How to deal with security and privacy threats in the context of data science?” is that data privacy and security have become an essential subject in recent years. The commercialization of data developed into an enormous market, and progressively more companies, as well as the everyday person, started to notice the true value that data contains. While companies want to gather as much data as possible from their user base, the everyday person is reluctant to give strangers their personal information. This is especially the case when this information is threatened by the ever-increasing cyber-attacks that do not even let tech giants like Facebook unaffected. In the heart of the subjects, privacy and security of data stand the profession of data science, which is confronted with these issues on a daily basis. The question arises, how to best deal with security and privacy threats in data science?

2.3 Aim of the Research Essay

As previously mentioned, the core question and topic of this research essay is how to deal with security and privacy threats in the context of data science. The specific aim of this research is to discuss this topic and pursue the following research questions.

I. How does data security and privacy impact data science?
II. What are the exact security and privacy threats in data science?
III. How to deal with security and privacy threats in the context of data science?

2.4 Delimitation and Research Methodologies

The methodology used in the following research essay is the review of related literature. In this approach, relevant information from multiple literature sources is extracted to resolve the aforementioned research questions. The current state of the art is explored, and possible solutions are given.

2.5 Organization and Structure

Section 4 presents the main body, which is the research implementation. The beginning of the main body is an introduction, followed by a general overview of the essential topics; privacy, security, and data science, which are the foundation of the research essay. It is then elaborated on how the terms privacy and security of data translate to the context of data science. The relationship between security, privacy, and data science is then explored. This includes the impact of data privacy and security on the profession of data science. At first, the security aspect is described in the segment of how data privacy and security impact data science. This includes common security threats that can occur on a day-to-day basis. The security segment introduces data science methods and strategies to deal with these security threats. Next follows the privacy aspect. This is in the form of government regulations, which impose constraints on the activities of a data scientist. The privacy segment elaborates on how data science can be used to increase a user’s privacy and what to take care of when handling data.

3. How to Deal with Security and Privacy Threats in the Context of Data Science

3.1 Difference between Data Privacy and Data Security

Before defining the terms data privacy and security, it is necessary to elaborate and distinguish the differences, as they are often used interchangeably but have completely different meanings. Data privacy governs the data of an individual or natural person and determines how the data is collected, shared, and used. On the other hand, data security is not restricted to the individual and can affect organizations and governments as well. It protects the data from external attacks or unauthorized access and deals with access control and encryption issues. (Sen & Basahel, 2019).

3.2 What is Data Privacy?

To fully define data privacy, it needs to be inspected from the organizational point of view as well as the perspective of an individual and as a law itself. In the case of an organization, data privacy, a subcategory of data protection, is about the proper handling of data based on its relative importance. This includes sensitive data like personal information, financial data, and intellectual property. This constitutes information that can be used to identify an individual and possibly harm them. The proper handling refers to the adherence to regulatory requirements to protect the confidentiality and immutability of this sensitive data (“SNIA,” 2021c).

The privacy laws organizations must adhere to are defined by the respective government and their legal definitions. In the case of the European Union, privacy is recognized as a fundamental right. It is the right of the individual to be free from unwanted intrusions. Privacy itself is a legal concept, and the term data privacy deals with the technical framework of adhering to the regulations. A legislation example is the GDPR, imposing obligations onto organizations, which collect data from EU citizens. Violations of these standards result in wide-ranging fines (Comm/dg/unit, 2017).

From the perspective of an individual, privacy is referred to the ability of a user to control what data and with whom it is shared with (Wieringa et al., 2021). This is done via applying access control. The user enters a relationship or trade with the organization, which is then also referred to as a data controller. The user gets access to the product or service often in trade with the respective user information, which the company now holds. The data can then be given to a third-party analyst, commonly a data scientist, to gain deeper insights. After the trade, the data also enters the public domain and becomes a threat to individual privacy. It is the duty of the data holder to warrant the privacy of personal information. However, the user also plays an important role. Not reading of terms and conditions, as well as allowing access to camera, files, and contacts, all contributes to a possible data leakage (Ram Mohan Rao, Murali Krishna, & Siva Kumar, 2018).

3.3 Why is Data Privacy important?

As previously elaborated, sensitive data constitute information that can be used to identify an individual either directly, through an identifier, like a name or identification number, or indirectly, via the combination of different factors, like physical or social identity (Wieringa et al., 2021). It is therefore of utmost significance to maintain data sovereignty, which refers to the abiding of the laws of the region in which the digital data is collected (Irion, 2012). Businesses need to show transparency in how they keep, share, and use the data to build trust with their consumer base and ensure regulatory compliance. If this is not the case, the resulting repercussions could be severe. Disclosing of personal information could range from potential harm to the individual reputation to even identity theft, blackmailing, or loss of employability (Altman et al., 2018). In addition to that, the organization could be fined for up to 20 million euros or 4% of total annual turnover (Comm/dg/unit, 2017).

3.4 What is Data Security?

(Lundin, 2019) defines *information security* as the practice of protecting information from unauthorized use, disclosure access, modification, or destruction, regardless of the form the information takes. Since the medium of data security is computer networks, the previous definition only applies to the realm of information technology. Therefore, the two main challenges of data security are information assurance and IT security.

Information assurance is the insurance that no data is lost and always correctly available to the user, regardless of the circumstances. It also encompasses risk management relating to the use, processing, storage, and data transfer. This includes the practices of confidentiality, integrity, availability, and authentification. At its core, it is the supply of the correct information to the right user at the right time (Sen & Basahel, 2019).

IT security includes the security measures applied to a computer network and the protection of computer systems from information disclosure, theft, or damage to software or hardware components. Through the expanded growth of computer systems and the internet, this field has also become more and more significant (Schatz, Bashroush, & Wall, 2017). IT security includes vulnerabilities and attacks of cybercriminals, as well as respective counter methods for computer and network protection. The magnitude of cyber attacks is wide-ranging. Threats of cyber attacks can be standard methods like malware, denial-of-service, or phishing, performed by single individuals conducted on a natural person (Biscontini, 2019). But, they can also be large-scale coordinated attacks on organizations with pervasive methodology, including data science methods. In this scenario, the term cyber attack could also be deemed cyber terrorism (Purdy, Elizabeth Rholetter, PhD, 2021). Countermeasures can be applied on a design, architecture, hardware network, or even user level. There are a broad range of methods that try to prevent a possible cyber attack, and techniques can also range from simple to extensive (Biscontini, 2019).

3.5 Why is Data Security Important?

Data security has become a crucial aspect in the increasingly modernized world. Nearly every person, public or private, stores data in some shape or form on electronic devices. The previously mentioned dangers of data privacy also apply here since the unauthorized breach in data always has a malicious intent behind it. In contrast to the importance of data privacy, which mainly focuses on the individual, data security is especially crucial in the public sector. A data breach on a corporation could put business data in the hands of a competitor, which can financially ruin the company. An attack on financial institutions can lead to potential losses in the millions. And getting unauthorized access to health institutes could lead to the disclosure of highly confidential data (Andress, 2011). In the worst-case scenarios, it can be termed cyber terrorism, where the target is an objective of monumental importance. Getting ahold of the controls of a nuclear power plant, for example, has the power to destroy a whole country. Cyber security's impact on the current world can really not be understated, which makes the proper use of data science all the more important (Matusitz, 2008).

3.6 What is Data Science and its Role in the Industry?

Data science is an interdisciplinary field of statistics, computer science, and business intelligence, which uses processes and algorithms to gain insights from data. These insights can then be used to solve problems and questions of the organization. It is a discipline that focuses on unlocking the actual value of data so that the everyday person can understand and use it for their benefit. This immense advantage, in combination with a world that is exponentially increasing in data is the reason for the rising popularity and importance of data science (Dhar, 2013). Progressively more data scientists take up the role of the “data controller” inside organizations or “data processors” as the third party. This, in turn, means that the data scientist has to take up the challenges associated with the processing of data, which include data security and data privacy (Comm/dg/unit, 2018).

3.7 What is Data Privacy and Security in the Context of Data Science?

In the previous chapters, the general definitions and importance of data privacy and data security, as well as the profession of data science, have been elaborated. Therefore, it is crucial to specify how the previous elaborated challenges translate to the context of data science.

As previously established, the data scientist takes up the responsibility of data controller or processor to process the given data and gain valuable insights. This is often achieved through the use of machine learning models and other data science methods on computer systems. Therefore, a data scientist faces multiple stages of challenges when processing the data. From a security perspective, there are the challenges of infrastructure security, software security, data protection, exploitation of machine learning algorithms, and other common vulnerabilities, which do not exclusively impact data scientists. On the other hand, the challenges of data privacy include transparency of the data, data privacy legislation as a whole, and data anonymization techniques (Tellenbach et al., 2019).

3.8 The Impact of Data Privacy and Security on Data Science

While data privacy legislation policies and an increase in software protection is a necessary step in the right direction, all these challenges provide considerable roadblocks to the data science field. While differing in their scope, data privacy laws include restrictions on the collection, sharing, use, and storage of data about a natural person. These regulations require organizations, and in further extension, the data scientist, to comply with an extensive list of laws and obligations. These requirements can obstruct organizations from collecting specific data, using data only with the individual's consent or the necessity to delete data after a particular time. The global trend is towards more and stricter privacy laws, which collides with the trend towards increasing importance of data and data science (Hintze, 2019b). While not being an exclusive threat to the data science field, data security also has an enormous impact. Data science being a relatively new field in combination with its rapid evolution leads to short development lifecycles of new software, hardware, and methods. This can result in severe security loopholes (Pauli, 2017). Data warehouses, which act as centralized storage of an organization, are also implemented increasingly, making them a prominent target for coordinated attacks. Even the machine learning models used by data scientists can be exploited. Data scientists face security threats at nearly every data processing stage, making the extraction of insights increasingly tricky. It is therefore vital to elaborate how these threats can be dealt with in the most effective way (Tellenbach et al., 2019).

3.9 What are the Security Threats in Data Science and How to Deal with Them?

At first, the security threats are elaborated in detail to get a thorough understanding of the different types of threats and challenges. Then after the threats have been elaborated, the respective countermeasures and best practices are explained.

3.9.1 Infrastructure Security and General Security

IT infrastructure is defined as a set of multiple components necessary to manage and operate an IT environment. The IT infrastructure can be a cloud computing system or a company's internal IT system. The three-component groups included are hardware, software, and networking components (“What does IT infrastructure mean?,” 2021d). The hardware components are, for example routers, computers, servers, and data centers. The software component aspect includes, for example operating systems and machine learning tools. And the networking component includes firewalls, internet connectivity, and network enablement (“What is IT Infrastructure?,” 2021e). Infrastructure security is then the protection of IT systems against physical intruders, virtual intruders, technical failures, and insider threats. This is the fundamental level of security for the everyday business environment and encapsulates a broad range of components, as well as their own challenges and threats. While data scientists are not directly involved in the creation or development of these systems, IT infrastructure is the foundation of the activities of a data scientist (Tellenbach et al., 2019).

Due to the broad range of components, with their own respective threats and challenges, infrastructure security is the most complex aspect to get right. Failing to build strong infrastructure security can lead to severe repercussions. For example, the cost of the security breach of Sony in the year 2007 is approximated to be $171 million (Schreier, 2011). Most of the protection tasks of infrastructure security are out of the hands of the data scientist, as the IT environment usually is not set up by the data scientist. But best practices can still be taken to avoid common threats.

3.9.2 Hardware Security Threats and How to Deal with Them

Hardware security issues have become an increasing threat in the last two decades. Due to modern computing hardware being crafted in different locations around the globe, with a different level of trust surrounding the devices. Hardware security threats can arise during all hardware development life cycle stages. This ranges from unintentional design flaws, system side effects to intended malicious design modifications. The addition of rich connectivity features of these devices can lead to remote attacks. The common threats in this area are architectural and system threats like secure boot attacks and firmware attacks, which manipulate key components and processes of the computer. Covert and side channels, like timing or power channels, leak information to an outsider. IP theft and counterfeiting threats generate a counterfeited version of the IPs/Ics to sell to a vendor. Hardware trojans are a malicious modification of the circuitry to create a back door to the device.

While it usually is not in the area of obligations for a data scientist to develop countermeasures, there are some criteria that can be considered when choosing devices and selecting tools that protect against these common threats. The hardware security properties are formal specifications of the security-related behavior of circuit designs, which can be taken into considerations when using hardware. The properties are dependability, confidentiality, integrity, isolation, constant time, and quantitative security properties. *Dependability* is the assessment of the trustworthiness of computing hardware to perform the expected function. *Confidentiality* is a security property, which states that secret information should never be able to be obtained or inferred by observing a memory location or public output. *Integrity* is the property that a trusted object should never be overwritten by an untrusted entity. *Isolation* is a two-way property, which states that two hardware components of different security levels should not directly communicate with each other. *Constant time* is a property that prevents learning any information about the inputs by observing the computing time. And at last quantitative security properties, which enable quantitative measurement of hardware design security.

Taking these properties and best practices into account and evaluating suitable hardware devices, either with tools or through comparison, is a possibility for a data scientist to reduce or minimize common hardware threats (Hu et al., 2021).

3.9.3 Software Security Threats and How to Deal with Them

Software security issues in the context of infrastructure are concerned with the correct choice of software with security in mind. Data scientists often have to select and deploy base technologies and programs. The software options begin with operating systems and include every program used to maintain the daily operations of a data scientist, such as MongoDB or Apache Spark. It is therefore crucial for a data scientist, like in chapter 3.9.2, to be aware of the security of these products (Tellenbach et al., 2019). An example of this is the 30 000 MongoDB instances that were compromised because the versions before 2.6.0 were insecure by default (Pauli, 2017). The problem was known and even documented, but many operators were oblivious to this (Matherly, 2015).

The basic software a data scientist has to choose even begins with the operating system since even that can have a significant impact on preventing threats. Most common operating systems like windows and IOS are popular targets for cyber attacks due to their popularity. These operating systems also were not created with security in mind. An alternative to these systems is, for example, Linux, which is already used as software for servers. There are also other operating systems, which were explicitly created with security in mind (Taylor, 2018). Due to data science being a relatively new field, much of the software is developed for easy experimenting and not necessarily for security. However, these design flaws can lead to costly outcomes, as mentioned before. Data scientists often have to select suitable databases and machine learning models. Machine learning threats are, for example, training data poisoning, which can lead to error-specific attack purposes, backdoors, and model inversion attacks can lead to theft of sensitive data (Xue, Yuan, Wu, Zhang, & Liu, 2020). Database threats are, for example, SQL injection, which is the insertion of unauthorized database statements into a vulnerable SQL channel or underlying platform vulnerabilities, like a zero-day attack, which is a security loophole at the release of a new update (Al-Sayid & Aldlaeen, 2013).

The general countermeasure for software security is to check the security features if the default configuration is secure and to read the documentation. Overall estimation of software security is also necessary to evaluate if the software was designed with security in mind and if the developer organization is trustworthy or had previous exploits (Xue et al., 2020). These general metrics can also be applied when data scientist develops their own software. SDLC need to integrate security in every stage of development and use security as an additional evaluation metric (Stewart, 2012).

3.9.4 Network Security Threats and How to Deal with Them

Network infrastructure security is the process of protecting the underlying network infrastructure through installing preventative measures and tools. This is the most frequent type of security threat, which includes risks like phishing, a type of online fraud, computer viruses downloaded from websites or emails, denial-of-service attacks, which hinder users from accessing services through overloading traffic, and many more. The focus here is to penetrate the corporate network and gain access

to internal systems (Wheelus & Zhu, 2020). The countermeasures against these attacks are a general understanding of the internet to avoid common attacks like phishing and scams. Basic tools like anti-virus software can be installed to reduce the risk of viruses. More advanced technology can also be implemented. For example, methods like NIDS monitor large networks and try to detect irregularities (Mendyk-Krajewska & Mazur, 2010).

3.9.5 Data Security Threats and How to Deal with Them

One of the core activities in data science is the processing of large amounts of data. Data is often stored and processed in unencrypted form, leading to two major disadvantages. If data theft occurs, the information can easily be extracted. For data processing, cloud computing is often employed. However, if the data contains confidential information or is subject to laws prohibiting the processing of this information, cloud computing is not possible (Prasanna & Akki, 2015). To circumvent these problems, data must always be stored in encrypted form. In this way, the data must be stolen during the processing where the encryption takes place. Another possibility is the processing in encrypted spaces through the employment of searchable or homomorphic encryption, which implements an encrypted keyword search index or allows operations on encrypted data (Brown, 2017).

3.10 What are the Privacy Challenges in Data Science?

Privacy Challenges in data science are derived from data protection laws, which differ depending on the respective region. The privacy challenges and solutions discussed in the following sections are more generally applicable.

3.10.1 Transparency and Consent

When the first internet-enabled software on mobile devices was released, awareness of what exactly happened with the data was very scarce. Many companies have had scandals over the years where sensitive data was sold to companies for commercial purposes (Madrigal, 2018). That is why most of the current data privacy laws have begun to establish obligations of transparency and consent. These laws state that the user must be notified what exact data is collected and for what purpose it is used. In addition to transparency, the laws require the consent of the user to collect and use their personal information. This problem can simply be circumvented by directly notifying the user if they conform to the data collection policies of the company and decide to what degree data collection and use are acceptable. While not being a major challenge, it is a significant roadblock when it comes to the analysis of data, which can result in data bias and skew research outcomes (Hintze, 2019a).

3.10.2 Right of Access and Deletion

Other principles, which are common in privacy laws, are the right of a user to access personal information and the ability to request data deletion. The right to access enables a user to learn not only what kind of information the organization collects but also to know specifically what personal information the organization currently has. The right of access can be handled by a verification and

search mechanism. While it does not directly interfere with data processing, making its potential impact relatively low, it can still lead to additional costs and overhead. A user also has the ability to decide if the information about them should be deleted or not, which has a greater impact. The only way for a data scientist to deal with this challenge is to comply and remove the requested information. Tools should be applied to make the cleanup as simple as possible to avoid costs (Hintze, 2019a).

3.10.3 Data Anonymization and Minimization

Data Minimization refers to the principle that the least number of actions should be performed on data as possible. This means, for example, that there should not be collected more data than needed or kept longer as necessary. This law comes into direct conflict with big data, which relies on massive amounts of information to gain insights and can present points of friction. To comply with this law, data scientists should think of this principle as risk reduction and not use data carelessly. For example, data unusable for research or processing should be discarded and not needlessly stored (Hintze, 2019a). Data anonymization refers to the process of de-identification, in which personal information is modified to make it more challenging to re-identify an individual. Several privacy preservation techniques like K anonymity, randomization, or cryptographic techniques can be applied to “anonymize” personal information. In this scenario, even if data is compromised, the information is not easily available (Ram Mohan Rao et al., 2018).

4. Conclusion

Due to the increasing amount of data, privacy challenges and security issues are more critical than ever before. Data scientists take up the role of a data processor and are confronted with the threats of data privacy and security. Due to the sheer diversity and complexity of the cyber security field, there is no one-size-fits-all approach. Therefore, data scientists need to be well versed in the basics of security, like infrastructure, but also be proficient in data science-specific security aspects, like choosing machine learning models with security in mind. Many common security issues can be circumvented by adhering to best practices and employing cyber security tools like NIDS. In data science, specific problems like data protection, more sophisticated methods must be employed directly by the data scientist, like encryption. On the other hand, privacy challenges are dictated by the laws of the specific region and create points of friction with the data science field. Data protection laws are often pervasive and hard to keep track of. However, these laws can often be broken down into general principles that data scientists need to adhere to. While principles like transparency and consent are minor obstacles, data anonymization and minimization need additional methods to be effective. And while privacy and security are an increasing challenge for data science, data science also helps to solve problems in the security and privacy fields through the development of new techniques and methods.

V. Bibliography

Al-Sayid, N. A., & Aldlaeen, D. (2013). Database security threats: A survey study. In *2013 5th International Conference on Computer Science and Information Technology, Computer Science and Information Technology (CSIT), 2013 5th International Conference on*. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.6588759&site=eds-live

Altman, M., Wood, A., O'Brien, D. R., & Gasser, U. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law*, *8*(1), 29–51. https://doi.org/10.1093/idpl/ipx027

Andress, J. (2011). *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. Saint Louis: William Andrew. Retrieved from https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=713616

Biscontini, T. (2019). Computer Security. *Salem Press Encyclopedia of Science*. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=ers&AN=87321231&site=eds-live

Brown, B. (2017). How to make Fully Homomorphic Encryption "practical and usable&quot. Retrieved from https://www.networkworld.com/article/3196121/how-to-make-fully-homomorphic-encryption-practical-and-usable.html

Comm/dg/unit (2017). Data protection. Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection_en

Comm/dg/unit (2018). What is a data controller or a data processor? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en

Cylance (2021a, December 1). Retrieved from https://www.blackberry.com/us/en/products/unified-endpoint-security/cylance-is-now-blackberry

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73. https://doi.org/10.1145/2500499

Hintze, M. (2019a). Science and Privacy: Data Protection Laws and Their Impact on Research. *Washington Journal of Law, Technology & Arts*, *14*(2), 103. Retrieved from https://digitalcommons.law.uw.edu/wjlta/vol14/iss2/3

Hintze, M. (2019b). Science and Privacy: Data Protection Laws and Their Impact on Research. *Washington Journal of Law, Technology & Arts*, *14*(2), 103. Retrieved from https://digitalcommons.law.uw.edu/wjlta/vol14/iss2/3

Hu, W., Chang, C., Sengupta, A., Bhunia, S., Kastner, R., & Li, H. (2021). An Overview of Hardware Security and Trust: Threats, Countermeasures, and Design Tools. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions On, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, *40*(6), 1010-1010-1038. https://doi.org/10.1109/TCAD.2020.3047976

Irion, K. (2012). Government Cloud Computing and National Data Sovereignty. *Policy & Internet*, *4*(3-4), 40–71. https://doi.org/10.1002/poi3.10

IU Learn (2021b, November 30). Retrieved from https://learn.iu.org/courses/DLBCSEBI01/books/DLBCSEBI01/version/001-2021-0406_2021-04-06T09-14-25-817Z/CYCLE_b1291231850113fb41c58fce273e5ca6

Lundin, L. L. (2019). Information security. *Salem Press Encyclopedia*. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=ers&AN=89677573&site=eds-live

Madrigal, A. C. (2018, December 19). Facebook Didn't Sell Your Data; It Gave It Away. *The Atlantic*. Retrieved from https://www.theatlantic.com/technology/archive/2018/12/facebooks-failures-and-also-its-problems-leaking-data/578599/

Matherly, J. (2015, July 18). It's the Data, Stupid! Retrieved from https://blog.shodan.io/its-the-data-stupid/

Matusitz, J. (2008). Cyberterrorism: Postmodern State of Chaos. *Information Security Journal: A Global Perspective*, *17*(4), 179-179-187. https://doi.org/10.1080/19393550802397033

Mendyk-Krajewska, T., & Mazur, Z. (2010). Problem of network security threats. In T. Pardela (Ed.), *3rd Conference on Human System Interactions (HSI), 2010: 13 - 15 May 2010, Rzeszów, Poland ; conference proceedings* (pp. 436–443). Piscataway, NJ: IEEE. https://doi.org/10.1109/HSI.2010.5514533

Pauli, D. (2017, January 9). MongoDB ransom attacks soar, body count hits 27,000 in hours. *The Register*. Retrieved from https://www.theregister.com/2017/01/09/mongodb/

Prasanna, B. T., & Akki, C. B. (2015, May 13). *A Comparative Study of Homomorphic and Searchable Encryption Schemes for Cloud Computing*. Retrieved from https://arxiv.org/pdf/1505.03263

Purdy, Elizabeth Rholetter, PhD (2021). Cyberterrorism. *Salem Press Encyclopedia*. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=ers&AN=89677539&site=eds-live

Ram Mohan Rao, P., Murali Krishna, S., & Siva Kumar, A. P. (2018). Privacy preservation techniques in big data analytics: A survey. *Journal of Big Data*, *5*(1), 1–12. https://doi.org/10.1186/s40537-018-0141-8

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data*, *7*(1), 1–29. https://doi.org/10.1186/s40537-020-00318-5

Schatz, D., Bashroush, R., & Wall, J. (2017). Towards a More Representative Definition of Cyber Security. *The Journal of Digital Forensics, Security and Law*, *12*(2). https://doi.org/10.15394/jdfsl.2017.1476

Schreier, J. (2011, May 23). Sony Estimates $171 Million Loss From PSN Hack. *WIRED*. Retrieved from https://www.wired.com/2011/05/sony-psn-hack-losses/

Sen, A. A. A., & Basahel, A. M. (2019). A Comparative Study between Security and Privacy. In *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), Computing for Sustainable Global Development (INDIACom), 2019 6th International Conference on*. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.8991376&site=eds-live

SNIA (2021c, December 1). Retrieved from https://www.snia.org/education/what-is-data-privacy

Statista (2021, November 30). Total data volume worldwide 2010-2025 | Statista. Retrieved from https://www.statista.com/statistics/871513/worldwide-data-created/

Stewart, J. M. (2012). *Cissp: Certified Information Systems Security Professional study guide* (6th ed.). *Sybex serious skills*. Indianapolis, Ind.: J. Wiley & Sons. Retrieved from https://learning.oreilly.com/library/view/-/9781118332108/?ar

Taylor, D. (2018). Why Linux is better than Windows or macOS for security. Retrieved from https://www.computerworld.com/article/3252823/why-linux-is-better-than-windows-or-macos-for-security.html

Tellenbach, B., Rennhard, M., & Schweizer, R. (2019). Security of Data Science and Data Science for Security. In M. Braschler (Ed.), *Applied Data Science: Lessons Learned for the Data-Driven Business* (pp. 265–288). Cham: Springer International Publishing AG. https://doi.org/10.1007/978-3-030-11821-1_15

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016, September 9). *Stealing Machine Learning Models via Prediction APIs*. Retrieved from https://arxiv.org/pdf/1609.02943

What does IT infrastructure mean? (2021d, December 13). Retrieved from https://www.redhat.com/de/topics/cloud-computing/what-is-it-infrastructure

What is IT Infrastructure? (2021e, December 15). Retrieved from https://www.ecpi.edu/blog/what-is-it-infrastructure

Wheelus, C., & Zhu, X. (2020). IoT Network Security: Threats, Risks, and a Data-Driven Defense Framework. *IoT*, *1*(2), 259–285. https://doi.org/10.3390/iot1020016

Wieringa, J., Kannan, P. K., Ma, X., Reutterer, T., Risselada, H., & Skiera, B. (2021). Data analytics in a privacy-concerned world. *Journal of Business Research*, *122*, 915–925. https://doi.org/10.1016/j.jbusres.2019.05.005

Xue, M., Yuan, C., Wu, H., Zhang, Y., & Liu, W. (2020). Machine Learning Security: Threats, Countermeasures, and Evaluations. *IEEE Access*, *8*, 74720–74742. https://doi.org/10.1109/ACCESS.2020.2987435