Vilnius University Faculty of Law
Department of Private Law

Radvilė Kaladinskienė,
II study year, LL.M International and EU Law programme Student

Master Thesis

Protection of genetic data under EU General Data Protection Regulation

Supervisor: assist. prof. dr. Julius Zaleskis

Reviewer: assist. prof. dr. Johanas Baltrimas

Vilnius
2023

ABSTRACT AND KEY WORDS

This master thesis analyses the EU General Data Protection Regulation and its application in practice as regards to the protection of genetic data. It also highlights the problematic aspects of the application of the EU General Data Protection Regulation to genetic data in relation to the aspect of protection it enshrines.

Key words: genetic data, data protection, data protection law, genetic code, GDPR

CONTENT

MAIN ABBREVIATIONS

DNA	Deoxyribonucleic acid
ECtHR	European Court of Human Rights
EDPB	European Data Protection Board
EU	European Union
GDPR	General Data Protection Regulation
GDS	Genomic Data Sharing
GINA	Genetic Information Nondiscrimination Act
OECD	Organization for European Cooperation and Development
PIMS	Privacy Information Management System
RNA	Ribonucleic acid
UK	United Kingdom
WP	ARTICLE 29 Data Protection Working Party

INTRODUCTION

Relevance of the topic. The protection of genetic data is an increasingly important topic in the world of health and technology. Our genetic information is valuable information about the personal health and background of data subjects and should be protected from unauthorised access and use. One of the main reasons for the importance of protecting genetic data is the sensitivity of the information it contains. Our genes can reveal sensitive information about our medical history, our predisposition to certain diseases and even our ancestry. This information, if accessed by unauthorised persons or entities, could be used against us in a variety of ways. Ensuring the protection of genetic information is directly linked to ensuring the fundamental rights and freedoms of data subjects (Costello, 2022) and the protection of genetic data is therefore a crucial aspect of individual privacy in today's world of rapidly advancing technologies (Bieker F. *et al.* 2016). With the availability of whole genome[1] sequencing and the collection of large amounts of genetic information, there is an increasing need for appropriate safeguards to protect individuals from the possible misuse or exploitation of their genetic information.

In this context, it is important to underline that Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and replacing Directive 95/46/EC (the General Data Protection Regulation). The Regulation became applicable in the EU on 25 May 2018. The GDPR was created to give individuals more control over their personal data and to standardize data protection laws across the European Union. Despite the clear added value and the ambition to harmonise data protection across the EU, a number of problems arise in practice in relation to the rapidly evolving field of genetic science and its main outcome - the vast amount of gathering genetic data (Quinn *et al.* 2018). Ensuring and reconciling the protection of individuals' personal data with rapidly evolving technologies is therefore an increasing challenge. In order to achieve full compatibility and compliance with the law, science and regulation should act in synergy and should not be too far apart from each other.

[1] The genome is the entire set of DNA instructions found in a cell. For more information: https://www.genome.gov/genetics-glossary/Genome

Aim of the thesis. The main aim of the Master's thesis is to analyse the provisions of the GDPR and their applicability in the context of ensuring the protection of genetic data, as well as to highlight the issues related to the application of these provisions in the context of the processing of genetic data.

Objectives of the thesis. To achieve the aim of the Master's thesis, the following objectives are set:

1. To explain the concept of genetic data and the main aspects related to technological advances in genetics.
2. To examine the definition and characteristics of genetic data as personal data.
3. To analyse the basic principles of the GDPR for the processing of personal data and their applicability in the context of the processing and protection of genetic data.
4. To highlight the significance of genetic data as a special category of personal data.
5. Analyse the safeguards applicable to the processing of genetic data and the purpose of ensuring the rights of data subjects.
6. To highlight the main problems related to the protection of genetic data in practice.

Object of the thesis. The object of this Master's thesis is determined by the aim and objectives set for the analysis of the topic. The object of the study consists mainly of the provisions of the GDPR and their application and how the protection of genetic data is ensured. The work first analyses the concept of genetic data and the impact of technology in the context of the protection of genetic data, followed by an analysis of the definition of genetic data in the context of the definition of personal data and the characteristic features. Next, it focuses on the basic principles of personal data processing enshrined in the GDPR and their applicability in practice in the context of the processing and protection of genetic data. Attention is then given to highlighting the significance of genetic data as a special category of personal data. An analysis of the safeguards in place for the processing of genetic data is then presented as a basis for the fundamental rights of data subjects. Finally, the object of the work consists of an analysis and overview of the main issues related to the protection of genetic data.

Research methods. The following research methods are used in this Master's thesis:

1. *Linguistic* – this method is used to analyse the concept of the definition of genetic data in the GDPR. It is also used to reveal the meaning and content of other provisions of the GDPR in the context of the topic under consideration.
2. *Historical* – this method is used to analyse the evolution of the concept of genetic

data. It is also used to review the evolution and changes in the relevant sources in relation to technological progress.

3. *Comparative* – this method is used to compare the provisions of some international instruments and legislation in the context of genetic data protection. This method is also used to compare the views of certain authors in the field of genetic data protection.
4. *Systematic* – this approach is applied in the interpretation of the GDPR provisions, as well as by linking the GDPR articles to the provisions of the GDPR Preamble, thus systematically seeking to reveal their interconnection.
5. *Teleological* – this approach is used in the thesis to provide an overview of the main objectives of the GDPR in terms of the protection of genetic data.
6. *Logical* – this method is used to summarise the parts of this thesis and to draw final conclusions.

Originality of the thesis. The authors who have addressed the issue of the protection of genetic data in scientific doctrine are indeed many. However, the analysis of this topic has been carried out mainly by foreign authors, whose researches, scientific articles etc. do not deal with the issue of the protection of genetic data in a separate manner, but rather with the analysis of the individual provisions applicable to the processing of genetic data, usually without distinguishing the practical aspects. Issues relating to the application of the provisions of the General Data Protection Regulation relating to the processing of genetic data and the evaluation of these provisions in the context of enhanced legal protection for the sharing of genetic data for research purposes have been addressed in an article “Processing of Genetic Data under GDPR: Unresolved Conflict of Interests” by the authors Sukhorolskyi P., V. Hutsaliuk V. (2020). Also in a scientific article “Big genetic data and its big data protection challenges“ by the authors Quinn P., Quinn L. (2018). In the article "Genetic Date and the Right to Privacy: Towards a Relational Theory of Privacy?" (2022), Costello R.A raised the issue that genetic data is not exclusively about the data subject and their right to privacy (individualistic approach), but also about the interests of the group and their privacy (relational effect). Therefore, in this case, any violation of the data subject's privacy rights affects the privacy of the persons to whom they are genetically related or with whom they have social or communication ties.

Meanwhile, in Lithuania, a somewhat more detailed analysis of the legal, medical, and ethical aspects of the protection of human genetic data has been conducted by Petkevičienė V.,

Pakutinskas P., Bitė V. (2020). „Asmens duomenų tvarkymo iššūkiai COVID-19 pandemijos metu“ also Lazauskienė R., Tamulionienė, D. (2020) „Asmens duomenų tvarkymo ypatumai nuotoliniu būdu teikiant paslaugas sveikatos priežiūros srityje“. However, these articles do not address the issue of genetic data in a separately manner.

A master's thesis on a similar topic related to genetic data is written by a student of the Faculty of Law of Vilnius University Dvarionaitė V. “Genetinių duomenų reguliavimas pagal ES Bendrąjį duomenų apsaugos reglamentą“. However, the work in question analyses how genetic data is regulated, mainly based on the provisions of the GDPR, without analysing in detail how genetic data is protected and what issues exist in practice. There is also a Master's thesis “Genetinių tyrimų etika ir teisinis reguliavimas“ written by a student Lekarauskaitė D. of the Faculty of Law of Mykolas Romeris University. However, the analysis of this Master's thesis focuses on the ethical aspects of genetic research without analysing the protection of genetic information.

This Master's thesis is different from the above-mentioned thesis because it systematises the specific issues related to the protection of the GDPR provisions for the processing of genetic data, examining foreign literature, research, key insights of the authors, and relevant scientific articles.

Main sources. The General Data Protection Regulation is the main source for this thesis. In analysing and interpreting the concept of the provisions of the GDPR, the guidelines of the EU Article 29 Data Protection Working Party established in 1995 under Article 29 of Directive 95/46/EC have also been extensively referred to as authoritative soft law sources on different aspects of the application of the GDPR. The thesis also draws mainly on existing foreign and Lithuanian academic doctrine and the jurisprudence of the European Court of Human Rights.

1. CONCEPT OF GENETIC DATA

Many would probably agree that data protection is one of the most debated topics these days. Increasing technological progress poses ever greater legal challenges in terms of how to adapt to changing circumstances and how to protect individuals' rights and freedoms. In this context, it is important to analyse the concept of genetic data considering the emphasis on individual rights and freedoms.

Genetics is a branch of biological science that studies the information encoded by genetic material (e.g., genes), the laws of heredity and variability, genetic data is information about human genes and the hereditary genetic material contained in them. Heritability, determined by the information encoded in the genetic material, is very important, but not the only factor influencing the phenotype of an organism (e.g., appearance, behaviour). Environmental influences play a role. How decisive heredity is can be assessed by comparing monozygotic ("identical") twins who have nearly identical DNA sequences but different phenotypes (Sharma *et al*. 2016).

Genetics is a branch of science that studies epigenetic modifications of the genome, methods of determining gene states, their transmission in the form of cells to other generations, and mechanisms of regulation of information inherited from generation to generation (without changing the sequence of nucleotides) in response to an external factor (Rebekah P.K. 2017).

Genetic data is the DNA molecular tagging system responsible for human gene regulation processes protected by law from their disclosure to third parties based on the GDPR. The specific nature of genetics makes research difficult. A blood sample is used for testing, which is used to determine a person's genotype, but blood DNA is not sufficient to determine most genetic markers, as it cannot identify the risk of chronic or oncological diseases (Tattersfield K. 2017).

Evolution of genetic data. The pioneer of genetics is Gregor Mendel, an Austrian scientist. In 1865 the scientist determined how the traits of heredity are transmitted from generation to generation. Dr. R. A. Waterland, a professor of pediatrics, nutrition and molecular genetics at Baylor, and a team of researchers identified the types of genome for which a blood sample can be informative. This allows scientists to study the genetic causes of diseases. To do this, they focused on the most stable form of genetic regulation, DNA

methylation. This process is the attachment of methyl groups to the DNA molecules of the embryo, which can affect human health. The activity of many genes is not constant: they are turned on (expressed) and turned off (repressed) depending on the influence of external factors. This change in gene activity, which does not affect the primary structure of DNA, but affects the manifestation of certain characteristics and traits, has become the subject of genetics research (Niu Z., *et al.*, 2017).

Genetics is a relatively young branch of science. The term genetics was first used by the English scientist C. Waddington in 1942. By studying many regularities, the scientist concluded that the functions of a living organism are determined not only by information encoded in genes, but in many ways, it serves as a response to environmental signals. The way in which certain genes are genetically determined to turn on and off has become one of the most important discoveries of our time, for which American scientists were awarded the Nobel Prize in 2006. Another scientist - G. Mendel noted that human genetics is based on the fact that changes in phenotypic traits are based on DNA mutations, that is, mechanical - random or induced - changes in the structure of hereditary information. Genetics is based on variations in the norm represented by modifications. Each of the genomic disorders is no less important than the genetic disorders and acts as the genetic equivalent of a genetic mutation. The gene is the main carrier of hereditary information, genetic mechanisms can control the work of only certain genes according to the available material. The genome is a control mechanism for the implementation of genetic information, which is carried out by modifying individual nucleotides. It should be noted that not all genes in a person are functional, some genes are active in one cell, inactive in another, and vice versa. There are certain regulatory elements that control gene activity. According to modern concepts, these elements include DNA methylation, histone modifications, acetylation, phosphorylation, glycosylation, various microRNAs and other structures/processes that "regulate" the human genome (Meler E, *et al.*, 2020).

Humans have 60,554 genes, which are fragments of the gene sequence that encode information about the amino acid sequence of a polypeptide or protein (which is provided in the form of RNA) or provide information about non-coding RNAs. Non-coding RNAs can be divided into ribosomal RNA (rRNA), transport RNA (tRNA), there are also long non-coding RNAs and small RNAs (snRNA, snoRNA, miRNA, etc.). Gene products (such as long non-coding RNAs or proteins) can influence certain characteristics of an organism (hair, eye colour,

height, etc.) and can partially influence the phenotype of an organism. Thus, a gene is a means of inheritance, a carrier of hereditary information. During reproduction, this genetic information is passed on to the next generation. Another type of genes - regulatory genes - interacts with regulatory molecules (non-coding RNAs, proteins) and controls the expression of genetic information (Coppedè F. 2019).

Human chromosomes contain about 25,000 genes that are located in a certain linear sequence and occupy a defined place (locus) on the chromosome. A gene consists of coding (exons) and non-coding parts (introns). In the human genome, coding DNA makes up only 1,5%. Introns usually separate exons, but there are human genes that do not have introns. During protein synthesis, according to the DNA sequence, RNA polymerase synthesizes information RNA (iRNA) - the so-called transcription takes place. Some genes encode information about the synthesis of various proteins in the cell. The latter are responsible for a certain biological function, determine certain traits and regulate other genes. If allelic genes (genes coded in the same loci and determining the same trait or function) equally determine the trait, the individual will be homozygous for this trait, if different, heterozygous. In addition, allelic genes can be dominant or suppressive in relation to each other (Deepak S., *et al.,* 2017).

For example, the gene for brown eyes is dominant over the gene for blue eyes (the latter is called recessive). Certain nucleotides (the basic unit of DNA), whether adenine (A), thymine (T), cytosine (C), or guanine (G), can be arranged in specific ways to form the FOXP1 gene, which in turn codes for a specific protein. A missense mutation occurs when a single DNA nucleotide is changed in a gene, sometimes the change is small and does not affect the protein the DNA encodes. Other changes result in an amino acid change in the protein that the gene encodes, which can sometimes fundamentally alter the protein's function (Mehrabani S. Z. N. 2019).

As is clear, the development of genetic science and the application of increasingly sophisticated techniques associated with it have provided the basis for understanding what genes are and the information they encode, as well as what is meant by the term genetic data.

1.1. The impact of technological progress in the context of genetic data

The impact of technological process in the context of genetic data is manifested in the emergence of new technologies that systematize and store a person's genetic data. This data

must not only be collected with the help of modern technologies, but also protected from its disclosure with the help of the GDPR (Bieker F., *et al.,* 2016).

There is no doubt that scientific inventions in the field of medicine, including genetic testing and the use of biological samples for the detection of genetic diseases[2] and the development of medicines[3], bring significant benefits to human health. In this context, it is also important to highlight that the use of advanced technologies contributes to the implementation of criminal justice[4].

One of the key technologies that has led to an increase in genetic data is Next Generation Sequencing (NGS), which allows the rapid and cost-effective sequencing of large quantities of DNA (Bahr *et al.* 2015). This has led to the accumulation of large amounts of data on the genetic structure of individuals, populations and species.

One of the major uses of genetic data is in the field of personalized medicine, where it is used to tailor medical treatments to the specific genetic makeup of an individual (Verma M. (2012). For example, genetic testing can be used to identify individuals who are at an increased risk for certain diseases, such as cancer, and to develop tailored treatment plans that take into account their unique genetic profile.

Another area where genetic data has revolutionised agricultural biotechnology (Montagu M.V. (2020) By analysing the genetic make-up of crops, scientists can identify traits that benefit agriculture, such as disease resistance or higher yields. This information can be used to develop genetically modified crops that are more productive and resilient.

The availability of genetic data has also facilitated the development of new methods for the diagnosis and treatment of genetic disorders (Goh. G. 2012). For example, whole exome sequencing, which sequences protein-coding regions of the genome, has been used to identify the genetic basis of rare diseases, leading to more accurate disease diagnosis and the development of targeted treatments.

[2] 1983 DNA polymorphism was used to map the first genetic disease - Huntington's. For more information: https://www.genome.gov/25520322/online-education-kit-1983-first-disease-gene-mapped

[3] 1982 the Food and Drug Administration approved Humulin, the first biosynthetic human insulin product and the first approved medical product of any kind that derived from this technology. For more information: https://www.fda.gov/about-fda/fda-history-exhibits/100-years-insulin

[4] DNA forensics was first reported in 1984 by Sir Alec John Jeffreys, a British geneticist, who developed genetic fingerprinting and DNA profiling techniques that are now used worldwide in forensic science. For more information: https://le.ac.uk/dna-fingerprinting/biography

Despite its undeniable benefits for humanity, the processing of genetic data poses major challenges - how to ensure the security of such data (protection aspect) without infringing on individuals' fundamental rights and freedoms (privacy concern).

Genetic data protection. Genetic data have certain characteristics that make the processing of genetic data reasonably require special legal protection in accordance with the GDPR (Mehrabani S. 2019):

- although genetic information is unique and distinguishes a person from other persons, it can simultaneously reveal information about him and affect that person's blood relatives (biological family), including relatives of both subsequent and previous generations. Furthermore, genetic data can characterize a group of individuals (e.g., ethnic communities); genetic data can reveal paternity and family relationships.
- genetic information is often unknown to its owner and does not depend on the individual's will since genetic data is immutable.
- genetic data can be easily obtained or isolated from raw material, although sometimes this data can be of questionable quality.
- considering the progress in the field of research, in the future genetic data can reveal even more information and can be used for different purposes by an increasing number of users.

As shown, the analysis of genetic data can reveal many unique characteristics of an individual. Therefore, there is no doubt that the security of genetic data is at high risk when such data is processed with the help of advanced technologies. The protection of genetic data is linked to the privacy of individuals.

Privacy concern. In the legal doctrine, privacy is defined as ,,the interest of a person who can maintain his personal space, free from any external interference" or "legally protected interest", which includes: protection of individual, family and home life, physical and mental integrity of a person, state of health, genetic data, honour and reputation, communication (communication, correspondence), restriction of access to personal facts, prohibition to independently collect, accumulate, disclose not only confidential, but also any other information with which there is no need to introduce outsiders (Petkevičienė *et al*., 2020).

Privacy - as a whole of personal information, separated into separate parts: ,,information privacy: body privacy; genetic data privacy; privacy of communications, communications, correspondence, territorial privacy" (Štareikė *et al.*, 2018).

The content of a person's right to privacy consists of 4 independent and at the same time interrelated elements: (a) informational privacy – related to the processing of personal data and is called personal data protection, i.e., when the person himself can dispose of his personal data, know about the processing of his data, familiarize himself with his personal and genetic data, etc.; (b) physical privacy - this can be called the inviolability of the body. No medical and scientific research may be performed on a person without his consent; (c) communicative privacy is the inviolability of a person's correspondence, chat phones, telegraph, Internet, and other forms of communication; (d) territorial privacy is the inviolability of a person's apartment or territory. (Lazauskienė *et al.*, 2020).

The essence of personal genetic data protection is to protect people's rights. The main objective is to ensure that personal genetic data is processed in a way that ensures privacy and other related human rights. Each of the listed elements of a person's right to privacy reflects certain identification of each person's genetic data, activities performed, his location, but all these data must be confidential and encrypted, and provided only by the person himself or with his permission (Lazauskienė *et al.*, 2020).

A new technological breakthrough or scientific advance is probably no longer a surprise. Today's world is based on rapidly evolving technologies and adapting to these inventions. Despite the benefits of advanced technological applications in the field of genetics, the processing of genetic data poses increasing challenges in terms of data protection and ensuring the privacy of individuals. Therefore, despite the great value of technological advances, the aim must be to maximise the protection of such data, while ensuring the fundamental rights and freedoms of individuals.

1.2. The importance of genetic code analysis

The first part of the thesis, "The concept of genetic data", analyses the concept and evolution of genetic information. It provides an overview of what genetic data is. This part will seek to provide an analysis of the genetic code and its meaning in the context of genetic data.

The genetic code can be thought of as a system of rules for mapping the information needed to synthesise proteins onto a DNA or RNA molecule by nucleotide sequences.[5] The genetic code should therefore be understood as a precise and unchanging set of rules that produce the correct synthesis of a protein's amino acid sequence. Earlier in this paper, it was analysed that genetic information is encoded in DNA. Therefore, when analysing the importance of the genetic code in relation to genetic data, we should realise that the genetic code is only part of the total information contained in DNA.

The genetic code makes it easier to understand DNA because it provides information, a specific nucleotide sequence. Therefore, the information that is encoded in any genetic material is carried by the genetic code. Furthermore, DNA is the genetic code of life. Deciphering these codes helps us understand how living things function and what mutations cause disease.[6]

An analysis of the meaning of the genetic code itself for genetic data would suggest that the system of rules itself and the particular sequence of information it contains would not qualify as genetic data for the purposes of the Regulation. In particular, according to GDPR: *'genetic data' means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question;" (Article 4(13)* of the GDPR) Also: „*Genetic data should be defined as personal data relating to the inherited or acquired genetic characteristics of a natural person which result from the analysis of a biological sample from the natural person in question, in particular chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis, or from the analysis of another element enabling equivalent information to be obtained" (*Recital 34 of GDPR). As the definition's makes explicit, genetic data are considered to be personal data that provide unique information about a person's physical or health condition and which is derived from DNA and RNA analysis. The genetic code itself does not provide such information. As noted above, the genetic code is the set of rules that living cells use to convert the information encoded in their genetic material into proteins. Although the genetic code itself does not directly reveal genetic information, only analysis of

[5] For more information: https://www.genome.gov/genetics-glossary/Genetic-Code

[6] For more information: https://esti.my/2021/01/08/reading-your-3-billion-genetic-code-how-sequencing-technology-changes-the-landscape-of-biological-science/#:~:text=DNA%20is%20the%20genetic%20codes%20of%20life.%20Deciphering,that%20involve%20in%20disease%20protection%20and%20immune%20system.

it can reveal some unique information. For instance, about diseases or mutations. A mutation should be understood as a change in the DNA sequence of an organism.[7] For example, if the genetic code is distorted, i.e., one or more nucleotides are removed or inserted, this will lead to an incorrect synthesis of the amino acid sequence of a protein.

To summarise the significance of the genetic code analysis, it should be noted that the protection of the genetic code is not directly incorporated into the GDPR. However, as has been analysed, the genetic code itself is only part of the total information contained in DNA. In this paper analysis of the definition of genetic data, we have found that genetic data is obtained by analysis of a biological sample of the natural person concerned, chromosomes, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), or any other element that provides equivalent information. Therefore, it is considered that the definition of the genetic code falls below the analysis of DNA and RNA, since it is part of that information, and its analysis may reveal some information related to genetic data.

[7] For more information: https://www.genome.gov/genetics-glossary/Mutation

2. EU GENERAL DATA PROTECTION REGULATION AS THE SOURCE FOR THE PROTECTION OF GENETIC DATA

The General Data Protection Regulation provides data protection rules for EU institutions to ensure that the data protection standards applicable to EU institutions and bodies comply with the data protection standards provided for in the GDPR. These rules reflect the same values, ensuring that EU citizens can enjoy the same, more protected rights when dealing with EU institutions that they enjoy when dealing with other companies, organisations, public bodies, personal genetic data under the GDPR.[8]

The General Data Protection Regulation is a 2016 Regulation of the European Parliament and of the Council adopted on 27 April 2016 (EU) 2016/679 "*On the protection of natural persons when processing personal data" in accordance with this regulation"*. This regulation encourages private and public institutions to accept responsibility for having the right of access and using the data in such a way that they cannot harm the persons to whom the data belongs, establishing the principle of responsibility for genetic data protection. In case of loss of personal genetic data, healthcare institutions protecting this data, must immediately notify the affected persons about the incident. After the GDPR regulation entered into force in Lithuania at the end of 2018, the State Data Protection Inspectorate published recommendations for the healthcare sector, ensuring compliance with the requirements in Lithuania (Asmens Duomenų Apsaugos Gairės Sveikatos...2019). The guidelines prepared by the inspectorate specify technical and organizational data security requirements. However, these requirements are considered sufficient only for o healthcare institutions that do not face high risks related to threats to the rights and freedoms of natural persons. Therefore, many healthcare institutions must implement not only these minimum requirements but must also take additional steps in order to properly protect the processed personal genetic data and ensure the required level of security, as explained by the General Data Protection Regulation (Štareikė *et al.*, 2018).

In conclusion, the entry into force of the General Data Protection Regulation is a major step forward in the effort to unify and strengthen the protection of individuals' data including genetic data on a global scale, thus safeguarding the fundamental rights and freedoms of individuals.

[8] For more information: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

2.1. Definition of genetic data under the General Data Protection Regulation

In the context of a comprehensive analysis of the definition of genetic data in the GDPR, it is important to note that the definition of genetic data is enshrined in a number of international instruments:

- Recommendation No R (97)5 of the Committee of Ministers of the Council of Europe of 13 February 1997 on the protection of medical data defines the concept of genetic data in paragraph 1, stating that it is *"any data relating to the inheritance of personal characteristics or to the pattern of inheritance of such characteristics in a given group of relevant individuals".*
- Article 2 (g) of the 2 August 2002 law of Luxembourg on the protection of persons with regard to the processing of personal data provided that *"any data concerning the hereditary characteristics of an individual or group of related individuals"*. *(*This act was replaced by the General Data Protection Regulation from 25 May 2018).
- Article 2 (i) of the International Declaration on Human Genetic data, UNESCO stating that it is *"Information about heritable characteristics of individuals obtained by analysis of nucleic acids or by other scientific analysis".*

It can be assumed from a further analysis of the definition of genetic data in the GDPR that all the above-mentioned definitions in international instruments have been transposed into the concept of genetic data in the GDPR. An analysis of this definition under the GDPR follows.

We can distinguish the concept of genetic data in two parts of the GDPR: paragraph 34 of the preamble to the Regulation and Article 4(13).

According to Recital 34 of GDPR: *„genetic data should be defined as personal data relating to the inherited or acquired genetic characteristics of a natural person which result from the analysis of a biological sample from the natural person in question, in particular chromosomal, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) analysis, or from the analysis of another element enabling equivalent information to be obtained".* The analysis of this broad definition reveals certain characteristics that are applicable to the definition. In particular, genetic data refers to data relating to a natural person; data relating to (1) the inherited genetic characteristics of a person[9] or (2) the acquired genetic characteristics of a

[9] For example, blood type inheritance. For more information:

person[10]; data obtained by analysis of a biological sample of the natural person concerned, in particular chromosomes and deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), or other elements from which equivalent information can be obtained. It can be considered that, in the context of the protection of the Regulation, only information from the subject concerned that meets the listed characteristics and has been collected by appropriate means constitutes genetic data.

However, it remains unclear what falls under "other elements enabling equivalent information to be obtained.". As stated above in recital 34, obtaining genetic information includes chromosomal, DNA or RNA analysis, or any other type of analysis that provides equivalent information. However, it is important to stress that not all genetic information is genetic data. First, there is always the question of whether genetic information is personal data. For example, a genetic sample itself is not personal data because we cannot read anything from it. Therefore, first, we need to carry out some analysis on the sample to be able to obtain certain data. As the very notion of genetic data implies, genetic analysis data is only personal data if it can be linked to an identifiable person. Despite this DNA and RNA analysis techniques do not raise major questions in this case, but it is not clear what falls within the definition of "other elements". Nevertheless, the methods used to analyse DNA and RNA genetic information do not pose major problems in this case, but it remains unclear what falls under the definition of "other elements". However, it can be assumed that the legislator uses the phrase "*or from the analysis of another element providing equivalent information*" to ensure that the law will adapt to new technologies and other ways of obtaining genetic information that may emerge in the future.

Regarding Article 4(13) of the Regulation: *'genetic data' means personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question".* Assessing both definitions - recital 34 of the GDPR and Article 4(13) of the GDPR – it can be observed that the legislator has essentially reiterated the same approach regarding the

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3595629/

[10] Acquired characteristics, by definition, are characteristics that are gained by an organism after birth as a result of external influences or the organism's own activities which change its structure or function and cannot be inherited. For more information: https://www.newworldencyclopedia.org/entry/Acquired_characteristics

definition of genetic data. Despite the similarities between the definitions, some differences can be identified.

An analysis of the definition in Article 4(13) of the GDPR indicates that it is more general, in this case, recital 34 is complementary to the concept of genetic data in an expansive sense (Kuner *et al.*, 2020). It could also be highlighted that the definition explicitly states that genetic data are obtained by analysing a biological sample of a natural person. On the contrary, recital 34 provides that genetic data may also be obtained *"from the analysis of another element enabling equivalent information to be obtained"*. It has already been analysed earlier in this thesis that the legislator has not limited itself to specific ways of obtaining genetic information but has looked to the future and to rapidly evolving technologies and has envisaged other ways. The reason for this decision is likely to be to keep the legal framework abreast of emerging technologies. It is also important to underline that when examining the concept of genetic data as set out in Article 4(13) of the GDPR, it can be considered that it is not necessarily the analysis of a biological sample that results in information that will be considered as genetic data in the context of GDPR.

As explicitly stated in Article 4(13) of the GDPR, the definition specifies that such information must include the genetic characteristics of a particular person and provide ***unique information*** about that person's ***physiology*** or ***health***. As can be seen clearly, the uniqueness of a person is distinguished, which can be seen as a characteristic, trait or attribute that is unique to that person in terms of physiology or health. Therefore, data which do not clearly distinguish a person from other persons in terms of physiological characteristics or state of health do not normally fall within the definition set out in Article 4(13), even though they may be obtained through the analysis of his or her biological samples (Kuner *et al.*, 2020).

An analysis of these two definitions in the GDPR suggests that the two concepts are complementary. Despite the similarities, the concept of Recital 34 is more adapted to future technologies and the need for regulation to keep pace with them.

2.2. Application of the Regulation

There is probably no doubt that technological progress has made life easier for all humanity. However, the rapid development of technology has left legal regulation far behind. Earlier in this thesis, it was mentioned that advances in technology have raised one of the major issue -

the right to privacy and how to protect it. In this respect, the legislator had to find a solution to keep pace with technological progress.

Despite the fact that, prior to the adoption of the Regulation, international instruments can be seen to protect genetic data by prohibiting any discrimination in the processing of such data: for instance, Article 21 of the EU Charter of Fundamental Rights enshrines the prohibition of "*any discrimination based on genetic characteristics*", and this prohibition is also enshrined in Article 11 of the Council of Europe Convention on Human Rights and Biomedicine, and in Article 6 of the UNESCO Universal Declaration on the Human Genome and Human Rights (ARTICLE 29 Data Protection WP… 2004). However, as can be seen, the prohibition of non-discrimination is generally enshrined in the above-mentioned instruments. Furthermore, given that we are living in an era of digitalisation, we should take a slightly broader view, i.e., not limit ourselves to prohibitions without taking appropriate measures to achieve the objectives pursued by such prohibitions. It is therefore not enough for legislators to established restrictions merely declaratively in law. It should be clear to the public what such restrictions cover, and the legislator's objective should be to seek to protect the interests of such a public in a comprehensive manner, by providing for specific conditions. Any declaratory provision is vaguer and leaves room for interpretation. There is therefore a risk that the interests of individuals may be prejudiced. As the WP noted, the main effectiveness of the prohibitions lies in the strict rules limiting the use of genetic data. And the protection of the right to health depends on ensuring that no genetic data is made known to third parties who could use it to discriminate and/or stigmatise the data subject (ARTICLE 29 Data Protection WP… 2004). The GDPR aims to protect the interests of the individual. It was adopted to protect the fundamental rights and freedoms of natural persons, in particular their right to the protection of personal data (Article 1(2) GDPR).

First chapter of this thesis analysed the impact of technology on the protection of genetic data. The analysis has shown that the accelerating pace of technological advances has a direct impact on the issue of sharing such data (an analysis of this issue will be presented in third chapter). The analysis of genetic data has undoubtedly contributed to medical advances in the invention of medicines, the diagnosis of diseases, etc. However, the sharing of such data (bearing in mind that genetic data is a special category of personal data) directly can infringe fundamental human rights and freedoms. Therefore, it is noteworthy that the EU, to protect the privacy of the individual, and by adopting the GDPR accordingly, has strengthened the rules

on cross-border data sharing, with appropriate safeguards in the Regulation. The GDPR regulates the processing of personal data of all natural persons, not only by natural persons, but also by companies and institutions established in the European Economic Area. In addition, the GDPR extends its jurisdiction to controllers and processors located outside the European Union in certain circumstances (Art. 2 GDPR). The territorial scope of the GDPR is established in Article 3 of the Regulation and represents a significant evolution of data protection in the EU (European Data Protection Board Guidelines…2018) The territorial scope of Article 3 of the GDPR can be distinguished:

a) Article 3(1) of the GDPR states that the *"Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not*." (Application of the establishment criterion) (European Data Protection Board Guidelines…2018).
b) Article 3(2) of the GDPR states that *"this Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to: (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or (b) the monitoring of their behaviour as far as their behaviour takes place within the Union."* This paragraph provides for the circumstances in which the Regulation applies to a controller or processor not established in the Union, considering the processing activities carried out by them (Application of the targeting criterion) (European Data Protection Board Guidelines…2018).

Article 3(1) of the GDPR defines the establishment not only of the controller but also of the processor. Therefore, the processing of personal data by a processor in the EU may also be subject to EU law by the processor has an establishment in the EU (European Data Protection Board Guidelines…2018). As regards the processing of genetic data, and in line with the provisions of the Article, for the processing of genetic data to fall within the scope of the Article, the processor must be established in the EU. Thus, it can be argued that the main criterion for this part of the Article is that of establishment (jurisdiction).

The 'targeting criterion' referred to in Article 3(2) for data subjects in the Union may be applied to a controller or processor not established in the Union in relation to processing

activities involving two distinct and alternative activities, provided that those processing activities relate to data subjects in the Union. In addition to the fact that it only applies to processing carried out by a controller or processor not established in the Union, the purpose limitation criterion focuses on what the 'processing activities' are 'related to', which has to be assessed on a case-by-case basis (European Data Protection Board Guidelines…2018). As the European Data Protection Board points out, a controller or processor may be subject to the GDPR in relation to some of its processing activities but not to the GDPR in relation to other processing activities. The relevant processing activities in question are decisive for the territorial application of the GDPR under Article 3(2).

It is also worth noting that paragraph 24 of the preamble to the GDPR states that the territorial scope aspect applies when data subjects may be monitored and profiled as a result of their online activities. In this context, various health apps (MyFitnessPal; Clue, etc.) process information about a person's habits and behaviours, some insights into physical and mental health, all of which can provide insights based on the data gathered.

Although genetic data do not fall within the definition of processing in this case. It could be assumed that the increasing advances in technology suggest that there may be a future need for such processing.

Article 2 of the GDPR sets out the substantive material scope of the GDPR. The first paragraph of that Article establishes to what the Regulation applies: the processing of personal data wholly or partly by automated means, and by means other than automated means, where the personal data form part of, or are intended to form part of, a file system. Article 2 does not distinguish between the public and the private sector and therefore applies to both sectors. The scope of application formulated in the first paragraph is limited by the second paragraph, which excludes certain data processing activities from the scope of the GDPR. Therefore, processing for purely personal or household purposes is excluded from the scope (Kuner *et al.* 2020). The other exemptions referred to in paragraph 2 relate to policy areas where the EU has no or limited competence or where specific Union rules apply. This includes the processing of personal data by competent authorities in the field of law enforcement (Kuner *et al.* 2020).

It is also important to mention the personal scope when analysing the scope of the GDPR. In particular, the GDPR applies both to the data subject whose data are processed (as defined in Article 4(2) of the GDPR) and to the controller who processes the data subject's data (as defined in Article 4(7) of the GDPR). In this case, it is also important to keep in mind the

processor who processes the data on behalf of the controller (the concept is set out in Article 4(8) of the GDPR).

It could be underlining that one of the main aims of the GDPR is to give individuals greater control over their personal data and to provide them with increased transparency about how their data is used. The GDPR defines personal data as any information relating to an identified or identifiable natural person. The GDPR sets out a number of principles that controllers and processors must adhere to when processing personal data. These principles include the requirement that personal data must be processed lawfully, fairly, and transparently; that personal data must be collected for specified, explicit, and legitimate purposes; and that personal data must be adequate, relevant, and limited to what is necessary for the purposes for which it is processed (an analysis of the principles is set out in Section 2.3 of this thesis). One of the key aspects of the GDPR is the requirement for controllers and processors to demonstrate compliance with the regulation. This includes the need for controllers and processors to implement appropriate technical and organizational measures to protect personal data, and to carry out risk assessments to identify any potential risks to the rights and freedoms of individuals (an analysis of the technical and organisational measures is given in Section 2.4.1 of this thesis). The Regulation also gives individuals certain rights with respect to their personal data, including the right to access their data, the right to rectification of their data, the right to erasure of their data (also known as the "right to be forgotten"), and the right to object to the processing of their data. It could be arguable that GDPR is a comprehensive piece of legislation that provides individuals with greater control over their personal data and sets out strict requirements for the processing of that data by controllers and processors. Its aim is to enhance the protection of personal data and to provide individuals with increased transparency about how their data is used.

In conclusion, the scope of application established in the GDPR is one of the most important features of the GDPR protection, covering both EU and non-EU entities' activities related to the processing of genetic data. Therefore, the definition of the scope of the GDPR is adopted in an expansive manner in order to fully protect the rights and interests of data subjects.

2.3. Basic principles of the Regulation

The previous analysis in this thesis (to be more specific - analysis of the definition of genetic data in the GDPR) emphasises that genetic data are *personal data*. Therefore, on the basis that genetic data are personal data, their processing should be lawful, fair, and transparent in relation to the data subject (Article 5(1) GDPR). These requirements are also reflected in another part of the GDPR (Recital 39 of the GDPR).

The principle of lawfulness, fairness and transparency of processing is considered to be the overarching and broadest principle of data protection law (Zaleskis, 2019). Although the above principles are quite broad, it is considered important to specify them in order to clarify the implications of their observance in terms of genetic data protection.

Lawfulness of data processing

Compliance with the law is a key aspect of this principle. Applying all other principles and rules of data protection law implements the principle under analysis (Zaleskis, 2019). It is also important to underline that this principle is fundamental, i.e., it underpins the other requirements of the GDPR. The principle of lawfulness essentially implies compliance with both normative sources of data protection law and international sources. In respect to this principle, data must be processed in accordance with the requirements set out in the legislation (Zaleskis, 2019).

In the analysis of the principle of legality, it is important to underline that the implementation of this principle is also reflected in other articles of the GDPR. For example, article 6 of the GDPR provides grounds for lawful processing of personal data without consent, including but not limited to the condition where "*processing is necessary for the purposes of legitimate interest*" (Shabani *et al.,* 2017). A parallel approach is taken in the conditions for transfers of personal data to third countries under Article 49(1), where such transfers may take place without consent where "*necessary for the purposes of the compelling legitimate interests of the controller, which do not override the interests or the rights and freedoms of the data subject"* (Shabani *et al.,* 2017).

In the context of the principle of legality, it is also important to mention Article 9 of the GDPR. The article lays down specific rules on the legitimate grounds for processing special

categories of data[11], which means that when one of the special categories of data is processed, the rules of Article 9 on the permitted use of sensitive data apply. The WP has clarified that a controller processing special categories of data can never rely solely on the general grounds for processing currently provided for in Article 6 GDPR. These rules "*will not prevail but will always apply in conjunction*" with the rules for processing special categories of data (ARTICLE 29 Data Protection WP… 2014).

Given the complexity and sensitivity of genetic information, there is a high risk that it may be misused and/or reused for different purposes by the controller or third parties. The risk of re-use may arise, for example, from the use of genetic information that has already been extracted, or from additional analysis of the underlying material (e.g., a blood sample). The Regulation prohibits further processing of the data which would be incompatible with the purpose for which the data were collected. However, it provides for exceptions to the prohibition of further processing for historical, statistical, or scientific purposes, provided that Member States provide for appropriate safeguards (ARTICLE 29 Data Protection WP… 2004).

To conclude, the lawfulness of data processing is one of the fundamental principles of the GDPR. Given that genetic data are particularly sensitive in relation to individuals, controllers are obliged to properly identify the grounds for processing such data and not to process the data for purposes unrelated to them. Otherwise, failure to implement this basic principle will not ensure the protection of genetic information processed by data subjects and is likely to result in a breach of fundamental rights and freedoms of individuals.

Fairness of data processing

Fairness is a fundamental principle of the EU data protection framework (Clifford *et al.*, 2017). This principle is reflected both in Article 8(2) of the Charter of Fundamental Rights of the European Union and in Article 5(1)(a) of the General Data Protection Regulation (GDPR). As the European Data Protection Supervisor has pointed out, the principle of fairness can be seen as a 'core' principle (European Data Protection Supervisor Opinion…2016).

However, the principle of fairness in data protection is often dealt with in a somewhat abbreviated way, despite the fact that it is considered a fundamental principle. Even though

[11] An analysis of the processing of special categories of data is provided in the next sub-section of this chapter.

there is not much literature dealing specifically with this principle, it is possible to identify both an explicit and an implicit role of fairness in data protection. Explicitly, fairness is linked to the notions of transparency and data collection, while implicitly, fairness is linked to protection against abuse by the controller and the notion of *'fair balance'* (Clifford *et al.*, 2017). In respect to *'fair balance'* aspect ECtHR in Gaughran v UK case ruled that the unrestricted retention of biometric data, which included a digital DNA profile, as well as fingerprints and photographs of persons convicted of a crime punishable by imprisonment, violated Article 8 ECHR (*Gaughran v The United Kingdom, 2020*). This is because the retention scheme required the applicant's personal data to be retained indefinitely, regardless of the seriousness of the offence, whether it had become a 'spent conviction', the need for indefinite retention, and without any real possibility of review. The Court therefore finds that such a procedure did not ensure ***a fair balance*** between competing public and private interests and constituted a disproportionate interference with the applicant's right to respect for his private life, which could not be regarded as necessary in a democratic society (Costello, 2022).

It can be concluded that the fairness aspect is reflected in the concrete assessment of the situation from the point of view of the balance between private and public interests (the aspect of proportionality and fair balance). Therefore, when assessing whether genetic data are processed in good faith, it is necessary to consider whether the individual's private interests have not been infringed. Therefore, when assessing the fairness of the processing of genetic data, it is necessary to consider not only whether a person's private interests have been infringed, but also whether the fundamental rights of the data subject, in particular data subject right to the protection of personal data, have not been violated.

Transparency of data processing

Transparency is a long-established feature of EU law. Its aim is to foster trust in the processes that affect citizens by enabling them to understand and, where necessary, challenge these processes. It also expresses the principle of fairness in the processing of personal data, as set out in Article 8 of the Charter of Fundamental Rights of the European Union (ARTICLE 29 Data Protection WP… 2018). Although the GDPR does not include a separate definition of transparency, recital 39 of the Regulation provides information on the meaning and impact of this principle in the context of data processing:

"It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed. The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used. That principle concerns, in particular, information to the data subjects on the identity of the controller and the purposes *of the processing and further information to ensure fair and transparent processing in respect of the natural persons concerned and their right to obtain confirmation and communication of personal data concerning them which are being processed..."*

In assessing the implementation of the principle of transparency in relation to the protection of genetic data, it is important to underline that the data which are processed should be clear, open and comprehensible to the data subject. However, it is important to note that implementation of this principle can be problematic when it comes to the protection of the interests not only of the data subject, but also of the persons genetically linked to the data subject (Costello, 2022). In such a case, controllers may face a significant challenge in ensuring the transparent implementation of this principle in relation to the data subject, with a significant risk of failing to ensure the protection of the interests of the persons concerned.

Purpose limitation principle

Article 5(1)(b) of the GDPR states that personal data shall be collected for specified, explicit and legitimate purposes and shall not be further processed in a manner incompatible with those purposes. The Data Protection Working Party noted that compliance with this principle is linked to transparency, legal accuracy and predictability of data processing. Therefore, the main objective of this principle is to protect the data subject by setting limits on the use of the data subject's data by data controllers and by ensuring the fairness of processing (ARTICLE 29 Data Protection WP… 2013).

Therefore, controllers should precisely and clearly define the purposes of the processing and select the appropriate lawful basis for the processing of genetic data before starting to process the data. However, in complex genetic research, it can be difficult to identify the precise purpose of the processing and the corresponding restrictions imposed on it, as the very nature of the processing operations that may be carried out cannot always be explained in a

simple and concise manner (Quinn *et al*., 2018). Therefore, given that researchers do not know exactly what correlations they intend to find (or even what correlations they are looking for) and what their potential significance is in terms of research or privacy, it may be difficult, to be precise or concise about the exact purpose of the research at the outset of the research. This can also be complicated by the 'opportunistic' nature of data mining operations. This is because the aim of such operations may be to discover unknown relationships and correlations between various genetic sequences and physical phenomena (Roshe *et al*., 2015). The discovery of these connections can raise new questions and point research in new directions. It can therefore be argued that data mining operations (and their purpose) may be subject to constant change as a result of new information, which may make it impossible to be precise and concise about the purpose of data collection. To formulate an objective that is too precise or restrictive is likely to severely limit many types of research projects in the field of computational genetics, given that such research relies precisely on the search for previously unknown relationships in the human genome, and on the use of such findings to stimulate further research (Quinn *et al*., 2018).

Data minimisation principle

On the basis of the principle of data minimisation, personal data must be adequate, relevant and only necessary for the purposes for which they are processed (Article 5(1)(c) GDPR). The principle of data minimisation is closely linked to the principle of purpose limitation. This principle also requires an appropriate relationship between the purposes of processing and the scope of the data. Otherwise, without clear, defined, and legitimate purposes for processing, it is not possible to ensure the application of the data minimisation principle (Zaleskis, 2019). The essence of the principle is to ensure that unnecessary personal data are not collected, thereby reducing the risk of harm to the privacy of data subjects (Quinn *et al*., 2018). Accordingly, this principle allows controllers to process personal data only for legitimate explicit purposes, while prohibiting unjustified increases in the amount of data processed. It is also important to highlight that the scope of data to be processed should be determined by the controller's predetermined purpose for the processing, and personal data should only be processed if the purpose of the processing cannot reasonably be achieved by other means (GDPR Recital 39).

In the context of genetic data protection, the implementation of this principle may be complicated. In particular because the granularity of genetic testing depends on the amount of data. In the context of such studies, maximising genetic data means using the entire genome of one or more individuals. While the result may be only a small sequence of DNA out of the total amount of data, the entire genome of individuals may need to be used to perform the relevant tests (Quinn *et al*., 2018).

Data accuracy principle

Article 5(1)(d) of the GDPR establishes the principle of accuracy, which obliges data controllers to collect only accurate and, where necessary, updated personal data. It also provides for the obligation to take all reasonable measures to ensure that personal data which are not accurate in relation to the purposes for which they are processed are erased or rectified without delay.

Analysing the aspect of accuracy in the processing of genetic data, we can see a link with the principle of data minimisation. As mentioned earlier in this thesis, genetic data is characterised by its quantitative nature, i.e., the more of it that is analysed, the more likely it is that the result will be more accurate. In this case, we can see the correlation and dependence between these two principles. However, there are reasonable doubts as to what could be considered as inaccurate genetic data in practice since the Regulation itself does not provide a definition of inaccurate genetic data.

Storage limitation principle

The principle of limitation of retention can be seen as a fundamental principle of data protection, which essentially implies that data should not be retained for longer than necessary. Therefore, when the reason for processing the data concerned ceases to exist, the data should be erased. In this way, the risk that the personal data will be misused at a later stage by the controller or another third party is reduced (Quinn *et al*., 2018).

This principle is enshrined in Article 5(1)(e) of the Regulation and which provides that personal data must be: *"kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or*

statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject". The requirements for data controllers set out in this Article are complemented by GDPR Recital 39 which sets out that: "*ensuring that the period for which the personal data are stored is limited to a strict minimum".* In respect to this principle implementation, in *S. and Marper v. the United Kingdom, 2008* case the ECtHR noted that the retention of personal data should be limited in time (they are retained only for as long as they are being processed) and proportionate to the purpose, and that they should be destroyed once that purpose has been achieved.

In the context of this principle, the processing of genetic data for the purpose of scientific research could be highlighted. As can be seen, the processing of genetic data for this purpose is subject to lighter requirements, such as the possibility to set a longer data retention period. *Quinn* identifies two main problems with the implementation of this principle in the case of medical genetic research. The first is that, the research may last longer than originally planned. In this case, the research period is prolonged because of further discoveries that occur after the data has been mined. Second, the implementation of accessibility of datasets. In practice, this is considered to be an element of good research practice - making datasets accessible to subsequent researchers (Quinn, *et al*., 2018).

As noted, the implementation of the principle under analysis in the context of scientific genetic research can be difficult, and despite the fact that the Regulation provides for an exception for a more flexible period of time for the processing of genetic data, controllers should not abuse this possibility and should attempt to maximise the implementation of the conditions of the principle in such a manner as to ensure the protection of data subjects.

Integrity and confidentiality principle

One of the most significant features of the data protection reform is the introduction of specific and broader requirements for personalised and standardised data protection. The principle of integrity and confidentiality requires that data shall be processed in such a way as to ensure, by appropriate technical and organisational means, adequate security of personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage (Article 5(1)(f) of the GDPR). Therefore, the main purpose of this principle is to oblige the controller or processor to take all feasible and lawful technical or

organisational measures to ensure the security of personal data. Under the GDPR regulation, the controller must put in place appropriate technical and organisational measures aimed at the effective implementation of the data protection principles and incorporate the necessary safeguards in the processing to comply with the provisions of the data protection law, ensure that only personal data which are necessary for the fulfilment of the purposes of each specific processing are processed in a standardised manner (GDPR Article 25(1-2)). These obligations explicitly include the requirement to comply with the other data protection principles of minimisation and proportionality, as well as with the limitations on data accessibility provided for.

When analysing the processing of genetic data, three main aspects can be identified in relation to the implementation of this principle: data confidentiality, security, and anonymity (Clayton *et al*., 2019).

Confidentiality describes a situation where information is disclosed in a relationship of trust (e.g., doctor-patient), with an explicit agreement that it will not be disclosed to others without the consent of the data subject. Confidentiality of genetic information is a core principle of many codes of ethics for the health professions and an element of many laws. However, the duty of confidentiality is not absolute; other interests, such as the safety and health of third parties, may prevail in certain circumstances recognised by law or codes of ethics (Clayton *et al*., 2019).

Security in the technology-driven digital space is becoming an increasingly important aspect today. Security is achieved by granting access to certain information to persons or entities with the appropriate authority to access it, while denying access to those without such authority. Security can be protected by a variety of means, such as training staff, adopting administrative procedures for handling sensitive information and implementing technical access control measures, including passwords and encryption (Clayton *et al*., 2019).

Anonymity is a form of privacy protection in which the identity of the source of certain genetic information is withheld or removed by data controllers. Anonymisation, de-identification and similar measures are often applied to genetic information in order to protect the privacy of the individual while preserving the scientific value of the information. The use of anonymised genetic information raises two main problems. First, technical methods may not be fully effective in preventing re-identification of genetic information. Second, there is a compelling argument that individuals' interest in autonomy should give them the ability to

know and control the use of even anonymised health information or biological samples (Clayton *et al.*, 2019).

In conclusion, it can be assumed that controllers may apply other measures to implement and enforce the security principle. In this case, it is important that the measures taken ensure the security of the genetic data processed and that the processing is lawful, thus ensuring the protection of the data processed against accidental loss, destruction or damage, in this case by implementing appropriate technical or organisational security measures.

Accountability principle

The principle of accountability is the cornerstone of all the principles enshrined in Article 5 of the GDPR. The accountability clause obliges the processor or controller to take responsibility for the processing of personal data, to comply with the data protection requirements and to be able to demonstrate that the processing is carried out in compliance with the Regulation. (GDPR Article 5(2)). Since controllers are liable for personal data breaches resulting in the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of personal data, unauthorised access to personal data, or any other unauthorised transfer, storage or processing of personal data, it is therefore up to the controller in this case to ensure that adequate security controls are in place in order to demonstrate liability.

Therefore, taking into account the essence of this principle, it can be distinguished that processors of genetic data should be able to justify their chosen method of operation and that it meets the requirements of the Regulation. However, as can be seen from the above analysis of the principles enshrined in the Regulation, controllers may face difficulties in enforcing all of them (given that this principle is the guarantor of the enforcement of all principles). As mentioned above, controllers may face the problem of minimising the data collected (in the case of genetic data collection, the quantitative aspect is very important). There are also data anonymisation issues (except for identical twins, each person’s DNA sequence is unique, which means a DNA sample can never be truly anonymized (National Human Genome Research Institute).

To summarise all the analysed principles in the aspect of genetic data protection, it can be distinguished that, above all, the protection of the mentioned data is particularly important for the legal basis of such data processing. Secondly, the validity of such data processing is directly related to the data subjects' right to respect for private and family life. Preservation of

special categories of data is necessary in a democratic society, therefore personal data processors must use all possible, legal technical and organizational measures to ensure that particularly sensitive information is not illegally transferred or disclosed, as this is incompatible with the guarantees of the rights of individuals provided in GDPR.

2.4. Requirements for data security

2.4.1 Security measures for the rights of the data subject

The previous analysis of this thesis has clearly highlighted that the processing of genetic data concerns highly sensitive personal information which may have a direct impact on fundamental rights and freedoms of the individual. To fully analyse the protection of genetic data, the further analysis of this thesis will focus on the specificity of the application of the safeguards contained in the GDPR, with the aim of ensuring the fundamental rights and freedoms of data subjects.

Recital 88 of the Regulation states that in order to ensure security and to prevent processing in breach of the GDPR, the controller or processor should assess the risks associated with the processing and implement measures to mitigate those risks. Those measures should ensure an adequate level of security, including confidentiality, taking into account the state of the art and the costs of implementation in relation to the risks and the nature of the personal data to be protected. In assessing data security risks, account should be taken of the risks arising from the processing of personal data, such as accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data, whether transmitted, stored or otherwise processed, which are likely to cause, in particular, physical, material or non-material damage. As can clearly be seen, Recital 88 of the GDPR explicitly lays down obligations for the controller and the processor, which include (a) ensuring the security of the processing; (b) assessing the risks associated with the processing; (c) taking measures to mitigate the identified risks. It should be emphasised that, as it has been previously stated in this thesis that genetic data are personal data, all the above-mentioned obligations also apply to data controllers and processors who process genetic data of data subjects.

Recital 88 of the GDPR, although it provides for obligations relating to ensuring the security of processing, does not provide for specific measures to be taken by the controller and processor. In this context, it is important also to analyse Article 32 of the GDPR, which states that: *"Having regard to the state of the art and the costs of implementation and taking into*

account the nature, scope, context and purposes of the processing as well as the risk of varying likelihood and severity for the rights and freedoms of individuals, the controller and the processor shall implement appropriate technical and organisational measures, to ensure a level of security appropriate to the risk, including inter alia, as appropriate: (a) the pseudonymisation and encryption of personal data; (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of systems and services processing personal data; (c) the ability to restore the availability and access to data in a timely manner in the event of a physical or technical incident; (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing".

The article also points out that *"in assessing the appropriate level of security account shall be taken in particular of the risks that are presented by data processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed".* It also mentions that compliance with an approved code of conduct (Article 40 of the GDPR) or an approved certification mechanism (Article 41 of the GDPR) may be used as an element of demonstrating compliance with the security requirements for data processing. Finally, the controller and the processor shall *"shall take steps to ensure that any person acting under their authority and having access to personal data, shall not process them except on instructions from the controller, unless otherwise required by Union or member state law".* According to the above provisions, security under the GDPR covers confidentiality, integrity and availability and should be assessed on the basis of a risk-based approach: the higher the risk (to the rights and freedoms of data subjects), the more stringent the measures to be taken by the controller or the processor (to manage the risks). Similarly, the security of data processing should be assessed in the context of the GDPR's general data protection accountability framework, which is also risk and impact based and aims to be appropriate to the specific context and practices of the organisation. Furthermore, Article 28(4) requires data controllers to conclude a written contract with the data processor that, among other things, requires the processor to take all measures required by Article 32, as well as to assist the controller in ensuring compliance with its own obligations under Article 32. This requirement only confirms that not only the controller, but also the processor is obliged to ensure the safeguards set out in Article 32 GDPR.

In the following, this thesis will seek to analyse in more detail the organisational and technical measures listed in Article 32(2) of the GDPR that controllers and processors must take when processing data subjects' data.

The pseudonymisation and encryption of personal data

GDPR clearly recommends the pseudonymisation and encryption of personal data as one of several ways to mitigate risks from the data subject's point of view, as well as to ensure the privacy of data subjects. This approach facilitates, among other things, the processing of personal data by controllers for purposes other than the primary purpose of collecting the personal data, as well as the processing of personal data for scientific and other purposes.

Article 4 of the GDPR defines pseudonymisation as :*"the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures"*. Therefore, in the light of the definition, pseudonymisation can be understood as a method used by controllers and processors to reduce the likelihood that personal data records and identifiers will be able to identify the natural person (data subject) to whom they belong. Only with the use of certain additional information can the data subject be identified. In this respect, it is important to note that pseudonymisation should not be confused with anonymisation. According to Article 29 Working Party, pseudonymised data cannot be equated to anonymised information as they continue to allow an individual data subject to be singled out and linkable across different data sets (ARTICLE 29 Data Protection WP...2014/05). The implementation of these measures (pseudonymisation) in the context of genetic data raises several issues. *Firs of all*, it was mentioned earlier in this thesis that Article 9(2) (j) of the GDPR enshrines the right for controllers and processors to process special categories of data for scientific or historical research purposes. In this respect, it is essential to use identifiable data for example in epidemiological studies. Therefore, pseudonymisation may be a limiting factor in the use of genetic data under the research exemption (Shabani *et al.* 2017). *Secondly*, the concept of pseudonymisation, as enshrined in Article 4 of the GDPR, focuses on the fact that the pseudonymisation method eliminates the possibility of directly attributing the data to a person, and therefore requires the provision of supporting information in order to find out to whom the data belong. Particular attention is therefore focused on whether third parties can identify a

person on the basis of such data (Sukhorolskyi *et al.* 2020). It is questionable whether it is really possible to correctly foresee and take into account further technological advances in order to arrive at an unbiased conclusion as to how much effort and resources will be needed to identify a person on the basis of pseudonymised data, even in the near future (Sukhorolskyi *et al.* 2020). Therefore, the provision under analysis can be interpreted as enabling the controller and the processor to rely on the use of pseudonymisation as a guarantee of the protection of the data subject's rights, even if this action is not justified. It is also not clear what action should be taken in the future when technological progress will allow for a significant reduction of the time and effort needed to identify a person on the basis of his or her pseudonymised data, given that such data will already be available to a wider range of persons (Sukhorolskyi *et al.* 2020).

The implementation of the pseudonymisation approach with regard to the protection of genetic data raises reasonable doubts, in particular as to the practical implementation of this approach (whether it is really possible to identify a person using separate information that is not directly attributable to the data subject), and also as to the way in which future technologies will impact on the performance of the relevant actions, once the actual process of identification will be shortened, and the data will be already known to a wider group of persons.

Recital 83 of the Regulation establishes the obligation of the controller and the processor to assess the risks associated with the processing to ensure the security of the data processed and to prevent it from processing data in breach of the GDPR. *Encryption* is one of the measures to mitigate the risks of processing.

Data encryption is one of the means to ensure the safety of electronic data or information (information systems). The protection of such value is receiving increasing attention both from a technical and a legal point of view all over the world. Genetic data security (safety) is understood as the protection of information and system infrastructure from accidental or intentional, natural or artificial effects that can cause damage to the owners and users of information or system infrastructure (Petkevičienė *et al.* 2020).

According to the Article 29 Working Party, the following encryption methods are commonly used: 1) *Encryption with a secret key.* In such a scenario, the key holder can trivially re-identify each data subject after decrypting the dataset, as the personal data is still in the dataset, albeit encrypted. Assuming that the latest encryption scheme has been applied, decryption can only be achieved by knowing the key; 2) *Hash function* that returns a fixed

output (the input can be a single attribute or a set of attributes) from any input of any size, which cannot be changed, meaning that the risk of reversal inherent in encryption is eliminated. However, if the range of input values held by the hash function is known, they can be reconstructed via the hash function to produce the correct value for a particular record; 3) *Salted – hash function.* Salted password encryption can be used to increase the security of passwords by adding additional layers of randomness to the encryption process. 4) *Keyed-hash function with a protected key:* this corresponds to a kind of hash function that uses a secret key as additional input. The controller can recover the attribute function using the secret key, but it is much harder for an attacker to recover the function without knowing the key, because the number of possibilities to be tested is large enough to be impractical; 5) *Deterministic encryption or key-hash function with key deletion:* this method can be compared to choosing a random number as a pseudonym for each attribute in the database and deleting the table of matches. 6) *Tokenisation.* This method is a derivative of the previous ones, usually based on the use of one-way encryption mechanisms or on the assignment of a sequence number or a randomly generated number that is not mathematically derived from the original data, using an index function (ARTICLE 29 Data Protection WP...2014/05).

While there are many encryption methods and techniques available, and while each is effective in protecting data at some stage in its lifecycle, no encryption method can guarantee the security of the data in perpetuity and preserve the underlying file structure. (Senf *et al.* 2021). In this context, it should be highlighted that the Global Alliance for Genomics and Health has endorsed the Crypt4GH File Encryption Standard, which allows to read encrypted data from a file or re-mote the Application Programming Interface and to decrypt only bytes in memory (GA4GH File Encryption Standard). This approach uses envelope encryption, a protocol that is relatively new in research and healthcare but is increasingly used in data security to improve the security of data transmission and storage (Global Alliance for Genomics&Health). The new Crypt4GH standard will ensure the security of genetic data throughout its lifecycle, from initial sequencing to sharing with external organisations. It can be concluded that the emergence of new tools (e.g., the Crypt4GH File Encryption Standard) demonstrates the attention paid to the protection of genetic data.

Considering the advances in technology and the fact that encryption is linked to information systems, which are also known to be evolving, the protection of genetic data should be given particular attention. In view of the particular sensitivity of genetic data and the

fact that any unauthorised disclosure of such data may infringe the fundamental rights and freedoms of individuals, it is therefore considered that controllers and processors should apply state-of-the-art encryption techniques in order to manage the risks associated with the processing of genetic data.

The ability to ensure the ongoing confidentiality, integrity, availability and resilience of systems and services processing personal data

In practice, three main aspects of information systems and genetic data security can be distinguished*:* 1) *availability* - protection against the risk of accidental or unauthorised loss of access to or destruction of personal data (ARTICLE 29 Data Protection WP…2016/679); 2) *integrity* – protection against unauthorised or accidental alteration of personal data (GDPR 5(1) (f)); 3) *confidentiality* – protection against unauthorised or accidental disclosure of, or access to, personal data (GDPR 5(1) (f)); 4) *resilience* - protection to keep systems running under adverse conditions[12] (Information Commissioner's Officer). The implementation of the measures analysed are also covered by the ISO 27701 standard: *"The organization shall apply the information security risk assessment process to identify risks associated with the loss of confidentiality, integrity and availability, within the scope of the PIMS".*

Confidentiality, integrity, and availability are the three key elements of information security, collectively known as the "CIA Triad" (Information Commissioner's Officer). The security of genetic data is therefore a combination of all the above. It is important to underline that a breach of any one of these three elements first constitutes a serious risk to the rights of data subjects, and second of all, a breach of these elements (or at least of one of them) gives rise to the liability of the data controller.

It should be noted that specific terms and concepts are not yet sufficiently explicit in the law and there is a lack of regulation in this area. Therefore, effective security of information managed in information systems should be one of the most important priorities of the state information policy. According to J. Januševičienė (2018), the Organization for European Cooperation and Development (OECD) was one of the first to regulate the security issues of information systems and the information processed in them on an international scale. The directives adopted by the OECD should be evaluated as legal acts of specific importance,

[12] For example due to a physical or technical incident.

indicating to the member states the main directions of activity in the field of information security regulation, implementing the policy of managing access to personal data (Januševičienė, 2018).

The Personal Data Access Management Policy sets out the rules for access to various systems, equipment and information. It should be noted that this policy aims to limit accidental or unauthorised access, including to genetic data. To achieve this objective, security measures are in place which technically act as access control measures. Therefore, this document can act as a preventive measure aimed at preventing information theft, fraud and, subsequently, litigation (Daigle *et al*., 2020).

It is important to underline that controllers and processor are required to implement a range of organisational and technical measures to ensure that the genetic data processed comply with the requirements of confidentiality, integrity, availability, and resilience, as referred to in Article 32(1)(b) of the GDPR. In order to comply with these requirements, controllers and processors (e.g., healthcare institutions) must ensure the security of the systems they use, as this is directly related to genetic data security.

The ability to restore the availability and access to data in a timely manner in the event of a physical or technical incident

Article 32(1)(c) of the GDPR imposes an obligation for controllers and processors to be able to restore access to data in the event of a physical or technical incident. In analysing this measure, it is important to detail what constitutes a physical or technical incident and what the legislator meant by "timely manner".

Firstly, it is essential to assess whether a physical or technical incident caused the personal data breach. GDPR defines a "personal data breach" in Article 4(12) as: *"a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed".* It could be emphasised that a breach is a specific security incident. However, as stated in Article 4(12), the GDPR only applies in cases where personal data are breached. The consequence of such a breach is that the controller will not be able to ensure compliance with the principles relating to the processing of personal data as set out in Article 5 of the GDPR. This highlights the difference between a security incident and a personal data breach. Therefore, while all personal data breaches are security incidents, not all security incidents are necessarily personal data

breaches (ARTICLE 29 Data Protection WP…2016/679). In the light of the above, controllers and processors should first assess whether a security breach of genetic data has occurred in the event of a security incident. In the event of a data breach, the controller has an obligation to notify the breach (Article 33 of GDPR).

Second, the measure under analysis implies that the controller or processor must be able to restore the availability of and access to data in a timely manner. The Regulation does not specify what constitutes 'in a timely manner', it could be argue that it leaves room for interpretation and ambiguity. Nevertheless, article 33 of the GDPR states that the controller must notify a data breach to the supervisory authority within 72 hours. In addition, article 34(1) of the GDPR provides that: *"When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay".* It should be noted that the term "undue delay" referred to in the Article under analysis is not clear in this context and the Regulation itself does not further define the time of notification. However, the Article 29 Working Party explained, that "undue delay" means as soon as possible (ARTICLE 29 Data Protection WP…2016/679). For this reason, it can be argued that in the event of a genetic data breach, the controller must notify the data subject as soon as possible.

In conclusion, it is important to underline that, as mentioned earlier in this thesis, genetic data is a special category of data subject to enhanced processing requirements and to a very high level of scrutiny and responsibility on the part of data controllers and processors. The interest to protect highly sensitive data is confirmed by the fact that the GDPR additionally imposes an obligation on controllers to notify a breach of security of genetic data to data subjects. This is in particular because the disclosure of genetic data is directly related to the rights and freedoms of data subjects.

A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing

Article 32(1)(d) of the GDPR provides that controllers and processors should regularly test the measures in place in order to ensure the security of the data processed. It could be assumed that the above requirement is cumulative of all the measures to be applied by controllers and processors in Article 32(1).

Although the GDPR does not specify testing tools and methods which ensure the safety of this data, vulnerability scanning and penetration testing could be used in practice (Information Commissioner's Officer). For instance, ENISA on the Handbook on Security of Personal Data Processing points out that: "*Vulnerability assessment, application and infrastructure penetration testing should be performed by a trusted third party prior to the operational adoption. The application shall not be adopted unless the required level of security is achieved"*, also " *During the development, testing and validation against the implementation of the initial security requirements should be performed"*. As can be clearly seen, the document in question sets out in more detail the requirements to be followed by controllers and processors when carrying out testing activities.

It is important to stress that computer technology and genomics professionals manage and store genomic data using a variety of computer systems and software. Increasingly, data analysis and coordination centres are part of research networks and provide these services. Many private and commercial cloud platforms hosting genetic data work in partnership with government and public bodies that define and provide the storage and computing infrastructure to host genomic data and, in particular, to ensure the necessary security and privacy protection for human genomic data. These data are made available to the wider scientific community for further data analysis, which includes information about the human genome, such as the location of genes and variants in DNA (National Human Genome Research Institute).

In this context, it can be concluded that the measure at issue is complementary to the measures to be implemented by the controller and the processor analysed above. Periodic testing of the systems used by the controller and the processor for processing genetic data is necessary to ensure the security of such data. In particular because a breach of the security of genetic data is a high-risk factor for the rights and freedoms of individuals, and because genetic data is mostly used for scientific purposes and their availability is attributed to the wider community, which also reflects a high data security risk.

3. SPECIFIC RULES FOR THE PROTECTION OF GENETIC DATA

3.1. Prohibition of processing of special categories of personal data

Personal data protection law provides for specific, stricter requirements for the processing of special categories of data. All general requirements apply to the processing of such data to the extent that the specific requirements do not provide otherwise (Zaleskis, 2019). Special protection should be afforded to such data, which by their very nature relate to fundamental rights and freedoms and are therefore sensitive, as the processing environment could result in a serious risk to fundamental rights and freedoms (Zaleskis, 2019).

The definition of special data is set out in Article 9(1) of the GDPR, which states that: *"Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of* ***genetic data****, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited".* As it can be observed, the GDPR provides a separate article for the processing of special categories of data, which only shows the high level of sensitivity of this type of data. Unlike Directive 95/46/EC, which did not distinguish what genetic data means and what legal value it has, the Regulation explicitly classifies genetic data as a group of special categories of personal data. Although Article 9(1) of the GDPR prohibits the processing of special categories of data, Article 9(2) provides for exceptions when such prohibition may not apply. The prohibition and the exceptions provided for in the GDPR replace the general requirement to base processing on at least one processing ground.

The Article 29 Working Party's opinion on genetic data argues that Directive 95/46/EC is understood as implicitly allowing genetic data to be recognised as 'health data', which are sensitive data, and hence subject to more stringent requirements (ARTICLE 29 Data Protection WP…2004). In the Working Party's view, in order to demonstrate that genetic data are sensitive data, it was necessary to establish that it should be understood not only as personal data but also as health data. This required demonstrating that the genetic data in question could be a "link" to the "health status" of the identifiable person. This condition implies a very wide possible interpretation of 'health data'. This is because the concept was not limited to data indicating the presence of a disease but could also include data suggesting the possibility of

the development of a disease (even if slight), and even data simply confirming that a person is 'healthy' (Quinn *et al*., 2018).

However, while there is some agreement that genetic data can be considered as health data, there is the challenge of defining health data too broadly. In this respect, according to the Working Party the wide range of personal data may fall under the category of health-related data and for this reason it could be considered that this category is one of the most complex areas of sensitive data, and one in which Member States are subject to considerable legal uncertainty. Therefore, specific measures are needed to protect health data with serious privacy implications from misuse (e.g., commercial use of patient data). Therefore, the Working Party considers that genetic data should be distinguished as a separate category of data for the purpose of greater clarity and definition of the legally established concepts (ARTICLE 29 Data Protection WP… 2011).

It has already been mentioned above that the processing of genetic data poses serious risks to the rights and freedoms of individuals. In this context, it is important to mention that Recitals 71 and 75 of the GDPR identify potential risks and likely discriminatory effects on natural persons on the basis of genetic or medical conditions. However, these provisions appear to lack clarity and detail on the risks themselves. It is important to underline here, the Article 29 Data Protection Working Party's opinion analyses this issue in more detail. It outlines the characteristics of genetic data that distinguish it from other categories: (1) genetic data reveal information not only about the data subject, but also about his or her blood relatives and certain groups of persons to which he or she belongs; (2) as a rule, genetic information is unknown to the bearer him/herself and does not depend on the bearer's individual will since genetic data are non-modifiable; (3) genetic data can be easily obtained from raw materials; (4) genetic data may reveal more information in the future and be used by an increasing number of agencies for various purposes (ARTICLE 29 Data Protection WP…2004). Also, taking into account potential risks arising from the processing of genetic data, it is highlighted that the most significant risks arise from the re-use of genetic data (where additional analysis of stored biological material is performed). The Working Party identifies the risks arising from the use of such data for employment, insurance, identification, medical and research purposes (ARTICLE 29 Data Protection WP…2004).

In conclusion, genetic data, although health-related, is interpreted too broadly in the context of health data. Therefore, the specificity of genetic data has given rise to a clear need

to consider this category of data not as 'part of' something, but as a distinct, independent and categorized as a special category of data in the context of the GDPR, thus underlining the enhanced protection of these data. It should be noted that, as highlighted earlier in this thesis, genetic data is highly sensitive information, and therefore sensitive personal data is subject to a higher regulatory burden than non-sensitive data, and the legal situation with regard to genetic data is constantly changing.

3.2. Exceptions to the processing of special categories of personal data

The processing of special categories of personal data, also known as sensitive personal data, is subject to stricter rules and regulations under the General Data Protection Regulation (GDPR). These categories include information such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and genetic or biometric data. However, there are several exceptions to the processing of special categories of personal data under the GDPR. These exceptions allow for the processing of sensitive personal data in certain circumstances, while still providing protection for the individual's rights and freedoms.

One exception is the explicit consent of the individual. Under Article 9(2)(a) of the GDPR, the processing of special categories of personal data is allowed if the individual has given their explicit consent. However, this exception is limited and only applies in specific situations. For example, an individual may give their explicit consent for their genetic data to be processed for the purpose of a medical study (ARTICLE 29 Data Protection WP…2004).

Consent from the data subject

The consent of the data subject to the processing of genetic data is a complex and multifaceted issue that has garnered significant attention in the field of data protection and privacy. At the heart of this issue is the question of whether or not individuals have the right to control how their genetic information is used and by whom. This question is particularly relevant in the context of genetic testing and research, where the use of genetic data can have significant implications for both the individual and society as a whole (ARTICLE 29 Data Protection WP…2004). One key factor that has been identified as influencing the consent of the data subject to the processing of genetic data is the level of trust that individuals have in the institutions or organizations involved. It was found that individuals were more likely to consent to the use of their genetic data if they had confidence in the security and confidentiality of the

data, as well as in the purpose and potential benefits of the research (Clayton *et al.* 2019). Also, individuals were more likely to consent to the use of their genetic data if they felt that their privacy and autonomy were respected, and if they had access to clear and accurate information about the research and its potential implications (Clayton *et al.* 2019).

The issue of consent to the processing of genetic data is also closely tied to the concept of informed consent, which requires that individuals have sufficient information about the risks and benefits of participating in a particular activity, and that they are able to make a fully informed decision about whether or not to participate. In the context of genetic testing and research, this means that individuals should be provided with clear and accurate information about the purpose of the research, the potential risks and benefits of participating, and the steps that will be taken to protect their privacy and confidentiality. However, obtaining informed consent for the use of genetic data can be a complex process, and there is ongoing debate about the best ways to ensure that individuals are fully informed about the risks and benefits of participating in genetic testing and research. It could be argued that the use of written consent forms is the most effective way to ensure that individuals are fully informed, but it could be also argued that more interactive approaches, such as face-to-face discussions or online resources, may be more effective in ensuring that individuals have a full understanding of the issues involved.

Overall, the consent of the data subject to the processing of genetic data is a critical ethical and legal consideration. Ensuring that individuals are fully informed and able to freely and willingly provide their consent is essential to protecting the privacy and autonomy of the individual, and to fostering trust in the use of genetic data for research and healthcare purposes.

An important public interest

There has been a significant public interest (Article 9(2)(g) GDPR)) in the processing of genetic data in recent years due to advances in genomics and the potential for personalized medicine. This interest has been fuelled by the availability of genetic testing services, such as direct-to-consumer genetic testing companies, and the increasing amount of genetic data being collected and shared by researchers and healthcare providers (Yamamoto *et al.* 2022).

One factor contributing to the public interest in genetic data is the potential for genetic testing to identify inherited conditions and diseases, such as breast cancer and Alzheimer's disease. According to a survey conducted by the National Human Genome Research Institute,

around 75% of people who have undergone genetic testing did so to learn about their risk of developing a particular health condition (NHGRI, 2018). This knowledge can help individuals make informed decisions about their health, such as seeking preventive measures or undergoing regular screenings.

Another reason for the public interest in genetic data is the potential for personalized medicine, which involves tailoring medical treatment and prevention strategies to an individual's genetic makeup. Personalized medicine has the potential to improve the effectiveness and efficiency of healthcare, as it allows for the targeting of specific therapies to individuals who are most likely to benefit from them (Mathur *et al*., 2017). However, there are also concerns about the potential for genetic data to be used for non-medical purposes, such as discrimination in employment or insurance (Godarn *et al*., 2004).

There are also ethical considerations surrounding the processing of genetic data, including privacy and consent. The collection and sharing of genetic data can raise concerns about the potential for this information to be misused or accessed by unauthorized parties (Clayton *et al.,* 2019). There are also concerns about the potential for genetic data to be used to make predictions about an individual's characteristics, such as intelligence or behaviour, which may be perceived as stigmatizing or unethical (Berrynessa *et al*., 2013).

In conclusion, the significant public interest in the processing of genetic data is driven by the potential for genetic testing to identify inherited conditions and diseases, and the potential for personalized medicine. However, there are also ethical considerations surrounding the collection and sharing of genetic data, including privacy and consent. Further research is needed to address these concerns and ensure the responsible use of genetic data in the pursuit of improving healthcare.

Scientific research

Genetic data processing has become a crucial aspect of modern science research **(**GDPR Article 9(2)(j)). The increasing availability of large-scale genomic data has allowed researchers to gain insights into a wide range of biological phenomena, including disease susceptibility, evolution, and population genetics (Bamshad *et al.,* 2003). However, the analysis of genetic data also poses significant computational challenges, as the amount of data generated by modern sequencing technologies can be enormous (Quin *et al*., 2018).

One of the major challenges in genetic data processing is the identification of single nucleotide polymorphisms (SNPs). SNPs are single base pair differences between individuals within a population, and they can provide valuable information about an individual's genetic makeup and disease susceptibility (Hindorff *et al*., 2009). To identify SNPs, researchers must analyze large amounts of genomic data, which can be time-consuming and computationally intensive.

To address these challenges, researchers have developed a variety of computational tools and algorithms for the analysis of genetic data. One such tool is the Short Oligonucleotide Analysis Package (SOAP), which is a software suite designed for the alignment and variant calling of next-generation sequencing data (Hintzsche *et al.*, 2016). Another tool is the Genome Analysis Toolkit (GATK), which is a widely used suite of software tools for the analysis of high-throughput sequencing data (McKenna *et al*., 2010).

In addition to these computational tools, researchers have also developed statistical methods for the analysis of genetic data. For example, the use of principal component analysis (PCA) has become increasingly common in the analysis of large-scale genomic data, as it allows researchers to identify patterns and trends in the data that may not be immediately apparent (Elhaik 2022). Similarly, the use of machine learning techniques, such as support vector machines and random forests, has also become increasingly prevalent in genetic data analysis (Ishwaran *et al*., 2012).

In conclusion, the analysis of genetic data has become an essential aspect of modern science research. However, the processing of large-scale genomic data poses significant computational challenges, which have been addressed through the development of computational tools and statistical methods. These tools and methods have allowed researchers to gain valuable insights into a wide range of biological phenomena, and they will likely continue to play a crucial role in the field of genetics for years to come.

Apparently publicly available data

The processing of genetic data has become increasingly important in recent years due to advances in technology and the growing interest in personalized medicine. One aspect of this process is the use of publicly available data (GDPR Article 9(2)(e)), which can be accessed by researchers, clinicians, and the general public. This data can come from a variety of sources, including genetic databases, clinical trials, and patient records.

One major source of publicly available genetic data is the National Center for Biotechnology Information's (NCBI) Genetic Testing Registry (GTR). The GTR is a database that provides information on genetic tests, including their purpose, methodology, and validity[13]. This data is useful for researchers and clinicians who are interested in understanding the different types of genetic tests available and their potential uses.

Another source of publicly available genetic data is the ClinicalTrials.gov database[14], which is maintained by the National Institutes of Health (NIH). This database contains information on clinical trials that are being conducted around the world, including those that involve genetic testing. This data is useful for researchers and clinicians who are interested in understanding the latest developments in genetic testing and how it is being used in clinical practice.

Publicly available genetic data can also come from patient records and electronic health records (EHRs). These records can contain a wealth of information about a patient's medical history, including their genetic makeup. This data is useful for researchers and clinicians who are interested in understanding how genetics can impact the course of a patient's treatment.

While the use of publicly available genetic data has many benefits, it is important to consider the potential ethical implications. One concern is the potential for privacy breaches, as genetic data is highly sensitive and personal. In order to address this concern, it is important to ensure that any publicly available genetic data is properly de-identified and that appropriate consent is obtained from the individuals whose data is being used.

Overall, the use of publicly available genetic data can be a valuable resource for researchers and clinicians who are interested in understanding the role of genetics in health and disease. However, it is important to carefully consider the potential ethical implications and ensure that appropriate safeguards are in place to protect the privacy of individuals whose data is being used.

3.3. Obligation to carry out a data protection impact assessment when processing genetic data

[13] For more information: https://www.ncbi.nlm.nih.gov/gtr/
[14] For more information: https://clinicaltrials.gov

Genetic data, also known as genomic data, is a complex and sensitive type of personal data that contains information about an individual's genetic makeup and inherited characteristics. The processing of genetic data has the potential to significantly impact an individual's privacy and personal autonomy, as it can reveal sensitive information about an individual's health, ancestry, and predisposition to certain diseases or conditions. As such, the processing of genetic data is subject to the GDPR, which requires organizations to carry out a data protection impact assessment (DPIA) when processing genetic data, or any other type of personal data that is likely to result in high risks to the rights and freedoms of individuals (Article 35 of the GDPR).

According to Article 35 of the GDPR, a DPIA is a process that involves assessing the potential risks to individuals' rights and freedoms resulting from the processing of personal data and determining the appropriate measures to mitigate those risks. A DPIA must be carried out when an organization plans to process personal data in a way that is likely to result in high risks to individuals, or when the processing involves new technologies or innovative uses of personal data.

The processing of genetic data is considered high-risk under the GDPR, as it has the potential to reveal sensitive and potentially stigmatizing information about an individual's health, ancestry, and predisposition to certain diseases or conditions. This can have significant consequences for an individual's privacy, personal autonomy, and dignity, as well as for their employment, insurance, and other areas of their life. As such, organizations that process genetic data must carry out a DPIA to ensure that they are complying with their obligations under the GDPR and protecting the rights and freedoms of individuals.

In this context, it is important to mention the list published by the State Data Protection Inspectorate of the Republic of Lithuania of when a data protection impact assessment is necessary. The list of data protection impact assessments stipulates that a data protection impact assessment is mandatory when personal data are processed for research purposes or when special categories of personal data are processed without the consent of the individual. It can be seen that the Inspectorate does not distinguish the multidimensionality of the data when processing special categories of personal data which may be based on genetic data, which in principle means that the processing of genetic data based on special categories of personal data will in practice always require a DPA, the list also foresees that an DPA will be mandatory where personal data will be processed on a large scale, where the personal data are not obtained

from the individual and where, in certain cases, the provision of information is not possible or would require a disproportionate effort, or where the provision of such information may make the achievement of the purposes of the processing impossible or very difficult. Thus, it should be noted that the Inspection Schedule, unlike the GDPR, does not only oblige the controller to carry out a DPA in respect of specific categories of data, but also in respect of any processing of data where it is carried out on a large scale and does not come from the data subject. This shows once again that large-scale processing of genetic data, even if it does not qualify as special category data under the GDPR, will require an DPA.

In conclusion, the processing of genetic data is subject to the GDPR's requirements for a DPIA, due to the high risks that it poses to the rights and freedoms of individuals. Organizations that process genetic data must therefore carry out a DPIA to ensure that they are complying with their obligations under the GDPR and protecting the privacy and personal autonomy of individuals. By following the steps outlined above, organizations can effectively manage the risks associated with the processing of genetic data and ensure that they are protecting the rights and freedoms of individuals.

3.4. The obligation to designate a DPO for the processing of genetic data

The obligation to designate a data protection officer (DPO) for processing genetic data has become a hot topic in recent years, as advances in genetic testing and the increasing availability of personal genetic information have raised concerns about privacy and the ethical use of this sensitive information. GDPR, which came into effect in 2018, requires organizations that process genetic data to appoint a DPO, and this obligation has been further clarified in subsequent guidance from the EDPB.

Article 37(1)(c) of the GDPR provides that the controller and the processor shall designate a DPO where "the controller's or processor's main activity is the processing of special categories of data on a large scale in accordance with Article 9, <...>."

According to the GDPR, genetic data is defined as "*personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question*" (Article 4(15)). This definition

encompasses a wide range of genetic information, including DNA sequencing data, genetic predisposition testing, and ancestry testing.

The GDPR imposes specific requirements on organizations that process genetic data, including the obligation to appoint a DPO. The DPO is responsible for ensuring that the organization complies with the GDPR and other data protection laws, as well as advising on data protection policies and strategies. The DPO must be an independent expert with knowledge of data protection law and practices and must be appointed on the basis of their professional qualities, rather than their position within the organization.

The appointment of a DPO is not mandatory for all organizations, but it is required for those that process large amounts of personal data, including genetic data, or that engage in high-risk data processing activities. This includes research institutions, biobanks, and genetic testing companies, as well as hospitals and other healthcare organizations that collect and process genetic data for diagnostic or treatment purposes.

The EDPB has issued guidance on the appointment and role of DPOs in the context of genetic data processing, stating that "*genetic data is considered as sensitive personal data and therefore requires a higher level of protection*" (ARTICLE 29 Data Protection WP… 2016). The EDPB recommends that organizations appoint a DPO specifically trained in the processing of genetic data, or at least a DPO with the necessary knowledge and expertise to provide advice on this topic.

The obligation to designate a DPO for processing genetic data is a key aspect of the GDPR's provisions on the protection of sensitive personal data. This requirement serves to ensure that organizations handling genetic data are held accountable for their data protection practices, and that individuals' privacy and rights are safeguarded.

4. GENETIC DATA PROTECTION ISSUES IN THE CONTEXT OF GDPR

4.1. Issues related to data sharing

The main objective of the GDPR is to protect the rights and freedoms of data subjects by giving individuals increased control over their personal data, including genetic data. It should be stressed that, as highlighted in this thesis, genetic data is unique and highly sensitive information that requires special safeguards to prevent misuse and unauthorised access. Although the GDPR lays down a number of requirements and provisions that data controllers and processors must comply with when processing data subjects' genetic data. Recital 6 of the GDPR states that: *"Rapid technological developments and globalisation have brought new challenges for the protection of personal data. The scale of the collection and sharing of personal data has increased significantly. Technology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Natural persons increasingly make personal information available publicly and globally. Technology has transformed both the economy and social life, and should further facilitate the free flow of personal data within the Union and the transfer to third countries and international organisations, while ensuring a high level of the protection of personal data".* Nevertheless, in practice, the protection of genetic data has become a critical issue in the era of big data and personal genomics. It should be noted that sharing of genetic data has become an important topic of debate in recent years, as genetic tests have become more readily available and the amount of data available has increased. On the one hand, sharing genetic data can be very useful for medical research and can help to better understand and treat genetic diseases. On the other hand, there are legitimate concerns about the privacy and security of genetic data, as well as the possible misuse of this sensitive information (Heuerman *et al.* 2017).

The following work of this the thesis will seek to outline the main issues related to the protection of genetic data.

The scope of personal data

The GDPR clearly defines „personal data" as data relating to an identified or identifiable individual. Article 4(1) GDPR states: "an identifiable natural person is a natural person whose identity can be determined directly or indirectly, in particular by means of attributes such as

name and surname, identification number, location data, identifier or one or more factors related to the physical, physiological, genetic, mental, economic, cultural or social identity of a natural person". The GDPR expanded the previously used catalog of special categories of personal data to include genetic data, biometric data when used to uniquely identify a natural person, and data related to criminal convictions and crimes" (Chassang, 2017).

The GDPR is extraterritorial in nature, which means that the rules of the Regulation, under certain conditions, apply to residents outside the European Union, accordingly, liability will be applied for their violation, regardless of where the company processing personal data is located, and which country's law applies to it. The first step for companies that are involved in some way with the use of personal data is to identify what personal data they collect and process. According to the GDPR, "personal data" is any information relating to a "data subject", i.e. an identified or identifiable natural person on the basis of which (based on one or more factors) such a person can be directly or indirectly identified. Even one factor is enough to clearly indicate the identity of this person (Peloquin *et al.*, 2020).

Regarding the scope of the GDPR, as already mentioned, it has an extraterritorial nature, i.e. is mandatory not only for EU residents, but also for non-residents related to personal data of EU citizens. The GDPR will apply to (Shabani *et al.*, 2015): (1) a company that collects and processes personal data, regardless of its registration or country of residence, if it operates in the EU, regardless of where exactly the work with personal data is carried out, i.e., the company's activities are carried out through any permanent structure in the EU: branch, representative office, partner (legal entity), agent/representative, etc.; (2) a company, regardless of the country of registration and residence, processing personal data of persons in the EU, activities related to the offering of goods and services to such persons, regardless of whether they are obliged to pay or not, or to monitor the behavior of such persons, if such behavior takes place in the EU.

It should be emphasised here that there are different views on when genetic data are sufficiently identifiable to be considered as 'personal data' and therefore fall within the scope of the GDPR (ARTICLE 29 Data Protection WP…2004). For example, it may assess whether the de-identified data is sufficiently anonymous or whether the genome sequences are substantially identifiable. It is also an assessment of whether combinations of genome and metadata are identifiable, or whether there is additional information that could identify the individual (ARTICLE 29 Data Protection WP…2004). This raises doubts as to whether it is

correct to consider anonymous genetic data as personal data. There is no unanimous opinion on this issue, which is controversial given that personal data are data from which a person can be identified, but anonymisation raises reasonable doubts as to the feasibility of this possibility.

4.2. Potential discrimination based on genetic data

The problem of misuse of genetic data is directly linked to the rapid development of technology and the increased availability of genetic information. The misuse of genetic data can therefore have serious consequences for individuals, communities and society as a whole.

One of the main problems associated with the misuse of genetic data is the potential for discrimination. For example, insurance companies may deny insurance coverage based on genetic information or charge higher premiums based on a person's genetic information (Joly *et al*., 2013). For instance, in 2008, the Genetic Information Nondiscrimination Act (GINA) was passed in the United States to protect individuals from genetic discrimination in employment and health insurance. However, GINA does not cover all forms of discrimination, such as life insurance and long-term care insurance (Feldman, 2012). It is important to highlighted that the Working Party is of the opinion that the processing of genetic data in the field of insurance should be prohibited and only allowed in genuinely exceptional circumstances, explicitly provided for in legislation (ARTICLE 29 Data Protection WP…2004).

Additionally, employers may use genetic information to make hiring and promotion decisions, leading to potential discrimination based on genetic predispositions (Chapman *et al*., 2020). The processing of genetic data in the field of employment should also be prohibited. In the Working Party opinion processing of such data should only be allowed in truly exceptional circumstances and in the light of the prohibition already in force in several Member States (ARTICLE 29 Data Protection WP…2004). In addition to discrimination in healthcare and employment, genetic data can also be used to discriminate against individuals based on their ancestry or race. This can occur using genetic ancestry testing, which has become increasingly popular in recent years. However, the interpretation and use of genetic ancestry testing results can be highly subjective and can lead to discrimination based on perceived genetic ancestry (Jorde *et al*., 2021).

In this context, the 2019 survey of respondents, which explored participants' willingness to share genetic data or biological samples for research purposes, could be mentioned. (M. E. Vidgen *et al.* 2020). The results showed that the majority of respondents wanted to be given the option to choose whether their genetic data from medical records would be used for research. Participants' expectations about whether they should seek consent for the use of their genetic data, and how often they should do so, also depended on whether the data were identifiable or anonymous. Respondents also expressed concerns about sharing genetic data which could lead to discrimination (M. E. Vidgen *et al.* 2020). In respect to the survey results, it can be seen that data subjects are concerned about whether genetic data will remain anonymous during research, which suggests that individuals do not want genetic data to be linked to them (in other words, that reading genetic data would allow the identification of a specific individual). Also, the disclosure of genetic data is associated by data subjects with potential discrimination against them. In this case, a correlation could be seen between ensuring the anonymity of data subjects and non-discrimination against data subjects. In particular, if the anonymity of genetic data is implemented, the possibility to identify a specific individual will be eliminated. Therefore, the presence or absence of discrimination is directly linked to the implementation of the anonymity of genetic data.

It can be concluded that discrimination in the processing of genetic data is a serious problem that needs to be addressed in order to ensure the ethical and fair use of genetic information. For data subjects, the occurrence of discrimination is directly linked to the lack of anonymity of genetic data. In view of the potential for discrimination against data subjects, it is not only necessary to strengthen the legal protection of genetic data, but also possible to carry out education and awareness-raising campaigns to promote the responsible use of genetic data.

Potential infringement of third parties' legitimate interests and rights

It can be argued that the sharing of genetic data without proper consent and without considering the rights and interests of the persons related with the data subject may be a contentious issue. While advances in genetic technologies have enabled extensive research and potential medical breakthroughs, the sharing of genetic data may violate the rights and interests of the data subject and those of persons linked to them.

First, it should be underlined that sharing genetic data without the proper consent of the data subject may violate the data subjects' right to privacy. Persons' genetic information is highly personal and sensitive, and sharing it without data subjects' knowledge or consent may violate his/her privacy. A study on Ethical concerns on sharing genomic data including patients' family members found that the collection, use and sharing of genetic data raises significant ethical issues, including the protection of personal privacy. (Takashima *et al*., 2018).

As stated above, the sharing of genetic data may also infringe the rights and interests of persons related to the data subject. This is because genetic information is often shared within families, and the sharing of genetic data of one person may affect the privacy and rights of his relatives. It could be argued that genetic information is not only important for the subject, but also for his or her relatives and descendants (ARTICLE 29 Data Protection WP…2004). This means that sharing a persons' genetic data without their consent could potentially affect the privacy and rights of their family members. Given the highly sensitive nature of this issue, the WP has noted that a balance must be struck between the right of the data subject not to disclose his or her genetic information and the potentially serious consequences that the disclosure and use of such information may have on members of the biological family (ARTICLE 29 Data Protection WP…2004). Despite these concerns, some argue that sharing genetic data can have valuable medical benefits. For example, it could facilitate the development of personalised treatments and improve our understanding of genetic diseases (Amorim *et al*., 2022).

However, despite the obvious medical benefits, there are potential risks for the persons related to the data subject. As analysed in the thesis, any collection of personal data (including genetic data) must have a lawful basis (lawfulness of processing), it is important to re-emphasise that genetic information is particularly sensitive and unique to each data subject, which is why the sharing of such data not only raises legal requirements, but also involves moral and ethical principles. Therefore, it is essential that any sharing of genetic data takes place with appropriate consent and in accordance with moral and ethical principles.

It is also should be stressed that genetic data can be collected not only from a living person, but also after his or her death, i.e., after a long period of time (Costello, 2022). These data can help to identify other data subjects related to the deceased person. In this context, "*genetic data can thus reduce privacy on a long-term basis and can reveal the genetic characteristics and relationships of an expanding group of individuals over successive*

generations" (Costello, 2022). This is arguably a significant ethical problem. First of all, a deceased person can no longer consent to the collection and use of his or her genetic information. As mentioned above, the sharing of genetic data may also infringe the interests of those related to the data subject. In this case, the practical question is how to properly implement the requirements of the Regulation when the data subject is no longer able to give consent. Although the death of the data subject removes the possibility of obtaining consent, the fact of death does not remove the possibility of extracting genetic information from the data subject, since, as Costello pointed out, the extraction of genetic information is also possible after a long period of time. Therefore, the genetic information of a deceased data subject can be considered as 'alive' simply because it may reveal a significant amount of information, not only about the data subject himself, but also about the persons related to him.

In conclusion, the sharing of genetic data without proper consent and consideration for the rights and interests of both the data subject and their related individuals can be a contentious issue. While advances in genetic technology have allowed for extensive research and potential medical breakthroughs, the sharing of genetic data can infringe on the rights and interests of the data subject and their related individuals. It is therefore important to ensure that proper consent is obtained and the rights and interests of both the data subject and their related individuals are safeguarded.

4.2. The problematic nature of consent to the processing of genetic data

In recent years, the issue of consent to the processing of genetic data has become increasingly problematic as technological advances have made it easier to collect and analyse large amounts of genetic information. While the use of genetic data for research and medical purposes can be beneficial, there are also serious questions about the ethics of collecting and using this sensitive information without the explicit consent of the individuals from whom it is collected.

Recital 40 of the GDPR states that: "*In order for processing to be lawful, personal data should be processed on the basis of the consent of the data subject concerned or some other legitimate basis…*" In this respect, it can be argued that consent is only one of the legal grounds on which personal data may be processed. Article 6 of the GDPR establishes six grounds for lawful processing: 1) on the basis of the data subject's consent; 2) on the basis of a contract to which the data subject is a party; 3) on the basis of a legal obligation imposed on the controller;

4) for the purpose of safeguarding the vital interests of the data subject or of another natural person; 5) for the purposes of the protection of the public interest or for the exercise of official authority vested in the controller; and 6) on the grounds of legitimate interests. Even though Article 6 of the GDPR provides for legitimate grounds for processing, genetic data fall within a special category of personal data (Article 9 of the GDPR).

Although Article 9(1) of the GDPR states that the processing of genetic data is prohibited, Article 9(2) of the GDPR provides for exceptions to this prohibition. Paragraph 2 (j) of analysing Article provides for the processing of special categories of personal data where necessary for archiving purposes in the public interest, for scientific or historical research purposes or for statistical purposes in accordance with Article 89(1). It can therefore be argued that the processing of genetic data is possible without the data subject's consent of the purpose of scientific research. Despite the exception, Recital 33 of the GDPR states that it is often not possible to fully determine the purpose of processing for research purposes. Also: "…*data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose".* According to EDPB: "*when special categories of data are processed on the basis of explicit consent, applying the flexible approach of Recital 33 will be subject to a stricter interpretation and requires a high degree of scrutiny"* (EDPB Guidelines on consent…2020). In the light of the provisions of the Regulation, it could be argued that there is a lack of clarity in this case, however, as to when the data subject will be required to give consent to the processing of genetic data in the context of research purposes, and when not.

It is also important to note that Article 9(4) of the GDPR provides that Member States may maintain or impose additional conditions, including restrictions on the processing of genetic data, biometric data, or health data. Although the GDPR provides requirements and rules for the use of personal data for research purposes, the law of the Member State is of great importance here. Although the GDPR is directly applicable and enforceable in all Member States, in the specific field of research, the GDPR provides only a few rules. Therefore, it is arguable that in the field of research, the rules still need to be determined by national or other EU law (Pormeister, 2018). It should be emphasised that the Regulation left many important aspects of data processing in the research field to the law of the Member States, and it is

therefore considered that in this case the GDPR did not harmonise the use of personal data in the research field across the EU, the reason for which is that the Member States are confronted with the problem of the application of their national law (Pormeister, 2018).

For instance, *Ancestry*[15] states that it uses personal data to personalise, improve, update, or extend its services. The list of uses of the data above includes scientific, statistical, and historical research[16]. Furthermore, Company provisions of the Privacy Statement provide that, after the processing of a biological sample sent by an individual, the data subject shall be given the opportunity to consent to the storage of that sample in Ancestry's Biobank for future research, subject to the data subject's informed consent to the research, or to any other consent to the testing of his or her own biological samples. As can be seen, the company indicates that the use of data subjects' genetic data for research purposes is subject to the data subject's consent. However, it is not clear from the rules what action is to be taken in cases where the data subject has not consented to the storage of the biological sample (i.e., neither consented nor objected to further storage of the biological sample). As Ancestry points out, if the data subject does not consent to the storage of the biological sample, they take action and delete the data. However, it is not clear what action the company takes when the data subject does not express any opinion on this matter.

Regarding the data subject's consent to the processing of genetic data for research purposes, it could be mentioned the Common Rule (45 CFR 46, Subpart A) which provides the basic elements of informed consent, which are relevant in the field of genomics. It also provides examples to guide the development of informed consent forms.

Main elements of informed consent: *1) purpose of the research; 2) description of the research; 3) risks; 4) benefits; 5) alternatives of the participation; 6) confidentiality; 7) potential benefits; 8) resources available in case of injury; 9) contact information; 10) voluntariness; 11) the collection of identifiable information or identifiable biospecimens; 12) compensation; 13) withdrawal from research; 14) use of biospecimens for commercial profit; 15) clinically relevant results; 16) whole genome sequencing.* According to the last element "whole genome sequencing" it could also be pointed that the NIH Genomic Data Sharing (GDS) policy also

[15] The world's largest for-profit genealogy company, it operates a network of genealogical, historical records and related genetic genealogy websites. For more information visit: https://www.ancestry.com/corporate/about-ancestry/our-story

[16] For more information visit: https://www.ancestry.com/c/legal/privacystatement

requires researchers generating large-scale genomic data to obtain consent for "*future research use and large-scale data sharing*". In this respect, it can be argued that researchers are held to a high standard when it comes to the implementation of informed consent. However, as mentioned earlier in this thesis, the GDPR provides that genetic data may be processed for research purposes without the data subject's consent. In this situation, Member States are free to lay down the relevant measures in their national legislation, leading to uncertainty and diverging practices between Member States.

In conclusion, Member States have a wide margin of discretion in the processing of genetic data for research purposes. In particular, Member States can do no further, i.e., follow the provisions of the GDPR in the processing of genetic data of data subjects and, secondly, impose additional conditions, such as stricter rules. In this case, it is not clear whether encouraging such practices is really the equitable approach. In the author's view, irrespective of the Member State in which the genetic information is processed, one key feature linking these data is data sensitivity. It is therefore considered that more attention should be paid to the processing of particularly sensitive data and to the harmonisation of practices. It is arguable that the application of different practices regarding genetic data poses a greater risk to the rights and freedoms of data subjects.

CONCLUSIONS

1. New technological breakthroughs or scientific advances have had a major impact on the development of genetic science and its increasingly sophisticated methods. Genetic science has provided the basis for understanding what genes are and what information they encode, as well as what is meant by the term "genetic data". Despite the advantages of using advanced technologies in the field of genetics, such as next-generation sequencing, personalised medicine, agricultural biotechnology, criminal justice etc., the handling of genetic data poses increasing ethical and practical challenges. One of the main concerns is privacy. With the increasing availability of genetic testing, there is a risk that sensitive genetic information may be accessed by unauthorised parties. This could have serious consequences for individuals, including discrimination. Another problem is the potential for misuse of genetic data. There have been cases where researchers have used genetic data for purposes other than those for which it was originally collected. This has led to calls for stricter regulation of the use of genetic data, including the need to obtain explicit informed consent from the participant. In addition, there are concerns about the accuracy and reliability of genetic data. Increasingly, direct-to-consumer genetic testing poses the risk of individuals receiving inaccurate or incomplete information. This may lead individuals to make decisions based on incorrect or incomplete information, with potentially serious consequences. Therefore, despite the value of significant technological advances, the aim must be to protect such data as much as possible, while ensuring the fundamental rights and freedoms of individuals.
2. Genetic data can be considered as personal data and stored accordingly, in compliance with the provisions of the GDPR and fulfilling certain criteria. Firstly, genetic data must be data relating to a natural person and must relate to genetic characteristics inherited or acquired by the individual which provide unique information about the physiology and health of that natural person. It must also be obtained in a specific way, i.e. by analysing a biological sample of that natural person. And finally, it must be able to identify the natural person directly or indirectly from the genetic data available.
3. The principles enshrined in the GDPR are of great importance in the context of the protection of genetic data. First, genetic data is highly sensitive and personal. It can reveal sensitive information about an individual's health, ancestry, and predisposition to certain diseases. As such, it is important that the processing of genetic data is carried out

in a way that respects the privacy and dignity of the individual concerned. The GDPR ensures that individuals have control over their own data and can make informed decisions about how it is used. Second, the GDPR sets out clear rules for the lawful processing of genetic data. It requires that data processors have a legal basis for processing genetic data, such as the explicit consent of the individual concerned or a legitimate interest. This helps to ensure that genetic data is not used for unauthorized or illegitimate purposes. Third, the GDPR requires that data processors implement appropriate technical and organizational measures to protect genetic data against unauthorized access, use, or disclosure. This is particularly important in the context of genetic data, as it is vulnerable to misuse or abuse if it falls into the wrong hands. Overall, the principles enshrined in the GDPR are essential for the processing of genetic data as they help to protect the privacy, dignity, and rights of individuals, and ensure that genetic data is used in a responsible and lawful manner.

4. While genetic data are health-related, they are too broadly interpreted in the context of health data. Therefore, the specificity of genetic data has given rise to a clear need to treat this category of data not as a 'part' of something, but as a separate, independent and specific category of data in the context of GDPR. In this case, the aim is to underline the special protection of these data. Genetic data are characterised by their sensitivity, which leads to the conclusion that sensitive personal data are subject to a higher regulatory burden than other categories of personal data. It is also important to underline that both the legal and technological landscape is constantly evolving, which makes the protection of genetic data particularly important.
5. Data controllers and processors are required to implement a range of organisational and technical measures to ensure that the genetic data they process meet the requirements of confidentiality, integrity, availability and resilience. In order to comply with these requirements, controllers and processors (e.g. healthcare institutions) must ensure the security of the systems they use, as this is directly related to the security of genetic data. One such measure is the implementation of strong security measures, such as encryption and access controls, to prevent unauthorized access to genetic data. Additionally, controllers and processors should have policies in place to ensure that only authorized personnel have access to genetic data, and that the data is only used for the purposes for which it was collected. Another important measure is the introduction of robust data

management systems, including data storage, retention and destruction processes. This helps to ensure that genetic data is properly organised and stored and that it is not kept longer than necessary. Also, controllers and processors should have strong privacy policies in place, including measures to ensure that individuals are informed about how their genetic data will be used and have the ability to give their consent or object to its use. Overall, the implementation of these organisational and technical measures is essential for the proper processing and protection of genetic data and helps to ensure that individuals' privacy and rights are respected.

6. Discrimination in the handling of genetic data is a serious problem that needs to be addressed to ensure the ethical and fair use of genetic information. Discrimination against data subjects is directly linked to the lack of anonymity of genetic data. In view of the potential discrimination against data subjects, it is not only necessary to strengthen the legal protection of genetic data, but also possible to carry out education and information campaigns to promote the responsible use of genetic data. Discrimination in the processing of genetic data is a major problem with potentially far-reaching consequences for individuals and society as a whole. Genetic information is increasingly being used in decision-making on everything from employment and insurance to healthcare and criminal justice. However, if not handled ethically and fairly, this information can lead to discrimination against certain groups of people. For example, if genetic information is used to discriminate against people with certain inherited diseases, it can perpetuate negative stereotypes and create barriers to employment, access to health care and other important resources. In addition, if certain groups do not participate in genetic testing, or do not have equal access to genetic testing and genetic information, this can perpetuate existing health inequalities and contribute to persistent discrimination. It is important to address and prevent discrimination in the handling of genetic data to ensure that all individuals are treated fairly and have equal access to genetic science.

7. The sharing of genetic data without proper consent and without taking into account the rights and interests of the data subject and those of the persons concerned may be a contentious issue. While advances in genetic technologies have enabled extensive research and potential medical breakthroughs, the sharing of genetic data may violate the rights and interests of the data subject and of the persons concerned. It is therefore

important to ensure that appropriate consent is obtained and that the rights and interests of both the data subject and the persons concerned are protected.

8. Member States have a wide margin of discretion in processing genetic data for research purposes. In particular, Member States may not only comply with the provisions of the GDPR when processing the genetic data of data subjects, but also, secondly, impose additional conditions, such as stricter rules. It is therefore not clear whether encouraging such practices is really the right approach. It is considered that, irrespective of the Member State in which the genetic information is processed, one of the main features linking these data is the sensitivity of the data. It is therefore arguable that more attention should be paid to the processing of particularly sensitive data and to the harmonisation of practices, as the application of different practices in relation to the processing of genetic data poses a greater risk to the rights and freedoms of data subjects.

LIST OF SOURCES

Special literature

1. Amorim M., *et al.* (2022). Benefits and Risks of Sharing Genomic Data for Research: Comparing the Views of Rare Disease Patients, Informal Carers and Healthcare Professionals. *Int. J. Environ Res. Public Health,* 19(14), https://doi.org/10.3390%2Fijerph19148788
2. Bahr A., Schlünder I. (2015). Code of practice on secondary use of medical data in European scientific research projects. *International Data Privacy Law*, 5 (4), 279–291, https://doi.org/10.1093/idpl/ipv018
3. Bamshad M. J. *et al.* (2003) Human Population Genetic Structure and Inference of Group Membership. *Am J Hum Genet.* 72(3): 578–589, https://doi.org/10.1086%2F368061
4. Berryessa C. M., Cho M. K. (2013). Ethical, Legal, Social, and Policy Implications of Behavioral Genetics. *Annu Rev Genomics Hum. Genet.* (14) 515-534, https://doi.org/10.1146%2Fannurev-genom-090711-163743
5. Bieker F. *et al.* (2016). A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation. *Privacy Technologies and Policy*, 9857 21-37, https://doi.org/10.1007/978-3-319-44760-5_2
6. Chapman C. R., *et al.* (2020). Genetic discrimination: emerging ethical challenges in the context of advancing technology. *Journal of Law and the Biosciences,* 7(1), https://doi.org/10.1093/jlb/lsz016
7. Chassang G. (2017). The impact of the EU general data protection regulation on scientific research. *Ecancermedicalscience,* 11(709), *https://doi.org/10.3332%2Fecancer.2017.709*
8. Chen X., Ishwaran H. (2012). Random forests for genomic data analysis. *Genomics*. (6) 323-329, https://doi.org/10.1016/j.ygeno.2012.04.003
9. Christopher K., *et al*. (2020). The EU General Data Protection Regulation (GDPR): A Commentary.

10. Clayton E. W. *et al*. (2019). The law of genetic privacy: applications, implications, and limitations (2019). *Journal of Law and the Biosciences*, 6,(1), 1–36, https://doi.org/10.1093/jlb/lsz007
11. Coppedè F. *et al.* (2019). Association of Polymorphisms in Genes Involved in One-Carbon Metabolism with MTHFR Methylation Levels. *International Journal of Molecular Sciences*. 20(15), https://doi.org/10.3390/ijms20153754
12. Costello R. A. (2022). Genetic Date and the Right to Privacy: Towards a Relational Theory of Privacy? *Human Rights Law Review*, 22(1), https://doi.org/10.1093/hrlr/ngab031
13. Daigle B., Khan M. (2020). The EU General Data Protection Regulation: An Analysis of Enforcement Trends by EU Data Protection Authorities. *The EU General Data Protection Regulation: An Analysis of Enforcement Trends by EU Data Protection Authorities.* Available at: *https://www.usitc.gov/staff_publications/jice/eu_general_data_protection_regulation_analysis*
14. Elhaik E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports.* 12, https://doi.org/10.1038/s41598-022-14395-4
15. Feldman E. A. (2012). The Genetic Information Nondiscrimination Act (GINA): Public Policy and Medical Practice in the Age of Personalized Medicine. *Journal of General Internal Medicine*, 27, 743–746, https://link.springer.com/article/10.1007/s11606-012-1988-6.
16. Godard B. *et al.* (2003). Genetic information and testing in insurance and employment: technical, social and ethical issues. *European Journal of Human Genetics.* (11)– 123-142, https://doi.org/10.1038/sj.ejhg.5201117
17. Goh. G., Choi M. (2012). Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. *Genomics&Informatics*, 10(4), 214–219. https://doi.org/10.5808%2FGI.2012.10.4.214
18. Heuerman T., *et al.* (2017). Open sharing of genomic data: Who does it and why? *PLoS One*, 12(5), https://doi.org/10.1371%2Fjournal.pone.0177158

19. Hindorff L. A. *et al.* (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS*.106(23):9362-7. https://doi.org/10.1073/pnas.0903103106
20. Hintzshe J. D. *et al.* (2016). A Survey of Computational Tools to Analyze and Interpret Whole Exome Sequencing Data. *Int J Genomics*. https://doi.org/10.1155%2F2016%2F7983236
21. Januševičienė J. (2018). Praktiniai asmens sveikatos duomenų tvarkymo aspektai pagal Bendrąjį asmens duomenų apsaugos reglamentą. *Teisė*, 107, https://doi.org/10.15388/Teise.2018.107.11826
22. Joly Y., Feze I. N., Simard J. (2013). Genetic discrimination and life insurance: a systematic review of the evidence. *BMC Medicine*, 11(25), https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-11-25
23. Jorde L. B., Bamshad M. J. (2021). Genetic Ancestry Testing What Is It and Why Is It Important? *Jama*, 323(11), 1089-1090, https://doi.org/10.1001%2Fjama.2020.0517
24. Lazauskienė R., Tamulionienė, D. (2020). Asmens duomenų tvarkymo ypatumai nuotoliniu būdu teikiant paslaugas sveikatos priežiūros srityje. *Jurisprudencija*. 27 (2), https://repository.mruni.eu/handle/007/17206?locale-attribute=lt
25. Mathur S., Sutton J. (2017). Personalized medicine could transform healthcare. *Biomedical reports*, 7(1), 3–5, 10.3892/br.2017.922
26. McKenna A. *et al.* (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research.* 20(9), 1297–1303, https://doi.org/10.1101%2Fgr.107524.110
27. Mehrabani S. Z. N. (2019). Association of SHMT1, MAZ, ERG, and L3MBTL3 Gene Polymorphisms with Susceptibility to Multiple Sclerosis. *Biochemical genetics*, 3(57), 355-370, https://doi.org/10.1007/s10528-018-9894-1.
28. Meler E., Sisterna S., Borrell A. (2020). Genetic syndromes associated with isolated fetal growth restriction. *Prenat Diagnosis*. 40(4), 432-446, https://doi.org/10.1002/pd.5635
29. Montagu M. V. (2020). The future of plant biotechnology in a globalized and environmentally endangered world. *Genetics and Molecular Biology*, 43(1.2), https://doi.org/10.1590%2F1678-4685-GMB-2019-0040

30. Niu Z., *et al.* (2017). Association of MTHFR, MTRR and MTR polymorphisms with breast cancer risk: a study in Chinese females. *International Journal of Clinical and Experimental Pathology*. 10(6), 7059-7066, https://doi.org/10.7754/clin.lab.2016.160917
31. Peloquin D., *et al.* (2020). Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics,* 28, 697–705, https://doi.org/10.1038/s41431-020-0596-x
32. Petkevičienė V., Pakutinskas P., Bitė V. (2020). Asmens duomenų tvarkymo iššūkiai COVID-19 pandemijos metu. *Jurisprudencija.* 27(2), 330-345, https://repository.mruni.eu/bitstream/handle/007/17204/6362-15294SM.pdf?sequence=1&is
33. Pormeister K. (2018). Genetic research and applicable law: the intra-EU conflict of laws as a regulatory challenge to cross-border genetic research. *Journal of Law and the Biosciences*, 5(3), 706–723, https://doi.org/10.1093/jlb/lsy023
34. Quinn P., Quinn L. (2018). Big genetic data and its big data protection challenges. *Computer Law & Security Review*, 34(2), 1000-1018, https://doi.org/10.1016/j.clsr.2018.05.028
35. Rebekah P.K. *et al.* (2017). Interaction between Maternal and Paternal SHMT1 C1420T Predisposes to Neural Tube Defects in the Fetus: Evidence from Case–Control and Family-Based Triad Approaches. *Birth Defects Research*, 109(13), 1020-1029, https://doi.org/10.1002/bdr2.23623
36. Roche M, I, Berg J. S. (2015). Incidental Findings with Genomic Testing: Implications for Genetic Counseling Practice. *Current Genetic Medicine Reports*, 3, 166-176, https://link.springer.com/article/10.1007/s40142-015-0075-9
37. Senf A. *et al*. (2021). Crypt4GH: a file format standard enabling native access to encrypted data. *Bioinformatics*, 37(17), 2753–2754, https://doi.org/10.1093/bioinformatics/btab087
38. Shabani M., Borry P. (2017). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *European Journal of Human Genetics*, 149–156, https://doi.org/10.1038/s41431-017-0045-7

39. Shabani, M., Knoppers, B. M., Borry, P. (2015). From the principles of genomic data sharing to the practices of data access committees. *EMBO Mol Med.*, 7, 507-509, https://doi.org/10.15252/emmm.201405002
40. Sharma D., Sharma P., Shastri S. (2016). Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clinical Medicine Insights: Pediatrics*, 10, https://doi.org/10.4137/CMPed.S40070
41. Sharma D., Sharma P., Shastri S. (2017). Genetic, metabolic and endocrine aspect of intrauterine growth restriction: an update. *The Journal of Maternal-Fetal & Neonatal Medicine*. 30(19), 2263-2275, https://doi.org/10.1080/14767058.2016.1245285
42. Štareikė E., Kausteklytė-Tunkevičienė S. (2018). Pagrindinės duomenų subjekto teisės ir jų užtikrinimas pagal ES Bendrąjį duomenų apsaugos reglamentą. Visuomenės saugumas ir viešoji tvarka.
43. Sukhorolskyi P., V. Hutsaliuk V. (2020). Processing of Genetic Data under GDPR: Unresolved Conflict of Interests. Masaryk University Journal of Law and Technology, 14(2), 151–176, https://doi.org/10.5817/MUJLT2020-2-1.
44. Takashima K., *et al.* (2018). Ethical concerns on sharing genomic data including patients' family members. *BMC Medical Ethics*, 19(61), https://doi.org/10.1186/s12910-018-0310-5
45. Verma M. (2012). Personalized medicine and cancer. *Journal of Personalized Medicine*, 2(1), 1–14, https://doi.org/10.3390%2Fjpm2010001
46. Vidgen M. E. *et al.* (2020). Sharing genomic data from clinical testing with researchers: public survey of expectations of clinical genomic data management in Queensland, Australia. *BMC Medical Ethics*, 21(119), https://doi.org/10.1186/s12910-020-00563-6
47. Yamamoto Y. *et al.* (2022). Current Status, Issues and Future Prospects of Personalized Medicine for Each Disease. *Journal of Personalized Medicine*. 12(3), 444. 10.3390/jpm12030444
48. Zaleskis J. (2019). Bendrasis duomenų apsaugos reglamentas ir asmens duomenų apsaugos teisė.

Legal acts

European Union legislation

49. 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. L 281, 23/11/1995.
50. Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Adopted by the Committee of Ministers on 19 November 1996.
51. Charter of Fundamental Rights of the European Union, published by the European Parliament, the Council of Ministers and the European Commission on 26 October 2012 (2012/C 326/391).
52. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Lithuania legal acts

53. Valstybinės duomenų apsaugos inspekcijos įsakymas dėl duomenų tvarkymo operacijų, kurioms taikomas reikalavimas atlikti poveikio duomenų apsaugai vertinimą, sąrašo patvirtinimo (2019). *TAR*, 4104.

US Legislation

54. The Genetic Information Nondiscrimination Act of 2008. Available at: https://www.eeoc.gov/statutes/genetic-information-nondiscrimination-act-2008
55. Code of Federal Regulations. Available at: https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46

***Soft law* sources**

56. ARTICLE 29 Data Protection Working Party (2004) Working Document on Genetic Data. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp91_en.pdf.
57. ARTICLE 29 Data Protection Working Party Guidelines on transparency under Regulation 2016/679. Adopted on 11 April 2018.
58. ARTICLE 29 Data Protection Working Party. Advice paper on special categories of data ("sensitive data") 20/04/2011. Available at: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf
59. ARTICLE 29 Data Protection Working Party. Guidelines on Personal data breach notification under Regulation 2016/679. Available at: https://ec.europa.eu/newsroom/article29/items/612052
60. ARTICLE 29 Data Protection Working Party. Opinion 03/2013 on purpose limitation. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf
61. ARTICLE 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques. Available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
62. ARTICLE 29 Data Protection Working Party. Overview of results of public consultation on Opinion on legitimate interests of the data controller (Opinion 06/2014). Available at: https://ec.europa.eu/justice/article-29/press-material/public-consultation/notion-legitimate-interests/files/20141126_overview_relating_to_consultation_on_opinion_legitimate_interest_.pdf
63. ARTICLE 29 Data Protection Working Party. Guidelines on Data Protection Officers ('DPOs'). 2016. Available at: https://ec.europa.eu/newsroom/article29/items/612048
64. European Data Protection Board Guidelines 05/2020 on consent under Regulation 2016/679. Available at: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

65. European Data Protection Board Guidelines 3/2018 on the territorial scope of the GDPR, 12 November 2019. Access on 5 November, Available at: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_3_2018_territorial_scope_after_public_consultation_en_1.pdf
66. European Data Protection Supervisor Opinion 8/2016. EDPS Opinion on coherent enforcement of fundamental rights in the age of big data Available at: https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_en.pdf
67. European Union Agency For Network and Information Security. (2017). Handbook on Security of Personal Data Processing, DOI 10.2824/569768
68. NIH Guidance on Consent for Future Research Use and Broad Sharing of Human Genomic and Phenotypic Data Subject to the NIH Genomic Data Sharing Policy. Available: at: https://osp.od.nih.gov/wpcontent/uploads/2015/08/NIH_guidance_elements_consent_under_gds_policy.pdf
69. Recommendation No R (97)5 of the Committee of Ministers of the Council of Europe of 13 February 1997. https://rm.coe.int/cmrec-97-5-on-the-protection-of-medical-data/1680a43b64
70. International Declaration on Human Genetic Data of the 32nd UNESCO General Conference of 16 October 2003.

Case law

71. *Gaughran v. the United Kingdom* [ECtHR], No. 45245/15, [13/06/2020]. https://hudoc.echr.coe.int/fre?i=001-200817
72. *S. and Marper v. United Kingdom* [ECtHR], No 30562/04 and 30566/04, [04/12/2008]. https://rm.coe.int/168067d216

Other sources

73. Global Alliance for Genomics and Health (GA4GH) Encryption Standard (21 Oct 2019). Available at: https://samtools.github.io/hts-specs/crypt4gh.pdf
74. Information Commissioner's Office. Guide to the General Data Protection Regulation (GDPR)/Security. Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/security/#10

75. Information Security Standard – ISO/IEC 27001:2022.
76. National Human Genome Institute. *Privacy in Genomics.* Available at: https://www.genome.gov/about-genomics/policy-issues/Privacy
77. National Human Genome Research Institute. *Genomic Data Science*. Available at: https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science
78. Tattersfield K. (2017). How universities have to adapt under the new EU General Data Protection Regulation (GDPR). Online access: https://www.fullfabric.com/articles/how-universities-have-to-adapt-under-the-new-eu-general-data-protection-regulation-gdpr

SUMMARY

Protection of genetic data under EU General Data Protection Regulation

The Master's thesis focuses on the application of the EU General Data Protection Regulation in the context of genetic data protection. The first part of the thesis analyses and explains the concept of genetic data from a historical perspective, highlighting and underlining the most important aspects from both the legal and the medical point of view. This part also analyses the approach to technology and its impact in the context of ensuring the protection of genetic data, highlighting the main issue related to the use of technology in the processing and protection of genetic data - ensuring the privacy of individuals. It also provides an analysis of the genetic code and its significance in the context of genetic data protection. The second part analyses the definition of genetic data enshrined in the GDPR in the light of the definition of personal data and its inherent characteristics. It also discusses the main principles for the processing of personal data enshrined in the GDPR, providing an analysis of each of them and their practical applicability in the context of the processing and protection of genetic data. At the same time, it focuses on the importance of genetic data as a special category of personal data and provides an analysis of the safeguards applicable to the processing of genetic data, which underpin the fundamental rights of data subjects. The second part analyses the definition of genetic data in the GDPR in the light of the definition and characteristics of personal data. It also discusses the main principles for the processing of personal data enshrined in the GDPR, providing an analysis of each of them and their practical applicability in the context of the processing and protection of genetic data. At the same time, it highlights the importance of genetic data as a special category of personal data. It also provides an analysis of the safeguards applicable to the processing of genetic data, underpinning the fundamental rights of data subjects, highlighting, and elaborating on each of the safeguards and detailing their practical applicability in the context of the processing of genetic data and the safeguarding of its protection. The last part of the thesis analyses issues related to the protection of genetic data. In particular, it identifies and analyses the problem of genetic data sharing, providing a general approach as well as concrete practical examples and analysis. It further identifies and analyses the issue of consent to the processing of genetic data for research purposes.