TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Social Sciences
Ragnar Nurkse School of Innovation and Governance

Liis Loorits

INTEGRATING CLINICAL AND GENOMIC DATA IN ESTONIAN HEALTH INFORMATION SYSTEM: TECHNOLOGICAL, LEGAL, ETHICAL AND SOCIAL PERSPECTIVES

Master's Thesis

Supervisor: Dr. Tarmo Kalvet

Tallinn 2016

I hereby declare that I am the sole author
of this master's thesis and it has not been
presented to any other university for examination.

Author
……(date) 2016

The master's thesis meets with the established requirements

Supervisor Tarmo Kalvet
…….(date) 2016

Accepted for examination…….(date) 2016
Board of examiners of Technology Governance master's theses
Prof. Dr. Robert Krimmer

Abstract

The use of genetic data gained wider attention when the Human Genome Project (HGP) was reaching its end. Furthermore, Estonia was among the first countries to launch a population based biobank and at the end of 2014, it was seen that it is possible to use the data in hospitals. This integration raises several technological, legal, ethical and social questions in theory and in practice. The current Estonian situation was studied in terms of these issues and the results confirmed that there is a need for standardisation guidelines for data storing and a technological solution for integrating unstructured data. The ethical issues of how closing the genetic data by a person affects their closest relatives and if a person should be compensated when their data is used need further discussion. Furthermore, there is a need for training and education for the healthcare professionals. Overall, it can be said that with the positive attitudes towards genomic medicine, existing nationwide IT infrastructures and the population based genome bank, Estonia is in a good position for realising the integration in the coming years.

Table of contents

Abbreviations

A – Adenine

ABMS – American Board of Medical Specialties

ACA – Affordable Care Act

ACT – Adult Changes in Thought

bp – Base Pairs

C – Cytosine

CAA – Computer-assisted Analysis

CAB – Community Advisory Board

CDC – Centers for disease Control and Prevention

CLIA – Clinical Laboratory Improvement Amendments

CSPro – Certified Service Providers

DNA – Deoxyribonucleic Acid

DTC – Direct To Consumer

EGCUT – Estonian Genome Center at the University of Tartu

EGP – Estonian Genome Project

EHIF – Estonian Health Insurance Fund

EHIS – Estonian Health Information System

EHR – Electronic Health Record

eMERGE - Electronic Medical Records & Genomics

EMR – Electronic Medical Record

ENCODE – Encyclopedia of DNA Elements

EPR – Electronic Patient Record

EU – European Union

FDA – Food and Drug Administration

G – Guanine

GINA – Genetic Information Nondiscrimination Act

GVF – Genomic Variation Format

HGNC – Human Gene Nomenclature Committee

HGP – Human Genome Project

HGVS – Human Genome Variation Society

HIT – Health Information Technology

HL7 – Health Level Seven International

HTA – Health Technology Assessment

ICT – Information and Communications Technology

IGMP – Implementation of Genomic Medicine Project

IP-A_CFs – Informed Permission, Assent and Consent Forms

IRB – Institutional Review Board

LOINC – Logical Observation Identifiers Names and Codes

NGS – Next-generation sequencing

NHIS – National Health Information System

NIH – National Institutes of Health

NSGC – National Society of Genetic Counselors

PHG – Public Health Genomics

PM – Personalised Medicine

SBS – Sequencing By Synthesis

SNOMED – Systematized Nomenclature of Medicine

STACC – Software Technology and Applications Competence Centre

T – Thymine

TIS – *Tervise infosüsteem*

US – United States

VHI – Voluntary Health Insurance

VUS – Variant of Uncertain Significance

WGS – Whole Genome Sequencing

WHO – World Health Organization

XML – Extensible Markup Language

1 Introduction

Combining different types of health related data of a population is already a huge asset, as argued generally. Whereas, adding data from emerging technologies, like molecular biology, creates a completely new value for the dataset. The information generated by combining different datasets can be used to diagnose diseases and identify new treatments or more specific drug targets (European Commission, 2015). Furthermore, genetic data is already used in the clinic with electronic health records (EHR) in certain settings, with one of the biggest initiatives being the Electronic Medical Records & Genomics (eMERGE) network in the US (Gottesman *et. al.* 2013). Other initiatives worth mentioning include the POLARIS programme in Singapore and Japan Implementation of Genomic Medicine Project (IGMP) (Milani *et. al.* 2015, p. 195). Moreover, technology capable of determining all protein coding DNA sequences in a persons genome is already a clinically validated method in Estonia since 2014 and routinely used in hospitals (Metspalu, 2016).

Estonia with Iceland has been among the first countries to establish a population based biobank in 1999 and as of 2015, the Estonian one is maintained by government under the University of Tartu as the Estonian Genome Centre at the University of Tartu (EGCUT) (Milani *et. al.* 2015, p. 188). Furthermore, at the end of 2014 the Estonian Ministry of Social Affairs presented the Estonian Government Pilot Project on Personalised Medicine for 2015-2018 (Kalda *et. al.* 2015). Whereas in the context of the Estonian personalized medicine pilot project, the report by Kalda *et. al.* 2015 defines personalized medicine as:

"Personalised medicine refers to prevention, diagnosis and treatment of health disorders, based on individual risk-tailored approach using computational decision support analysis of person's phenotype and genotype data. The goal of personalized medicine is to contribute towards preventive, predictive and participatory health system." (Kalda *et. al.* 2015, p 13).

Moreover, Estonia has identified eHealth system and integrating genomic data to it, as one of the smart specialization perspectives in health technology sector (Eesti Arengufond, 2014, p 21).

The thesis analyses the technological, ethical, legal and social issues that have risen in theory and in practice. These issues are transferred to the Estonian context and the thesis will aim to answer to the following research questions:

- What kind of technological, legal, social and ethical issues are related to integrating genomic data with clinical?
- What is the situation in Estonia towards the integration of genomic data with clinical in the EHIS?

The main goal for the paper is to determine the nature of current situation in Estonia related to integration of health information to genome data. Furthermore, based on the research already carried out on the subject of genetic data, its integration to EHRs and technology management in Estonia, a hypothesis can be set for the current paper: *Since by 2016, the technological issues are replaced with softer issues like social acceptance, Estonia's success in generating public acceptance for the use of new technological solutions can serve as an important advantage.*

The data collection for the study can be divided into three phases. Firstly, an analysis of research done by scholars on similar topics was conducted. Secondly an analysis was conducted on the studies done by Estonian experts for the Estonian Government Pilot Project on Personalised Medicine for 2015-2018 and opinions of Estonian experts from the fields of medical research, medical research ethics and medicine. The resulting information was used to compare the situation of Estonia to the rest of the world. Thirdly, in order to get further insight to the results of the analysis, a phone interview was carried out with the director of Estonian Genome Center professor Andres Metspalu and an interview was also carried out with Estonian eHealth expert professor Peeter Ross.

The analysis on the topic is presented in five chapters. The first chapter gives an overview on the importance of the topic and of the current paper. The second chapter gives a historical overview on the development of the technology and different theoretical and practical issues that the technological change has given rise to. The third and core chapter gives an overview of the current situation in Estonia on the basis of analyses conducted by Estonian experts. The fourth chapter links the theoretical and empirical parts and presents a discussion over the validity of results. Finally, the main findings and resulting suggestions are re-emphasised in the final chapter.

2 Theoretical framework

Integrating genomic data with clinical data in the EHIS can be classified as a "wicked" issue because it is an extremely complex subject (Pikani *et. al.* 2015, p. 30). In order to analyse different aspects of it, a comprehensive view is necessary and the following theoretical framework will give an overview of numerous technological, ethical, legal and social issues rising from it.

2.1 Emergence of the integration of genome data with electronic health records

In 1990, the US Department of Energy announced a project with an ambitious goal to sequence the whole human genome (Cordero & Ashley, 2012, p. 1001). The project brought with it a technological race between public and private entities, for being the first to announce a draft of a first full human genome. As a result of the competition between the Human Genome Project (HGP) and the commercial effort by Celera, two useful drafts of the human genome were published in 2001 (Venter *et. al.* 2001; Lander *et. al.* 2001), that was years before the initial deadline of 2005 (Shendure *et. al.* 2004, p. 336). Furthermore, after the ambitious milestone was reached in 2001, the sequencing industry was further accelerated and this has resulted in the gradual drop in costs (Cordero & Ashley, 2012, p. 1001). The sequencing cost of the first human genome was around US $300 million (Shendure *et. al.* 2004, p. 336), whereas by October 2015 the sequencing cost per genome had dropped to about US $1200 (National Human Genome Research Institute, 2015).

Already before the project for sequencing the whole human genome was announced in 1990, researchers had realized that with sufficient effort they could uncover the code that holds the key to an organisms physical traits (Cordero & Ashley, 2012, p. 1001). An individuals' genome is stored across 46 chromosomes consisting of 3 billion DNA base pairs (bp). Whereas Whole Genome Sequencing (WGS) is a laboratory process for determining most, if not all of these base pairs making up an individual's genome (Vassy *et. al.* 2014, p. 2).

During the sequencing of the first human genome, researchers were astonished about how few traditional genes encoding proteins were scattered across these 3 billion base pairs. When at first it had been estimated 100 000 genes to be present, the initial analyses found about 25 000 (Lander *et al.,* 2001, p. 931). Whereas the number has decreased since, reaching 21 000 by 2012 and at first it seemed that in between were megabases of "junk" DNA. However, by the year

2012, in the course of a decade long project Encyclopedia of DNA Elements (ENCODE), it had been found that about 80% of the human genome serves some purpose, biochemically speaking (Pennisi, 2012, p. 1159).

By 2013, it was known that only about 1-1.5% of the human genome is represented by protein coding sequences, that is 50 million base pairs of the 3 billion (Samuels *et. al.* 2013, p. 593). Since a lot of information about the whole human genome is still uncovered, clinical decision-making based on the data is a complex task. When there is no data to support the role of a particular genetic variant in the disease phenotype, it is called a variant of uncertain significance (VUS). In this case, the VUS might aid with clinical decision making when research supporting association is published and this has been used in practice for example in the Genomics Medicine clinic of the Medical College of Wisconsin, USA. As a result, WGS has enabled to improve the rate of diagnosis to 27% for rare or undiagnosed diseases (Jacob *et. al.* 2013, p. 2). However, returning whole genome and whole exome (all protein coding areas of the genome) sequencing results to clinicians and patients remains under debate (Green *et al*., 2013a), since less than 2% of adults have relevant actionable findings from whole exome sequencing (Dorschner *et. al*. 2013). At the same time it should be considered that every fully sequenced genome provides a new reference for an individuals family and the general population (Jacob *et. al.* 2013, p. 4).

Today, already some patients and physicians in certain settings have access to WGS (Worthey *et. al.* 2011; Egan *et. al.* 2012; Altman, 2013; Green *et. al.* 2013b; Jacob *et. al.* 2013), but its impact on health-care utilization, patient well-being and clinical decision-making remains largely unstudied (Vassy *et. al.* 2014, p. 1). As for rare diseases, the economic case for WGS is relatively apparent because patients and their families would have to go see different physicians, from one hospital to another and have numerous expensive tests. In the case of common diseases like type 2 diabetes, cancer or heart diseases, the cost-effectiveness of WGS is questionable. This due to the probabilistic nature of genomic data which brings about the need of follow-up analyses for secondary findings that increase the costs (Jacob *et. al.* 2013, p. 4). Furthermore, as the need for knowledgeable professionals for interpreting the test results is emphasized, it is perceived that low volumes of applicable cases would not make hiring such professionals cost-effective (Hamilton *et. al.* 2014, p. 241).

WGS can also expand the accuracy of family history in preventive care, because everyone has a risk to develop at least one disease (Jacob *et. al.* 2013, p. 4). Nevertheless, WGS has been in the past regarded as not economical in prevention because many people have to be screened to find the few at risk (Cohen *et. al.* 2008). Decision makers and payers show increased demand for

cost-effectiveness data in fields like pharmacogenomics related test-treatment approaches (Faulkner *et. al.* 2012, p. 1167). Whereas generally the methods of economic evaluation in personalized medicine (PM) have no significant difference from the standard cost-effectiveness analysis with captured health gain and cost-offsets, there are additional ways of creating value for the patient that can be identified. For example the "value of knowing" – reducing the patients uncertainty about the outcome of the treatment, improving adherence and through that results of the treatment and thirdly raising overall utilization at a population level (Towse & Garrison 2013, p. S40). Furthermore it has been noted that there is evidence to support the "added value from information" to clinicians and patients (Payne & Annemans, 2013, p. S34) but the fact that the effect of preventative testing might be a restrict for a subgroup and reduce the absolute quality-adjusted life-year gains from treatment is also brought out. This meaning that although targeting is cost-effective, some patients for whom the drug would also have been effective do not receive treatment because of the false positives and patients misclassified as nonresponders (Towse & Garrison 2013, p. S40). Consequently there is an ongoing need for identifying the best practices for economic modeling in fields like pharmacogenomics. (Faulkner *et. al.* 2012, p. 1167).

In his classic work the *Structure of Scientific Revolutions* Thomas Kuhn has brought out that for scientific progress the key drivers are not dramatic advances, like the sequencing of the human genome, but instead the incremental learnings which illuminate the shortcomings of existing scientific dogma that in turn will prepare the ground for revolutionary change. Today in this incremental process that the advances of DNA sequencing technology have brought about, there is a potential for a scientific revolution in which health economics research professionals in academia, industry and payer systems will play a central role (Kuhn, 1962 cited in O'Donnell, 2013, p. S2).

2.2 Technological barriers

The data set obtained from WGS is large, up to about 1 terabyte per genome, which means the need for computer-assisted analysis (CAA) (Jacob *et. al.* 2013, p. 3). As of today, biological research has transformed into a big data science, with the emergence of next-generation sequencing (NGS) technologies enabling high throughput sequencing. As a result many computer scientists and biologists are emphasizing the urgent need for computing power, storage and bioinformatics software to analyze large quantities of sequence data (Kwon *et. al.* 2015, p. 490). Furthermore, storage development cannot keep up with NGS data production, since the

cost of NGS data production is reduced by half every 5 months and in contrast the cost of data storage is reduced by half every 14 months (Baker, 2010, p. 495).

As a result of the revolution in molecular biology brought about by NGS technologies, bioinformaticians nowadays have to interact with e-infrastructure consisting of high-performance computing. It is common for data analysis to include multiple software tools, used in sequential manner on input data (Spjuth *et. al.* 2015, p. 1). For executing these workflows and pipelines, local clusters or distributed computing clouds are used (Bux *et. al.* 2013). Whereas, in academia these high-performance resources usually consist of compute clusters with Linux operating system (Spjuth *et. al.* 2015, p. 2), but recently cloud computing has emerged as an additional technology (Kwon *et. al.* 2015).

In addition to potentially being very large, next-generation sequencing data has uneven depth of coverage. This brings about new challenges in storing the information based on the quality and coverage of the data (Masys *et. al*. 2012, p. 420). Furthermore, because before very recent developments, sequencing techniques have been partial and not fully accurate, it is expected that one patients DNA will be sequenced more than once in their lifetime (Ury, 2013, p. 781). Since the first human genome project a variety of different technological solutions for NGS have been realized in a commercial product. For example, the 454 Genome Sequencer by Roche Applied Science based in Basel Switzerland was the first successful next generation system. It uses a pyrosequencing technology and was the market leader in the years 2005 to 2012. However, the relatively high cost of reagents and high error rate in terms of poly-bases longer than 6 bp remained a challenge for Roche 454 (Liu *et. al.* 2012, p. 2).

The competitors have developed their systems parallel to Roche and by 2012 the sequencing on Illumina systems using sequencing by synthesis (SBS) technology was the cheapest on the market (Liu *et. al.* 2012, p. 4). Also, because of the high quality in sequencing accuracy, the U.S. Food and Drug Administration (FDA) announced in 19th of November 2013 that they will allow marketing of four Illumina NGS systems for clinical use (FDA, 2013), whereas previously NGS systems were only used for scientific research. Furthermore, on 14th of January 2014 Illumina announced the production of a new system promising to deliver full coverage human genomes for less than $1000 and with a high quality standard of 30x, which means each base will be read an average of thirty times by the machine (Hayden, 2014).

2.3 Data security

Today, the readily available and affordable genomic sequencing brings about great opportunities and significant concerns, mainly in personal privacy. At the same time the regulatory frameworks to address these issues are emerging through international efforts, such as the Global Alliance for Genomics and Health. This is a coalition comprised of more than 140 member organizations working towards a goal of effective and responsible data sharing (Global Alliance for Genomics and Health, 2015).

The advances in WGS technology are paving the way to progress in healthcare, but at the same time raise serious concerns. This due to the fact that a person can be uniquely identified by their genome and it also contains information about one's ethnic heritage, different phenotypic traits and predisposition to different types of physical and mental health conditions. At the same time, genomic privacy is often viewed with skepticism because everyone constantly leaves behind biological material like hair, skin and saliva that can be collected and used for DNA sequencing. However, this would be an attack against a targeted individual which is incomparable to privacy threats of accessing large numbers of digitized genomes. (Ayday *et. al.* 2013, p. 5).

The data privacy problem regarding the genome is a complex problem since the genome itself is the ultimate identifier (Homer *et. al.* 2008). Furthermore, it has already been demonstrated in a study by Gymrek *et. al.* 2013 that it is possible to re-identify donors from a public research database by using popular genealogy Web sites and other available information. As a conclusion it can be said that traditional approaches such as aggregation and de-identification (Malin, 2005, p. 29) are not effective in the genomic context (Ayday *et. al.* 2013, p. 5).

Moreover, the privacy issue for genomic data is unique since the genomes of related individuals are highly similar and disclosure of a persons genome can possibly leak information about all their close relatives. As future generations inherit most of their ancestor's DNA, the genomic information disclosure can become an endless curse for them (Ayday *et. al.* 2015, p. 62). This can be illustrated with the case of Henrietta Lacks (August 1, 1920 - October 4, 1951) whose cancer cells were researched nearly five decades ago. This research found that the cells are highly suitable for biogenetic research and resulted in the use of now extremely popular HeLa (in honor of Lacks' name) cell line. The DNA was sequenced and published online without the consent of her family (Humbert *et. al.* 2013, p. 1142).

High security in the use, storage and sharing of genetic data is of utmost importance in the case of incorporating genetic data with EHR. Breach of confidentiality can bring about social harm for the individual like stigmatization, identification of misattributed parentage and family

conflict. In addition to that, discrimination may occur when employment and obtaining different types of insurance is considered. However, health information technology (HIT) professionals have a key role in the protection against these harms by ensuring high security and data governance (Shoenbill *et. al.* 2014, p. 174). It is important to develop security measures that protect not only the information stored within a single healthcare institution but also the exchange of data between institutions and patients themselves (Hazin *et. al.* 2013, p. 812). In order for interoperable EHR systems to maintain privacy, appropriate security measures and access controls have to be implemented throughout the whole workflow (Ruotsalainen *et. al.* 2011). Furthermore, taking into account that genetic information itself is the unique identifier, encryption seems like an ideal answer for the data security question. However, the encryption schemes that are considered strong now may weaken with time, whereas the genome's sensitivity will not. Further measures and legal guidelines are needed for unforeseen weaknesses that might result in early decryption (Ayday *et. al.* 2015, p. 63).

All in all it can be said that because genomes are a new kind of personal health information which raises numerous issues, the technology must work with legal and professional guidelines that determine how it is transmitted, stored, processed and eventually disposed (Ayday *et. al.* 2015).

2.4 Policy and regulation

It is commonly assumed that the field of genetics is advancing faster than ethical standards to help guide its use. However the Committee for the Study of Inborn Errors of Metabolism of the National Academy of Sciences proposed guidelines on ethically responsible genetic screening already in 1975 and these were reiterated in an Institute of Medicine report in 1994 (Shoenbill *et. al.* 2014, p. 171). The World Health Organization (WHO) and other groups have proposed similar guidelines (Andermann *et. al.* 2008).

For the quality of the data, there are three central concepts in classifying genetic tests and their use in the clinical practice. The first is analytical validity that can be defined as the probability that the results reported by the test are correct. The second is clinical validity, which can be defined as the rate of correct assessment of the risk of health or disease. Furthermore, the third one is clinical utility, which can be defined as the degree of safe, effective and available medical interventions to the individual being tested. Therefore, when incorporating genetic data into EHRs the data has to be evaluated according to these three central concepts. Otherwise the

potential risks can outweigh the benefits, resulting in harming patients because of inaccurate or unreliable data. (Shoenbill *et. al.* 2014, p. 171).

When clinical tests in the USA have to be performed by CLIA (clinical laboratory improvement amendments)-certified laboratories, it does not apply to research genetic tests or direct to consumer (DTC) tests offered by for-profit companies (Eng & Sharp, 2010, p. 2). Whereas, the research laboratories are exempt from CLIA's requirements if they do not report patient specific results for the diagnosis, prevention or treatment of any disease or the assessment of an individuals health (Prince *et. al.* 2015, p. 837). However, for the implementation of genomic data use in EHRs, HIT developers have to agree on standards for genetic terminology and methods for data transfer. The study by Shoenbill *et. al.* 2014 has resulted with a diagram (Figure 2) showing the integration of genetic data to EHR and complemented it with examples of standards available for data messaging and genetic data annotation. As already previously mentioned in section 2.2, different sequencing technologies generate data with variable quality. Because of that, the Genome Variation Format (GVF) and the Human Genome Variation Society's nomenclature have been proposed to provide coordination among different sequencing technologies (Masys *et. al.* 2012, p. 420).

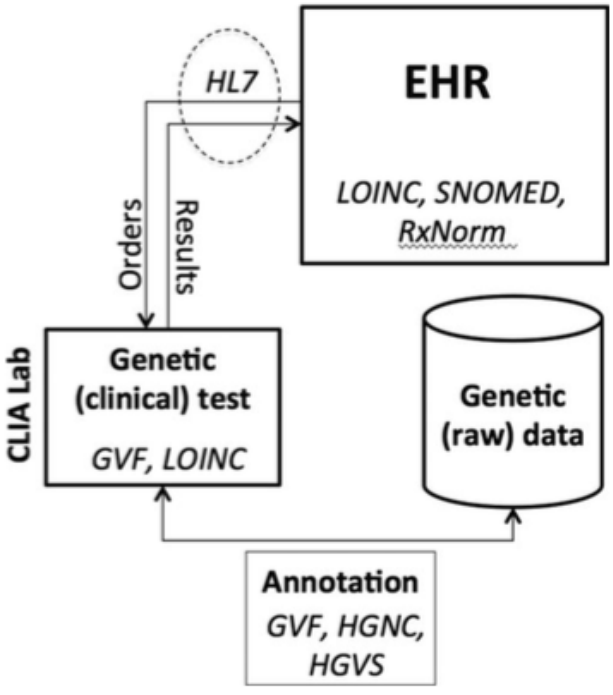

Fig. 2 Integration of genetic data to electronic health record (EHR). CLIA – clinical laboratory improvement amendments; HL7 – Health Level Seven International, data messaging standard; GVF – genomic variation format; HGNC – Human Gene Nomenclature Committee's terminology; HGVS – Human Genome Variation Society; LOINC – Logical Observation Identifiers Names and Codes; SNOMED – Systematized Nomenclature of Medicine; RxNorm – National Library of Medicine (Shoenbill *et. al.* 2014, p. 176)

The majority of different laws and standards relating to data protection in clinical research rely on the fact that a reduction in the potential to identify an individual correlates with increased protection to the privacy of individuals (Elger et al., 2010, p. 232). Whereas existing laws already protect genomic data privacy to some degree and in 2008, the US government established the Genetic Information Nondiscrimination Act (GINA) that does not allow health insurance or employment discrimination on the basis of genetic information (Ayday *et. al.* 2015, p. 63).

Additionally, the Council of Europe has drawn up legal standards focusing on the ethical and legal issues raised by different applications of genetics. On 7 May 2008 the first international legally binding instrument concerning genetic testing for health purposes was passed by the Committee of Ministers of the Council of Europe, that among other things lays down principles on the quality of genetic services, prior information, consent, genetic counseling, protection of private life and the right to information obtained by genetic testing (Lwoff, 2009).

Although the regulations for ethical and data privacy issues has already been put in place, there has been critique that it does not contain enough technical information on how to safely process and store digitized genomes. One of the problems behind this is that these issues for individual genomes and genome collections are not well understood. Moreover, as brought out in section 2.3, the genome is a unique identifier and de-identification is clearly unsuitable. Furthermore, guarding against surreptitious DNA testing with the need of an informed consent from a person is also been highlighted as overlooked. (Ayday *et. al.* 2015, p. 63).

However, on 27 April 2016 the European Parliament and the Council passed regulation 2016/679 on the protection of natural persons on the case of processing personal data and on the free movement of such data. Whereas article 53 of the regulation addresses the processing of personal data in special categories for health-related purposes and concludes it should only be done for the benefit of natural persons and society as a whole in the area of public health. However, member states are allowed to introduce their own limitations regarding the processing of genetic and biometric data, at the same time considering that free flow of personal data and cross-border processing of such data within the Union is not hampered. Continuing, the article 54 states that processing of special categories of personal data can be carried out without the consent of the data subject for reasons of public interest in areas of public health. The article 54 also concludes that the processing of data for health reasons for public interest should not result in data being processed in the interests of third parties such as employers, insurance or banking companies. Moreover, the article 83 of this regulation states that in order to maintain security,

the processor must evaluate the risks and implement further measures such as encryption to mitigate these risks. (The European Parliament & The Council of the European Union, 2016).

2.5 Ethical challenges

Sharing genomics data alongside the EHRs brings about a number of unique challenges, among those ethical (Berger, 2015, p. 129). Related to this, many commentators have voiced their concerns on the issue that genomic research results which lack clinical validity or utility may end up in the medical record. This in turn can cause anxiety and misunderstanding for patients and can also be misinterpreted by general practitioners with little knowledge and training in genomic analysis (Prince *et. al.* 2015, p. 833).

As potential risks of reporting inaccurate genetic data to patients is concerned, it can be illustrated with specific cases. For example, an unvalidated genetic test led to false diagnosis of a metabolic disorder with severe neurodevelopmental symptoms in 27 newborns and if the parents had followed the physicians' recommendations, children would have been kept on a special diet throughout their lives (van Calcar *et. al.* 2007). Another example illustrates the change in reproductive choices for parents of children who were carriers of the gene variation for cystic fibrosis because of the false belief that their child had the disease (Mischler *et. al.* 1998). The question of what types of data should be returned to the patient or to the parents in the case of children are particularly important to consider. For example in the case of WGS should parents receive information about their child's status for the *ApoE4* allele[1] associated with Alzheimer's disease or should parents receive only the data directly related to the disease for which the patient is being treated at the time (Jacob *et. al.* 2013, p. 3)?

Furthermore, since not all genetic information have direct clinical utility, there is an on-going debate concerning the issue of who has the responsibility to determine where to place genetic information in the patient's medical record. How is the information managed when it is necessary to mark some of it as sensitive and how should sharing of genetic results between family members be organized (Shirts *et. al.* 2015)? Whereas in the study by Shirts *et. al.* in 2015 conducted to explore the storage and display of genetic information in the eMERGE in US, it was also shown that over 65% of the respondents said the laboratory performing the test was deciding where the genetic information should be placed in a medical record. Furthermore, it is also debated whether researchers should place research results directly in a person's medical

[1] A variant form of a gene.

record. However in a study by Prince *et. al.* in 2015 the debate has come to a conclusion that there is no clear ethical or legal duty for a researcher to add research results in a medical record unless requested by a participant themselves.

Taking the complexity of genetic information into account, the patients' understanding of the purpose of testing and the significance of the results for their care is of utmost importance (Haga *et. al.* 2014, p. 1). Furthermore, commentators have long recognized that the patient understanding is critical for competent decision-making and not the clinicians or researchers disclosure (Manson & O'Neill, 2007 cited in Parker, 2008, p. 68). Whereas the principle of the informed consent is widely recognized as a pillar of bioethics and it allows individuals to exercise their fundamental right of deciding whether and how their body, body parts or data associated with them will be used (Cambon-Thomsen *et. al.* 2007, p. 376). However, when the completion of the HGP was near, the power of biobanks for epidemiological research was recognized, with Estonia and Iceland being among the first countries to initiate a population-based biobank (Milani *et. al.* 2015, p. 188). This in turn has brought about the difficulties of applying informed consent in the case of large-scale biobanks where a sample is used over a number of years. Consequently a lot of discussions on broad consent and the secondary use of samples were generated (Knoppers, 2005, p. 34). Whereas, a lot of countries are placing the responsibility of deciding on the necessity of a new consent on the research ethics committees (Maschke, 2006, p. 194). Furthermore, it has been concluded that informed consent is necessary in the general process of building trust that is central to all kinds of biobank related projects (Cambon-Thomsen *et. al.* 2007, p. 378). However, with the construction of population-based human genetic databases which are an immense public resource, ethicists increasingly claim that instead of emphasizing the individual's rights, a communitarian view for facilitating research progress crucial to the health of others, should be used (Chadwick & Berg, 2001 cited in Sutrop, 2011, p. 9). In contrast to that, the critics who oppose the broad consent and open consent, believe that they should only be allowed when combined with "opt-out" consent (Kaye, 2004 cited in Sutrop, 2011, p. 10), that would allow the person to restrict the use of their data afterwards.

Concerning the ethical decision-making related to cost-effectiveness of treatment, the doctors have to maintain a balance between the benefits of the individual and the benefits for the public. Moreover, the limited resources require denying the genuine health needs of individuals when it is considered that not enough good from the social perspective is yielded at too high price (Fleck, 2012, p. 759). Today a considerable amount of healthcare costs in the US is covered by privately financed health care: about 33 percent of total health care spending in 2012. Moreover,

in 2013, about 64 percent of US residents received their health insurance coverage from private voluntary health insurance (VHI). Nevertheless, the situation in the US is changing, since in 2010 the Patient Protection and Affordable Care Act (ACA) was established as a "shared responsibility" between the government, employers and individuals, to ensure all Americans have access to affordable and quality health insurance. As a consequence, the number of uninsured is expected to decrease by 26 million by 2017 (Mossialos *et. al.* 2015, p. 153).

Attention has to be paid in what kind of genetic information and how is displayed in the EHR. Additionally, patients understanding about the purpose and outcomes of testing can adversely influence their medical care (Haga *et. al.* 2014, p. 6). Furthermore, the use of genetic data and its linkages to close relatives complicates the situation on whose duty it is to warn patients and their relatives (Hazin *et. al*. 2013, p. 814). It can be concluded that without careful attention to different ethical issues concerning the integration of genetic information to EHR, the promise of personalized, genetic medicine at the individual patient and population level may not be fully realized (Shirts *et. al.* 2015, p. 1240).

2.6 Social acceptance

In order to reach the full potential of genomic medicine there will in the near future surely be a push for all patients to undergo WGS and integrate the results into their electronic medical record (EMR). In order to address the ethical issues arising from integrating the exceptionally personal genomic information into the EMR, patient perspectives will surely be required (Kimball *et. al.* 2014, p. 16). Furthermore, for realizing the potential of genomic medicine and establishing the relationship of disease risk to genomic variants derived from sequencing translational and clinical research studies are required. This in turn will need enrollment of thousands of clinical research participants and new methods and approaches to interact and communicate with them. Whereas, since all patients are likely to have several deleterious variants, it is important for clinicians to know their motivations and expectations in regards to the return of results (Facio *et. al.* 2011, p. 1213).

Today, already a large number of people participate in genomic research and mostly their attitudes towards it are supportive and positive, in particular towards the genotype-guided drug therapy (Facio *et. al.* 2011; Gollust *et. al.* 2012; Harris *et. al* 2012; Olson *et. al.* 2013). More precisely, in a study by Facio *et. al.* 2011, 322 individuals enrolled in the whole-genome sequencing (ClinSeq) study were surveyed and two main themes in their motivation were identified: promoting research and a desire to learn more about the contributors of their own

health. The individuals in this study were between 45 and 65 years of age and located in Washington DC and Baltimore, like most "early adopters" they were more likely to have a college degree or higher and 23% reflected scientific literacy in their responses. Furthermore, in a study by Gollust *et. al.* 2012 where 369 individuals who registered for the Coriell Personalized Medicine Collaborative in the US were questioned, the respondents were motivated by curiosity, finding out disease risk and improving their health. Whereas in this study, overall 32% of respondents stated misperceptions about personal genomics and some expressed unrealistic expectations like the belief, that one of the benefit of the study would be to gain access to gene therapy (13%). In addition to that, the concerns about risks were modest and 31% believed there were no risks.

Nevertheless, there are several surveys and focus groups with participants in genomic studies, which reveal that some individuals indeed have concerns about discrimination from health insurers and employers (Hartzler *et. al.* 2013; Kimball *et. al.* 2014; Ludman *et. al.* 2010; Trinidad *et. al.* 2010). Moreover, the Mayo Clinic in Rochester, Minnesota, USA has been at the forefront of eMERGE Network- a National Human Genome Research Institute-funded consortium participating in the development of best practices for using the EHR as a tool for genomic research (Gottesman *et. al.* 2013). As the social and ethical issues emerged from the integration of genomic information into the EHR, they decided input from the local community was necessary. More precisely in 2007 twenty citizens from the local community were chosen to represent the local population. The members of the resulting Community Advisory Board (CAB) read background material, got information from scientists, patient advocates, legal experts and other stakeholders (Mayo Clinic, 2016). The study by Kimball *et. al.* 2014 was carried out using one of the Mayo Clinic CAB monthly meetings. Although they were positive on hoping the research would improve medicine and expressed hope for the development of new drugs on the basis of genetic information, they also raised many concerns. Among others their comments included the need to make information to prospective study participants comprehensible and they expressed strongly that possible risks regarding privacy and confidentiality have to be made clear in a straightforward way. They were aware that there are existing laws like the GINA, but also recognized limits of those laws and voiced concerns about job discrimination saying it will just take a little leak before people with medical issues will be targeted. Serious concerns about life insurance discrimination and the limits of governmental protection against this were raised. It was said that due to the changing laws it would be difficult to predict the genetic discrimination in the future and some even wondered why a healthy person would risk discrimination by having genomic data incorporated into their medical record. They also voiced

their concerns about the issue that a life-threatening or untreatable diagnosis would result in discontinuation of medical care. Moreover, one of the board members summarized the groups concerns:

"It was moving the information into my personal health record that took it from "this is for the better of everyone in the universe, hooray" to "I am not going to get life insurance." I mean, your life may be better ten years from now, but I don't have any insurance or I get fired from my employer. Science is one thing, but this was taking it down to the individual risk that me as an older person could see right now, right here, tomorrow affecting me." (Kimball *et. al.* 2014, p. 20).

All in all having identified the risks, the board members in the study by Kimball *et. al.* 2014 agreed that there is a need for balancing risk and benefit in using genomic data in clinical care. Whereas some of the members voiced their opinions that due to the personal nature of the information being collected and placed in the EMR, the risks of this kind of studies outweigh the benefits.

Additionally, in the study by Trinidad *et. al.* 2010 performed on an existing dataset from the Adult Changes in Thought (ACT) as part of the eMERGE Network in the US most of the participants endorsed the value of data sharing and while they acknowledged the risks, it was mostly considered that the benefits would outweigh them. Whereas most participants expressed distrust of the ethics, motives, research and marketing practices of pharmaceutical companies and the youngest respondents (18 – 34 years) advocated direct control over how much and which parts of their medical records would be used for research purpose. Furthermore, in a study by Ludman *et. al.* 2010 telephone surveys with 400 individuals who had their data submitted in to the US federal database of Genotypes and Phenotypes (dbGaP) showed that 69% found that re-consent to use their data was very important and to 21% it was somewhat important.

At this point, even when the lay public expresses their support and motivation for genomic research and medicine, there are still some with lingering concerns. Consequently, when medical institutions begin to integrate patient's genomic information into the clinical setting, they must consider both the public support on the use of genomic data to improve patient care and public concerns about potential harms (Kimball *et. al.* 2014, p. 22).

2.7 Healthcare professionals awareness and expectations

When early adopters among consumers in 2012 were already prospectively enthusiastic about personal genomics (Gollust *et. al.* 2012, p. 22), the appreciation by the clinicians still remains a challenge (Manolio *et. al.* 2013, p. 260).

Whereas the lack of interest by clinicians and health-care institutions is on the one hand determined by lack of clinical validity and utility for genomics to improve patient care. On the other hand, there is lack of adoption even for genomic applications with proven validity and utility, such as family history, showing that the lack of evidence is not the only barrier. (Manolio *et. al.* 2013).

The challenge for uptake of genetic services is widely recognized because of the considerable gap between genomic discoveries and their utilization as genetic services (Rogowski *et. al.* 2009 cited in Hamilton *et. al.* 2014, p. 238). The facts that health professionals are not typically educated in genetics and have little time to use genetic services in their practice are contributing factors for the lack of adoption of genetic services in the clinic (Hamilton *et. al.* 2014, p. 238). Furthermore, physicians have rated their knowledge of genetics poor (Haga *et. al.* 2012, p. 390) and a study by Edwards *et. al.* 2009 showed that fifty-five per cent of the 147 respondents did not answer correctly to any of the knowledge questions.

However it is beyond the human mind to manage all known genetic tests and their interpretations without the aid of computerized supports. To illustrate, as of 2013, a genetic test was available for about 3000 diseases and one analysed gene may possess about 1200 to 2000 known variants. It is not possible for a physician to know all the variants a gene may have, thus it is necessary to implement CAA when integrating genetic data to EHR. As for the problem with the lack of time for health professionals, WGS can save a lot of time in contrast to individual genetic tests. More precisely, as mentioned there are nearly 3000 diseases for which individual genetic tests are available, but WGS can retrieve the whole information by one single genetic test, eliminating the need for clinicians to order several single gene tests until a particular diagnosis can be confirmed. (Welch & Kawamoto, 2013).

The psychological impact of the information on patients or their families is also among the risk factors of using genetic data by physicians. In the past, this issue has been addressed with the informed consent process and genetic counselling. However, since the use of genomics in medicine is expanding rapidly, there will soon be a need for additional genetic counsellors (Manolio *et. al.* 2013).

Genetic professionals in the US today can be divided into two specialties that have arisen to provide knowledge about clinical genetics and help manage the diseases (Welch & Kawamoto, 2013, p. 310). Medical geneticists are physicians who are trained to assess genetic risks, manage diseases with a genetic predisposition and provide counselling for individuals and families at risk for diseases with genetic basis (American Board of Medical Specialties (ABMS), 2016). Genetic counsellors however are masters-level health professionals who work with physicians by helping in assessment of genetic risks and communicating genetic information to patients and their families (National Society of Genetic Counselors (NSGC), 2008). By 2013 there was one medical geneticist per 262,000 U.S. citizens and one genetic counsellor per 105,000 U.S. citizens. Whereas the problem of insufficient supply of genetics experts is not expected to change in the near future because aspiring clinicians are not choosing the genetics profession at the rate that is needed (Welch & Kawamoto, 2013, p. 310).

As for the burden to clinicians and patients in obtaining, interpreting and managing the results of genetic testing, there is a need for new innovative models using social media tools to provide counselling and education (Manolio *et. al.* 2013, p. 262).

2.8 Framework of the main technological, legal, ethical and social issues

When the project to determine the whole sequence of a human genome was first announced in 1990 in the US, there were significant technological barriers for achieving it. However, as mentioned in section 2.1 the desire to reach this ambitious goal pushed the technological developments in the field of genetics. By the beginning of the 2010's, the use of full genome sequences was already available in certain hospitals, mostly in the US. Whereas, even though the genomic data is very large, up to about 1 terabyte for one persons genome as mentioned in section 2.2, the availability of compute clusters and cloud computing solves this technological issue. Furthermore, by 2014, the technological issue for DNA sequencing quality was solved with technological developments by the private company Illumina. However, since this development is so recent, the genome sequences existing in different databases still have varying quality.

Because of the technological developments in this field, the technological barriers are no longer a significant problem since 2014. However, they were gradually replaced with numerous softer issues like social acceptance, ethical challenges and personal privacy. Furthermore, all the functions of the human genome have not yet been determined as mentioned in section 2.1 and thus returning whole genome data to patients remains under debate. Moreover, the healthcare

system is restricted by the amount of financing they have for treating one single person and because of that the cost-effectiveness of its use needs to be further studied.

As the fast, high quality and affordable genome sequencing becomes more and more available a lot of opportunities arise, but also concerns about data security and personal privacy. As mentioned in section 2.3 a person can be identified on the basis of their genome data and it also contains information about a person's ethnic background, diseases they are susceptible to and information on mental diseases. Furthermore, since related individuals have similar genome, the disclosure of one person's genome can also leak information about their closest relatives. Because of that, it is necessary to evaluate risks when healthcare institutions process genome data and determine if implementing further security measures like encryption if necessary.

For addressing the issues related to the analytical validity, clinical validity, clinical utility and protection of genetic data, a series of regulations and standards have already been put in place, as mentioned in section 2.4. Clinical tests in the US have to be performed by CLIA-certified laboratories and in order to address the issue of integrating data with variable quality the standardisation according to the Genome Variation Format and the Human Genome Variation Society's nomenclature have been proposed.

Genetic Information Nondiscrimination Act has been established to prevent health insurance or employment discrimination based on genetic information. Similar developments are taking place in Europe and on 27 April 2016 the European Parliament and the Council passed regulation 2016/679 on the protection of natural persons on the case of processing personal data. The regulation states that when central national health authorities are processing special categories of personal data with genetic data among them, it has to be assured that the data is protected against processing by third parties like employers or insurance and banking companies. However, as the Human Genome Project has brought with it a change in bioethics, as mentioned in section 2.5, the ethical considerations have moved from the benefit of the individual to the benefit of the society. Because of that, the new European regulation described in section 2.4, states that processing of health related data can be carried out without the consent of the individual for the purposes of common good and improving public health.

Continuing with the ethical considerations, as already mentioned in previous section there has been a communitarian turn in bioethics, that has been brought about by the technological advances in the field of genetics and consequently established biobank projects in 1999. As already mentioned in section 2.5 the debates on the need for informed consent, when public benefit is considered were initiated by the changes but there were also ethicists who opposed this and emphasised that an "opt-out" option is necessary. Furthermore, as ethical issues are

concerned, the cost-effectiveness of the treatment methods creates another ethical dilemma on how to balance the interests of the individual with the interests of society. Since in the US about 64 per cent of residents received their health insurance from private voluntary health insurance, this can bring with it important differences from the Estonian system.

In order to reach the full potential of genomic medicine a push for all patients to integrate their results in the EMR is expected. Because of that patients opinions and acceptance has a decisive role in the successful integration. As shown in section 2.6 in the US, people's opinions towards it can be considered mostly positive but there are also concerns about insurance and job discrimination related to data leakages. Whereas, it can be brought out that the younger respondents have advocated more control over their genome data.

As shown in section 2.7 the uptake of genetic tests by clinicians remains a challenge because of their lack of training and time. When some aspects of these issues can be solved by computerized support, there is still the need for medical geneticists to assess the risks, manage diseases with genetic predisposition and provide counselling. In terms of this there is an insufficient supply of genetics experts in the US and by 2013 there were no foreseeable changes in regards to this issue.

In terms of theoretical framework, it can be concluded that a holistic view is necessary to comprehend the dynamics of how the emphasis has moved from the technological issues towards the softer issues through time. Whereas, when by 2016 the technological barriers are no longer an issue, the proper management of risks and cultivating social acceptance have a central role in the uptake of this new solution.

3. Integration of genomic data in Estonia

3.1 Emergence of the use of genome data in Estonia

The benefits of biobanks and epidemiological research related to them was recognized and implemented in Estonia starting from 1999. As a consequence, Estonia together with Iceland was among the first countries to initiate a population based biobank for the study of common diseases and traits in the context of genomic medicine. Whereas when the Icelandic initiative was owned by a private enterprise recently acquired by a US multinational biopharmaceutical company Amgen, Inc., the Estonian biobank is maintained by the government (Milani *et. al.* 2015, p. 188). Furthermore, as the technological advances in DNA sequencing continue to lower its costs it becomes more and more available and Estonian government has already understood the potential benefits of genomic data integrated to the EHIS. This can be achieved by implementing personalized medicine principles into Estonian health care system and by creating an innovative computerized infrastructure for research and development. In the end of 2014, the Estonian Ministry of Social Affairs presented the Estonian Government Pilot Project on Personalised Medicine for 2015-2018. Two months later, on 30 January 2015 President Barack Obama introduced the US Precision Medicine Initiative (Kalda *et. al.* 2015, p. 4).

Andres Metspalu, the Director of Estonian Genome Project, elaborated on the benefits of using WGS technology in the hospital at the Estonian Medical Students' Associations' vision conference in 2016. He said that in the past 5 years, technologies that allow the analysis of the whole genome even within 24 hours have become available and this can be used in intensive care situations, especially for children. Whereas, he says that technology transfer from a research laboratory to the hospital takes about 5 years in Tartu and stresses that the ethical and legal implications of any technology have to be analysed before it can be used. Furthermore, the technology has to be accepted by general public and informing people is important. He mentions that exome sequencing that is already a clinically approved method, is used frequently in Estonia. (Metspalu, 2016).

During a phone interview with professor Andres Metspalu he agreed that public acceptance for new technological solutions gives Estonia an advantage. Furthermore, he says that Estonia has the best advantages for developing a nationwide solution where genomic data is integrated to the eHealth system. Estonia has national eHealth database, ID card and Estonian Genome Center with a database of genetic variants of the population. He also says that all the northern countries and Netherlands have high readiness.

3.2 Technological barriers in Estonia

The Estonian Biobank now belongs under the University of Tartu as the Estonian Genome Center at the University of Tartu (EGCUT) and the research institute that was established over a decade ago and is now starting to yield valuable longitudinal follow-up data for a large number of individuals (Milani *et. al.* 2015, p. 188). Today, the EGCUT has the necessary infrastructure for storing biological samples and several DNA sequencing platforms suitable for WGS (EGCUT, 2016). Their dataset consists of biological samples from 50 000 individuals and 20 whole-genome sequence files have been given to Cypher Genomics (San Diego, US) (Reisberg *et. al.* 2015, p. 38). Furthermore, in a phone interview, professor Andres Metspalu said that by May 2016 they have full genome sequences with the high quality accuracy of 30x for 2500 people.

The analysis results for exome sequencing are used in the Tartu University Hospital with decision support software in order to find potential genetic causes for children with specific conditions (Reisberg *et. al.* 2015, p. 38). Although the full genome sequences have been sent to a US company that uses scalable cloud architecture for clinical-grade interpretation of genomes (Cypher Genomics, 2016), there is an existing bioinformatics competence in Estonia. Quretec LLC (Tartu, Estonia) together with Department of Computer Science (University of Tartu) forms a joint Bioinformatics, Algorithmics and Data Mining group (Quretec, 2016). Additionally, in regards to the necessity of high performance computing power and available storage, there is a high performance computing (HPC) farm located in Tallinn University of Technology (Tallinn University of Technology, 2016) and the necessary infrastructure for scientific computing in the High Performance Computing Center of University of Tartu (High Performance Computing Center, 2016).

Thus, whether it is more effective to use outsourcing from other countries or using the architecture and competences already existing in Estonia, the integration of genomic data to EHR is technologically achievable. The report made in collaboration of Software Technology

and Applications Competence Centre (STACC), Quretec and University of Tartu analyses the current status and future needs of the Information Architecture and Data Management solutions for the realization of the national personalized medicine pilot project. One of the conclusions of this report is that handling unstructured input information like research articles properly for high-quality decision support algorithms is far more challenging than the technical side (Reisberg *et. al.* 2015, p. 29).

On the technological feasibility, the director of Estonian Genome Project professor Andres Metspalu commented that as of 2016, the technological readiness is there, but the integration has not been carried out because of a legal issue. However, the eHealth expert professor Peeter Ross commented that the technological issues are not 100% solved and the main problem lies in the difficulties of integrating unstructured data from genetic analyses. He elaborates on the issue and says that one of the main problems with this issue is that both genome sequencing data and summaries for doctors visits are not presented in a structured form. Because of this, the integration of the data cannot be automated and that can be regarded as a technological issue. Furthermore, he states: "For me this is a technological issue…the integration would be doable but not in a fully automated way."

3.3 Data security in Estonia

Today Estonia has an opt-out type of system for EMR and EHIS, which mean that the state collects the data automatically but an individual has the right to close their data. Whereas, for underage individuals, aged 0-18 there will be a legal representative who makes transactions and activities in the EHR on behalf of them. There is also an attending physician concept, which means that the data in the EHIS is accessible to any healthcare professional but all their actions in the system are traceable and they can only view a patient's data if they are treating them. (Ross, 2015b).

Whereas in the past there has been critique that the general level of security of sensitive personal data in Estonia is rather low and most often, the breaches of data are caused by malevolent or negligent employees of the institutions handling the data not by hackers (Sutrop & Simm, 2004, p. 260). Furthermore, Dr. Peeter Ross an Estonian expert on e-health has said in his 2015 lecture on basics of e-health at Tallinn University of Technology that there have been cases where healthcare professionals have been let go because it was identified they had looked at the personal data of individuals who they were not treating.

Pertaining to the privacy and data security issue, one of the main issues is the fact that genome itself is a unique identifier and thus there is a necessity for encryption of the data and further measures to prevent early decryption (Ayday *et. al.* 2015, p. 63). The report by Reisberg *et. al.* 2015 addresses the issue of data security and states that genetic data is individual's property and any access to it should be regulated and tracked. Whereas, when using cloud computing all sensitive data will have to be encrypted before entering the cloud. They also conclude that more advanced encryption measures and access control schemes have to be implemented using cloud computing environments.

The data exchange will occur over the Estonian X-Road and authenticated, encrypted health information will be sent directly to known, trusted recipients via the internet (Reisberg *et. al.* 2015, p. 3). The Estonian eHealth expert Peeter Ross commented on the issue during the interview and said that when secure data exchange is considered, there are no technological issues.

3.4 Policy and regulation in Estonia

In relation to policy and regulation in Estonia, the report by Pikani *et. al.* 2015 states that a number of measures has been already established within the Estonian Health Technology Assessment (HTA) program concerning screening evaluations and this will be an important basis for the personalized medicine program. As for the data accuracy challenges related to the use of genetic data in the clinical decision-making in Estonia, the report by several Estonian experts addresses the issue of the need for a genetic test to be analytically valid, clinically valid and have clinical utility (Pikani *et. al.* 2015, p. 138). Whereas it is pointed out that guidelines for evaluating genetic tests can be followed on the Centers for disease Control and Prevention website where they have established and tested systematic, evidence-based process for evaluating genetic tests and other genomic technology transitioning from research to clinical and public health practice (CDC, 2016).

Andres Metspalu, the Director of Estonian Genome Project, stated in the Estonian Medical Students' Associations' vision conference in 2016 that starting from 2014, exome sequencing has been approved by Estonian Health Insurance Fund as a diagnostic method (Metspalu, 2016). The website of the Estonian Genome Center states that their Core Facility is certified by Illumina (Biotechnology company with headquarters located in San Diego, US) Certified Service Providers (CSPro) for genotyping, gene expression and sequencing (Estonian Genome Center, 2016). Some commercial companies in Estonia that are offering genetic testing services have

pointed out on their website that they are CLIA certified (Asper Biotech, 2016). However, by 2012 accreditation of medical laboratories in Europe is primarily according to ISO15189 (Huisman, 2012) and when this is not mandatory, by 2011 1/3 of Estonian medical laboratories were accredited by this standard (Bakhoff *et al.* 2012, p. 8) and others not accredited also follow these requirements (Bakhoff *et al.* 2012, p. 30). Professor Andres Metspalu has brought out that the problem for integrating existing genetic data to EHIS originates from the fact that the laboratory of Estonian Genome Center is a research laboratory and not a medical laboratory.

The requirements for CLIA are also not mentioned in the report by Pikani *et. al.* 2015, despite the fact that they determined the lack of world scale standardization as one of the main risks for entrepreneurs. Furthermore, the entrepreneurs have to be confident, that when they develop something in Estonia, they will be able to transfer their solutions to bigger markets.

Moreover, as standards in data messaging are concerned, the report by Estonian experts has shown that in architectural decisions, messaging will be implemented through XML based international HL7v3 (extended) messages (Reisberg *et. al.* 2015, p. 21). Additionally, the report by Ross *et. al.* 2015a also mentions that laboratory analyses in most hospitals' electronic patient records (EPR) already have LOINC classification and pathology locations are captured as SNOMED codes (Reisberg *et. al.* 2015, p. 41). In this respect it can be noted that attention has been paid on the fact that international standards are necessary for transferring the data or solutions over the borders in the future.

However a considerable amount of data is presented in non-compliant structures currently but regarding this a strategy to establish an organization responsible for the normalization and distribution of already gathered genome, health and medical data has been proposed (Ross *et. al.* 2015a, p. 5).

3.5 Ethical challenges in Estonia

The report by Pikani *et. al.* 2015 states the importance of informed consent and the protection against stigmatization and discrimination, as well as the importance of the fact that tests are acceptable for the population and the results made comprehensible. In addition to that they conclude that ethical, legal and social issues associated with personalized medicine have to be integrated throughout the process of its uptake in the healthcare system. Furthermore, based on the research article by Vogenberg *et al.* 2010, the report states that commercialization of personalized medicine tools will surely require further research related to its ethical, legal and social implications.

In Estonia, an ethical review of research on the Estonian population genetic database and e-health databases is legally required (The Parliament of Estonia, 2000; 2001). Dr Andres Soosaar, a specialist in ethics of medical research and previous member of the Estonian Council of Bioethics among others (University of Tartu, 2016), has said in the Estonian Medical Students' Associations' vision conference in 2016 that for the Estonian case, the situation is *Harju keskmine* (somewhere in the middle). Whereas there are necessary existing institutions and guidelines are being followed. On the topic of ethics committees, the Estonian eHealth expert professor Peeter Ross commented that there is a separate ethics committee for the EHIS called TIS (*Tervise infosüsteemi*) and many hospitals have their specific rules for conducting studies. Moreover, professor Andres Metspalu has said in a phone interview that adding genome data to the everyday work in hospitals does not raise more complicated ethical questions than using other health related data.

Dr. Andres Soosaar also expressed an opinion that in the future, the medicine ethics and legal side will merge and there will be an increasing need for ethics committees and their approval for different procedures. He also believes that in the case of the autonomy of the individual vs. public health benefit, the centre of gravity will move more towards the collective and social benefit (Soosaar, 2016). Whereas, within the studies related to University of Tartu Centre of Ethics project "The ethical aspects of genetic databases and new technologies: individual versus common values and goods" it has been concluded that there is no point in comparing the individual and common values because a value can be both individual and common (Sutrop & Simm, 2008, p. 513). Furthermore, the study revealed that trust plays a central role and there is a need for rational trust which is based on people's autonomous choices that they can make when they have information (Sutrop, 2011, p. 10).

Dr Andres Soosaar comments on the Additional Protocol to the Convention on Human Rights and Biomedicine concerning Genetic Testing for Health Purposes, passed in 2008 in the EU, saying that the healthcare monitoring was severely disrupted because of it, with registries closing because of the strict policy on informed consent. He also mentions that for this were exemptions and the informed consent is not required if data is used for legal reasons or for public health purposes. Dr Andres Soosaar continues with describing that for the Estonian Genome Center the open informed consent adjusted for biobanking, was used. Whereas, this allows the person to give consent to use his/her biological material if the research is up to certain standards and has been approved. He believes that this kind of mitigating measures for the strict informed consent will be used more in the future for the benefit of social good over individual. (Soosaar, 2016).

During the interview with the Estonian eHealth expert Peeter Ross, he brings out the Estonian ethics committee decision for an "opt-out" system in EHIS. He continues that the ethics committee decided that from the point of view of the individual "opt-in" system would be right but from the point of view of public health "opt-out" is right, in order to collect enough data for decision making on public health. However, Peeter Ross continues that in the case of genetic data things are more complicated and "opt-out" solution is not suitable. This due to the fact that when a person decides about closing their data, they will also decide for their parents and children, because as mentioned in section 2.3 close relatives share high similarities in their genome. Furthermore, he continues with an example of BRCA mutation associated with breast cancer. If the person decides they do not want to know about their genetic predispositions, the person analysing the data is faced with ethical dilemma. Moreover, they have to decide whether to keep quiet about it or find out if this person has a mother or daughter who have the same genetic predisposition and may be interested about their genetic risks. Because of these issues, the discussion concluded that when "opt-out" system is used, all the data should be visible to the family doctor, who can address the issues with relatives, even when one person has said they don't want to know their risks.

Professor Peeter Ross also brings out that one of the reasons an "opt-out" system was preferred in Estonia is the fact that a person's decisions on their health differ according to their health. It is expected that a young person is not able to make adequate decisions on concerning their needs in a later age. In the case of "opt-in", when a twenty-year-old person says they do not want their data to be collected, it would be comparable to destroying data because the data would not be available for use in the future.

Professor Peeter Ross concludes that for the Estonian Government Pilot Project on Personalised Medicine for 2015-2018 they have decided to take informed consents from all participants. Since Estonia has pronounced personalised medicine as one of their priorities, the ethical side will be thoroughly looked over in the future. On the one side there is the individual and their relatives privacy and on the other side there is the question on whether the person should benefit something for the use of their data.

Furthermore, cost-effectiveness has an important dimension in weighing the decisions related to the benefit of the individual vs. social benefit (Pikani *et. al*. 2015, p. 148). Taking into account that in Estonia health insurance system covers about 95% of the population and about two-thirds of total health care expenditure comes through solidarity-based mandatory health insurance contributions in the form of a social payroll tax (Lai *et. al.* 2013, p. xix), the consideration of public benefit is important.

At the Estonian Medical Students' Associations' vision conference in 2016 Andres Metspalu, the Director of Estonian Genome Project, talked about the common benefits of using clinically approved exome sequencing technology in the hospitals in Estonia and that about 30% of cases it contributes to finding the cause of the disease. Furthermore, he stresses that two thirds of these causes are de novo mutations in the genome, this means that this has developed in the genome of the child and is not related to the parents genomes. This has positive effects on reproductive decisions of people. (Metspalu, 2016). Furthermore, during the phone interview with professor Metspalu he comments on the European regulation 679 where it is stated that health related data can be used without a persons consent for public health benefit. He says that Estonia could benefit by studying the 250 000 existing genome maps for children. Whereas the mothers have given, consent for the analysis but there is no consent from the individuals themselves. He sees great potential in this because in that case there would already be usable genetic data for quarter of the population.

3.6 Social acceptance in Estonia

The data collection for the Estonian Genome Project (EGP) started in October 2002 and related to this the Centre of Ethics at the University of Tartu carried through a survey in December 2002. The results of the study showed that the general attitude towards the project was very positive and about 62% of respondents said that they have heard about the project, whereas only 7% regarded themselves as well informed. Additionally, about 83% of the EGP participants said their motivation was to know their own genetic risks. Furthermore, there was a general positive attitude and trust towards science and scientists. However, the least trusted sources of information about the project were politicians and journalists. In addition to that, large number of people had concerns about the misuse of data and mistrust in the institutions that store and handle personal data. All in all it can be said the opinions were ambiguous and when on the one hand there was mistrust then on the other hand trust in science and the notion that the project will make Estonia rich and famous. (Sutrop & Simm, 2004).

As for recent social opinions towards genomic data and genomic medicine in the Estonian context, the report by TNS Emor from spring 2015 where 1213 individuals aged 16 and older were questioned, showed that 76%-82% agrees with the benefits of genetic tests. Moreover, about 66% agrees that the knowledge about genetic testing is low. Whereas, 6% of Estonian population has had a genetic test with most of them having a higher education and their motivations have been learning about their health risks and genetic background. In the issue of

who should decide about the necessity of a genetic test, 32% agrees that the patient should decide, 27% agrees that it should be a part of a routine check up and 26% thinks that the doctor should be the one to decide. Moreover, the younger part of the respondents advocated the role of the patient as the one who decides for themself. (TNS Emor, 2015a).

In the light of general public acceptance towards the use of genetic data, it can be brought out that in the past Estonia has been successful in implementing a radical new solution of e-Voting by successfully generating public acceptance (Kalvet, 2009, p. 515).

3.7 Healthcare professionals awareness and expectations in Estonia

In addition to the opinions of the public, the views of the healthcare professionals have also been reported in Estonia. Whereas 40 healthcare professional over Estonia were interviewed in July 2015 and most of them felt that they do not know enough about genetic data and its use, except for oncologists who were familiar with it. The field itself was considered very expensive and there were doubts about whether it is affordable in Estonia. In contrast to this, the opinions on personalized drugs based on genetic data were positive and healthcare professionals believed it would make treatment more effective and help save money on the drugs that would not work. Additionally all healthcare professionals, oncologists among them, agreed that thorough education and training for doctors and healthcare professionals is necessary. (TNS Emor, 2015b).

As already reflected by the opinions of the healthcare professionals, genomic medicine is expensive and its cost-effectiveness for social good has to be evaluated. Furthermore, as mentioned by Dr Andres Soosaar, concerning the ethical questions the balance moves from the individual benefits to the benefits of public health and cost-effectiveness has an important dimension in this. In order to understand cost-effectiveness and survival impact of new technologies, new methodologies and new approaches are needed, as concluded in a report by Estonian experts (Lewis *et. al.* 2013 cited in Pikani *et. al.* 2015, p. 148).

4. Discussion

The analysis of previous research on technological, legal, ethical and social perspectives on the use of genome data in hospitals enabled to create a framework which could be applied in the Estonian context.

As concluded in the section 2.8, the technological advances in the field of genetics and ICT enable the integration of genomic data to EHRs. In Estonia this integration is doable already in 2016, but not in a fully automated way. Furthermore, it can be concluded that today there is not 100% readiness in terms of technological issues, because of the problems in connecting unstructured data like genetic data reports and reports from doctors visits. However, in terms of data security, it can be concluded that there are no technological issues and attention has been paid to the need for encryption of genetic data.

When regulation is concerned, it can be brought out that international standards are being followed in terms of data messaging and for laboratory analyses in the hospitals. The European regulation 2016/679 brought out in section 2.4 addresses the free flow of personal data and cross-border processing of such data within the Union. In terms of this Estonia should pay attention to standardizing the unstructured data brought out in section 3.4, for it to be interoperable for cross-border solutions in the future. As for now there is still a significant amount of data that is in noncompliant structures and this also raises a technological issue as mentioned in the previous section. For a solution to this, an establishment of an organization responsible for the normalization and distribution of already gathered genome, health and medical data has been proposed in section 3.4.

Furthermore, when legal issues are concerned, as also mentioned in section 2.4 the medical laboratories in the European union are accredited to the ISO15189 standard. One of the issues of integrating genetic data to EHIS lies in the fact that Estonian Genome Center has a research laboratory and they do not follow the requirements necessary for the medical ones. Although the 2500 full genome sequences have the high accuracy of 30x, they cannot be directly integrated.

As for ethical issues are concerned, the framework in the section 2.8 brings out that there has been a communitarian turn in bioethics. Furthermore, Estonia has also considered the benefit of

the public when introducing the "opt-out" system in the EHIS. However, as genetic data is concerned, just "opt-out" system is not enough, because as brought out in the section 2.8 the genetic risks of the individual also affect their relatives. So, when a person decides they do not want to know about the risks, they also decide on behalf of their closest relatives. Further discussion on the ethical issues is required but as for now one of the solutions that have been pointed out, is that the family physician should have access to all the data and inform the family members who are interested about their health risks.

As brought out in the section 2.8, there is a need for large amount of patients to incorporate their genome data to their EPRs for the advancement of genomic medicine. It can be concluded by this, that patient perspectives and social acceptance will drive the further developments in this field. There is a general positive attitude towards genomic medicine, both in the US and in Estonia. However, it can be highlighted that in the US, there are also several concerns about discrimination mostly when health insurance is concerned. This can be attributed to the fact that the US has a primarily privately financed health insurance as mentioned in section 2.5 but Estonia has a solidarity-based health insurance brought out in section 3.5.

Furthermore, younger people both in the US and in Estonia advocated more control over their data. This issue has been discussed in Estonia and professor Peeter Ross brings out that one of the reasons why the "opt-out" system was preferred was the fact that when younger people make the decision to close their data, they can always reopen it in the future. It was considered that younger people are not able to make far-reaching decisions about their future health needs and in this case, the data would not be lost.

As seen in section 2.7 the lack of uptake of genetic data by clinicians in the US can be attributed to the fact that there is lack of knowledge and training. In Estonia the healthcare professionals opinions on personalised drugs based on genetics have been reported positive as shown in section 3.7. However, both healthcare professionals in the US and Estonia bring out the need for additional training and education in this field.

Overall, it can be concluded that the hypothesis set in the first part of this thesis held up partially. As for the Estonian case, the technological issues are not 100% solved in terms of achieving a fully automated solution but the integration is technologically doable nevertheless. On the second part of the hypothesis it can be said that over time the main focus has moved from the technological issues towards softer issues like personal privacy, ethical questions and social acceptance. In terms of this, there is a general positive attitude towards genomic medicine in Estonia.

5. Conclusion

This thesis focused on analysing the technological, legal, ethical and social perspectives of integrating genetic data to the Estonian Health Information System (EHIS). As a result the main findings can be highlighted as following: there is a need for standardizing data in different databases, the laboratories conducting genetic analyses for the integration should follow the requirements for medical laboratories, further discussion is needed on the ethical issues, there is a need for training and education of healthcare professionals and there is an overall positive attitude towards genomic medicine.

In terms of unstructured data, there is an overall need to introduce international standards in the working environments of institutions handling data. This is important both in terms of integrating data within the scope of the national system, but also for the interoperability of cross-boarder healthcare solutions in the European Union. Not all data problems can be solved with standardisation and since there are various future sources of unstructured data, like for example new published research, Estonia should use a technological solution. For example, the supercomputer IBM Watson cognitive technology is already used in practice in several cancer centres in the US (Doyle-Lindrud, 2015). Although it would take several years to teach Estonian to the supercomputer, it can serve as an important tool in the future.

In terms of the requirements for the medical laboratories, Estonia should on the one hand keep in mind that the data would comply with international standards, but on the other hand find a way to use high quality data already existing in the Estonian Genome Center laboratory.

When ethical issues are considered a thorough discussion is still required to find the best solution. Furthermore, there is a need for discussion on what the person should benefit if their data is used to develop new treatment methods or medicine. As for now during the pilot project the informed consent approach will be used.

In conclusion it can be said that the current position is a good starting point for Estonia and because there are already existing nationwide IT solutions and mostly positive attitudes towards genetic data the realisation of the project within the coming years is possible.

Bibliography

American Board of Medical Specialties (ABMS). Medical Genetics and Genomics. Available at: http://www.abms.org/member-boards/contact-an-abms-member-board/american-board-of-medical-genetics-and-genomics/ [Accessed 27. January 2016].

Altman, R. B. (2013). Personal genomic measurements: the opportunity for information integration. *Clin Pharmacol Ther*, 93, 21–23.

Andermann, A., Blancquaert, I., Beauchamp, S., Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization* 86, 317–19.

Asper Biotech. Available at: http://www.asperbio.com, [Accessed 23. March 2016].

Ayday, E., De Cristofaro, E., Hubaux, J-P., Tsudik, G. (2013). The Chills and Thrills of Whole Genome Sequencing. *Computer*, no. 1, pp. 1, PrePrints.

Ayday, E., De Cristofaro, E., Hubaux, J-P., Tsudik, G. (2015). Whole Genome Sequencing: Revolutionary Medicine or Privacy Nightmare. *IEEE Computer Magazine*, 48, No. 2.

Baker, M. (2010). Next-generation sequencing: adjusting to data overload. *Nature Methods* 7, 495-499.

Berger, M. L., Lipset, C., Gutteridge, A., Axelsen, K., Subedi, P., Madigan, D. (2015). Optimizing the leveraging of real-world data to improve the development and use of medicines. *Value in Health*. 18(1), 127-30.

Bux, M., Leser, U. (2013). Parallelization in Scientific Workflow Management Systems. *ArXiv e-prints*. 1303.7195.

Cambon-Thomsen, A., Rial-Sebbag, E., Knoppers, B. M. (2007). Trends in ethical and legal frameworks for the use of human biobanks. *European Respiratory Journal* 30, 373-382.

Centers for Disease Control and Prevention (CDC). Genomic testing. ACCE. Available at: http://www.cdc.gov/genomics/gtesting/index.htm, [Accessed 22. March 2016].

Chadwick, R., Berg, K. (2001). Solidarity and equity: new ethical frameworks for genetic databases. *Nature Reviews. Genetics* 2, 318–21.

Cohen, J. T., Neumann, P. J., Weinstein, M. C. (2008). Does preventive care save money? Health economics and the presidential candidates. *New England Journal of Medicine* 358, 661–663.

Cordero, P., Ashley, E. A. (2012). Whole-Genome Sequencing in Personalized Therapeutics. *Clinical Pharmacology & Therapeutics* 91(6), 1001-9.

Council of Europe. Details of Treaty no.203. Available at: http://conventions.coe.int/Treaty/EN/Treaties/Html/203.htm, [Accessed 19. March, 2016].

Cypher Genomics. About us. Available at: http://cyphergenomics.com/about-us/, [Accessed 22. March 2016].

Dorschner, M. O., Amendola, L. M., Turner, E. H., Robertson, P. D., Shirts, B. H., Gallego, C. J., et al. (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. *The American Journal of Human Genetics* 93, 631–640.

Doyle-Lindrud, S. (2015). Watson will see you now: a supercomputer to help clinicians make informed treatment decisions. *Clinical Journal of Oncology Nursing* 19(1), 31-32.

Dranseika, V., Gefenas, E., Cekanauskaite, A., Hug, K., Mezinska, S., Peicius, E., Silis, V., Soosaar, A., Strosberg, M. (2011) Twenty years of human research ethics committees in the Baltic States. *Developing World Bioethics* 11(1), 48-54.

Edwards, Q.T., Maradiegue, A., Seibert, D., Saunders-Goldson, S., Humphreys, S. (2009). Breast cancer risk elements and nurse practitioners' knowledge, use, and perceived comfort level of breast cancer risk assessment. *Journal of the American Academy of Nurse Practitioners* 21, 270–277.

Eesti Arengufond. (2014). Smart specialization report on the growth area of health care technology. p 21. Available at: http://ns.arengufond.ee/med-report, [Accessed 16. May 2016].

Egan, J. B., Shi, C. X., Tembe, W., Christoforides, A., Kurdoglu, A., Sinari, S., Middha, S., Asmann, Y., Schmidt, J., Braggio, E., Keats, J. J., Fonseca, R., Bergsagel, P. L., Craig, D. W., Carpten, J. D., Stewart, A. K. (2012). Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood* 120, 1060–1066.

Elger, B. S., Iavindrasana, J., Iacono, L. L., Müller, H., Roduit, N., Summers, P., Wright, J. (2010). Strategies for Health Data Exchange for Secondary, Cross-institutional Clinical Research. *Computer Methods and Programs in Biomedicine* 99(3), 230-51.

Eng, C., Sharp, R. R. (2010). Bioethical and clinical dilemmas of direct-to-consumer personal genomic testing: the problem of misattributed equivalence. *Science Translational Medicine* 2, 17-5.

Estonian Genome Center, University of Tartu (EGCUT). Core Facility. Equipment. Available at: http://www.geenivaramu.ee/en/core-facility/equipment, [Accessed 21. March 2016].

Estonian Genome Center, University of Tartu (EGCUT). Core Facility. Available at: http://www.geenivaramu.ee/en/core-facility, [Accessed 22. March 2016].

Estonian e-health Foundation. Available at: http://www.e-tervis.ee/index.php/en/, [Accessed 5 May 2015].

European Commission. Public Health Genomics. Available at: http://ec.europa.eu/health/screening_genetics/genomics/index_en.htm, [Accessed19. April 2015].

Facio, F. M., Brooks, S., Loewenstein, J., Green, S., Biesecker, L. G., Biesecker, B. B. (2011). Motivators for participation in a whole-genome sequencing study: implications for translational genomics research. *European Journal of Human Genetics* 19(12), 1213-7.

Faulkner, E., Annemans, L., Garrison, L., Helfand, M., Holtorf, A. P., Hornberger, J., Hughes, D., Li, T., Malone, D., Payne, K., Siebert, U., Towse, A., Veenstra, D., Watkins, J. (2012). Personalized Medicine Development and Reimbursement Working Group. Challenges in the development and reimbursement of personalized medicine-payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. *Value in Health* 15(8), 1162-71.

Fleck, L. M. (2012). Pharmacogenomics and Personalised Medicine: Wicked Problems, Ragged Edges and Ethical Precipices. *New Biotechnology*, *Molecular Diagnostics & Personalised Medicine* 29(6), 757–68.

FDA. FDA news release Nov. 19, 2013. FDA allows marketing of four "next generation" gene sequencing devices. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm375742.htm, [Accessed 14 January 2016].

Global Alliance for Genomics and Health. Framework for responsible sharing of genomic and health-related data. Available at: http://genomicsandhealth.org/node/6611, [Accessed 5 May 2015].

Gollust, S. E., Gordon, E. S., Zayac, C., Griffin, G., Christman, M. F., Pyeritz, R. E., Wawak, L., Bernhardt, B. A. (2012). Motivations and perceptions of early adopters of personalized genomics: perspectives from research participants. *Public Health Genomics* 15(1), 22-30.

Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W. A., Li, R., Manolio, T. A., Sanderson, S. C., Kannry, J., Zinberg, R., Basford, M. A., Brilliant, M., Carey, D. J., Chisholm, R. L., Chute, C. G., Connolly, J. J., Crosslin, D., Denny, J. C., Gallego, C. J., Haines, J. L., Hakonarson, H., Harley, J., Jarvik, G. P., Kohane, I., Kullo, I. J., Larson, E. B., McCarty, C., Ritchie, M. D., Roden, D. M., Smith, M. E., Böttinger, E. P., Williams, M. S. (2013). eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genetics in Medicine* 15(10), 761-71.

Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., McGuire, A. L., Nussbaum, R. L., O'Daniel, J. M., Ormond, K. E., Rehm, H. L. Watson, M. S., Williams, M. S., Biesecker, L. G. (2013a). ACMG recommendations for reporting of incidental find- ings in clinical exome and genome sequencing. *Genetics in Medicine* 15, 565–574.

Green, R. C., Rehm, H. L., Kohane, I. S. (2013b) Clinical genome sequencing. In Genomic and Personalized Medicine. *Vol. 1. 2nd edition. Waltham, PA: Academic Press*, 102–122.

Guttmacher, A. E., McGuire, A. L., Ponder, B., Stefansson, K. (2010). Personalized genomic information: preparing for the future of genetic medicine. *Nature Reviews Genetics* 11(2), 161-5.

Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. and Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science*, vol. 339, no. 6117, 321-324.

Haga, S. B., Burke, W., Ginsburg, G. S. Mills, R. Agans, R. (2012). Primary care physicians' knowledge of and experience with pharmacogenetic testing. *Clinical Genetics* 82, 388–394.

Haga, S. B., Mills, R., Pollak, K. I., Rehder, C., Buchanan, A. H., Lipkus, I. M., Crow, J. H., Datto, M. (2014). Developing patient-friendly genetic and genomic test reports: formats to promote patient engagement and understanding. *Genome Medicine* 6(7), 58.

Hamilton, A. B., Oishi, S., Yano, E. M., Gammage, C. E., Marshall, N. J., Scheuner, M. T. (2014). Factors influencing organizational adoption and implementation of clinical genetic services. *Genetics in Medicine* 16(3), 238-45.

Harris, E. D., Ziniel, S. I., Amatruda, J. G., Clinton, C. M., Savage, S. K., Taylor, P. L., Huntington, N. L., Green, R. C., Holm, I. A. (2012). The beliefs, motivations, and expectations of parents who have enrolled their children in a genetic biorepository. *Genetics in Medicine* 14(3), 330-7.

Hartzler, A., McCarty, C. A., Rasmussen, L. V., Williams, M. S., Brilliant, M., Bowton, E. A., Clayton, E. W., Faucett, W. A., Ferryman, K., Field, J. R., Fullerton, S. M., Horowitz, C. R., Koenig, B. A., McCormick, J. B., Ralston, J. D., Sanderson, S. C., Smith, M. E., Trinidad, S. B. (2013). Stakeholder engagement: a key component of integrating genomic information into electronic health records. *Genetics in Medicine* 15(10), 792-801.

Hayden, E. C. (2014). Is the $1000 genome for real? *Nature news* Available at: http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530, [Accessed 14 January 2016].

Hazin, R., Brothers, K. B., Malin, B. A., Koening, B. A., Sanderson, S. C., Rothstein, M. A., Williams, M. S., Clayton, E. W., Kullo, I. J. (2013). Ethical, legal, and social implications of incorporating genomic information into electronic health records. *Genetics in Medicine* 15, 810-816

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., Craig, D. W. (2008). Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays. *PLoS Genetics* 4(8), e1000167.

Huisman, W. (2012). European medical laboratory accreditation. Present situation and steps to harmonisation. *Clinical Chemistry and Laboratory Medicine* 50(7), 1147-52.

Humbert, M., Ayday, E., Telenti. A., Hubaux, J-P. (2013). Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy. *Proc. 20th ACM Conf. Computer and Communications Security (CCS 13)*, 1141-1152.

Jacob, H. J., Abrams, K., Bick, D. P., Brodie, K., Dimmock, D. P., Farrell, M., Geurts, J., Harris, J., Helbling, D., Joers, B. J., Kliegman, R., Kowalski, G., Lazar, J., Margolis, D. A., North, P., Northup, J., Roquemore-Goins, A., Scharer, G., Shimoyama, M., Strong, K., Taylor, B., Tsaih, S. W., Tschannen, M. R., Veith, R. L., Wendt-Andrae, J., Wilk, B., Worthey, E. A. (2013). Genomics in clinical practice: lessons from the front lines. *Science Translational Medicine*

5(194), 194cm5.

Kalda, R., Kathiresan, S., Kiivet, R-A., Lotman, E-M., Padrik, P., Reppo, I., Saretok, M., Sund, R., Viigimaa, M., Volke, V., Themas, A., Themas, E. (2015). Feasibility Study for Personalised Medicine in Estonia: Clinical Approach. *Tartu University*. Available at: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Personaalmeditsiin/feasibility_study_for_personalised_medicine_in_estonia.pdf, [Accessed 21. March 2016].

Kalvet, T. (2009). Management of Technology: The case of e-Voting in Estonia. *International Conference on Computer Technology and Development 13-15 November 2009, Kota Kinabalu, Malaysia. IEEE Computer Society*, 512-515.

Kaye, J. (2004). Broad consent – the only option for population genetic databases? In: Árnason G, Nordal S, Árnason V, eds. Blood and Data: Ethical, Legal and Social Aspects of Human Genetic Databases. *Reykjavik: University of Iceland Press,* 103–9.

Kimberly, M. B., Hoehn, K. S., Feudtner, C., Nelson, R. M., Schreiner, M. (2006). Variation in standards of research compensation and child assent practices: a comparison of 69 institutional review board-approved informed permission and assent forms for 3 multicenter pediatric clinical trials. *Pediatrics* 117(5), 1706-11.

Kimball, B. C., Nowakowski, K. E., Maschke, K. J., McCormick, J. B. (2014). Genomic Data in the Electronic Medical Record: Perspectives From a Biobank Community Advisory Board. *Journal of Empirical Research on Human Research Ethics* 9(5), 16-24.

Knoppers, B. M. (2005). Consent revisited: points to consider. *Health Law Review* 13, 33–38.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions. *Chicago: University of Chicago Press.*

Kwon, T., Yoo, W. G., Lee, W-J., Kim, W., Kim, D-W. (2015). Next-generation sequencing data analysis on cloud computing. *Genes & Genomics* 37, 489-501.

Lai, T., Habicht, T., Kahur, K., Reinap, M., Kiivet, R., van Ginneken, E. (2013). Estonia: health system review. *Health Systems in Transition* 15(6), 1–196.

Lander, E.S., Linton, L. M., Birren, B., Nusbaum, C. Zody, M. C., Baldwin, J., Devon, K., Dewar, K. *et al*. (2001). International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lewis, J. R., Lipworth, W. L., Kerridge, I. H., Day, R. O. (2013). The economic evaluation of personalized oncology medicines: ethical challenges. *The Medical Journal of Australia* 199(7), 471-3.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lihua, L., Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology* 2012, Article ID 251364

Ludman, E. J., Fullerton, S. M., Spangler, L., Trinidad, S. B., Fujii, M. M., Jarvik, G. P., Larson, E. B., Burke, W. (2010). Glad you asked: participants' opinions of re-consent for dbGap data submission. *Journal of Empirical Research on Human Research Ethics* 5(3), 9-16.

Lwoff, L. (2009). Council of Europe adopts protocol on genetic testing for health purposes. *European Journal of Human Genetics* 17(11), 1374-1377.

Malin, B. (2005). An evaluation of the current state of genomic data privacy protection technology and roadmap for the future. *Journal of the American Medical Informatics Association* 12(1), 28-34.

Manolio, T. A., Chisholm, R. L., Ozenberger, B., Roden, D. M., Williams, M. S., Wilson, R., Bick, D., Bottinger, E. P., Brilliant, M. H., Eng, C., Frazer, K. A., Korf, B., Ledbetter, D. H., Lupski, J. R., Marsh, C., Mrazek, D., Murray, M. F., O'Donnell, P. H., Rader, D. J., Relling, M. V., Shuldiner, A. R., Valle, D., Weinshilboum, R., Green, E. D., Ginsburg, G. S. (2013). Implementing genomic medicine in the clinic: the future is here. *Genetics in Medicine* 15(4):258-67.

Manson, N. C., O'Neill, O. (2007). *Rethinking Informed Consent in Bioethics.* Cambridge University Press.

Maschke, K. J. (2006). Alternative consent approaches for biobank research. *The Lancet Oncology* 7, 193–194.

Masys, D. R., Jarvik, G. P., Abernethy, N. F., Anderson, N. R., Papanicolaou, G. J., Paltoo, D. N., Hoffman, M. A., Kohane, I. S., Levy, H. P. (2012). Technical desiderata for the integration of genomic data into electronic health records. *Journal of Biomedical Informatics* 45, 419– 422.

Mayo Clinic. Governance and Oversight. Available at: http://www.mayo.edu/research/centers-programs/mayo-clinic-biobank/about/governance-oversight, [Accessed 28 March 2016].

McCarthy, M. (2011). European health research and globalization: is the public-private balance right? *Globalization and Health* 7, 5

Metspalu, A. (2016). Personaalmeditsiini võimalused praegu ja tulevikus. Visioonikonverents tervise tuhat tahku. Tallinn. Available at: http://www.eays25.ee/visioonikonverents/ulekanne/#, [Accessed 22 March 2016].

Milani, L., Leitsalu, L., Metspalu, A. (2015). An epidemiological perspective of personalized medicine: the Estonian experience. *Journal of Internal Medicine* 277(2), 188-200.

Mischler, E. H., Wilfond, B. S., Fost, N., Laxova, A., Reiser, C., Sauer, C. M., Makholm, L. M., Shen, G., Feenan, L., McCarthy, C., Farrell, P. M. (1998). Cystic fibrosis newborn screening: impact on reproductive behavior and implications for genetic counseling. *Pediatrics* 102, 44–52.

Mossialos, E., Wenzl, M., Osborn, R., Anderson, C. (2014). International Profiles of Health Care Systems. *Commonwealth Fund* pub. No. 1802.

National Human Genome Research Institute sequencing costs. Available at: http://www.genome.gov/sequencingcosts/, [Accessed 13. January 2016].

National Society of Genetic Counselors (NSGC). (2008). Making Sense of Your Genes: A Guide to Genetic Counseling. *NSGC, Chicago, IL, USA*

Olson, J. E., Ryu, E., Johnson, K. J., Koenig, B. A., Maschke, K. J., Morrisette, J. A., Liebow, M., Takahashi, P. Y., Fredericksen, Z. S., Sharma, R. G., Anderson, K. S., Hathcock, M. A., Carnahan, J. A., Pathak, J., Lindor, N. M., Beebe, T. J., Thibodeau, S. N., Cerhan, J. R. (2013). The Mayo Clinic Biobank: A building block for individualized medicine. *Mayo Clinic Proceedings* 88, 952-962.

O'Donnell, J. C. (2013). Personalized medicine and the role of health economics and outcomes research: issues, applications, emerging trends, and future research. *Value in Health* 16(6 Suppl), S1-3.

Parker, L. (2008). Review of Neil C. Manson and Onora O'Neill, Rethinking Informed Consent in Bioethics. *The American Journal of Bioethics* 8(8), 68-71.

Payne, K., Annemans, L. (2013). Reflections on market access for personalized medicine: recommendations for Europe. *Value in Health* 16(Suppl.), S32–8.

Pennisi, E. (2012). Genomics. ENCODE project writes eulogy for junk DNA. *Science* 337(6099), 1159-1161.

Pikani, J., Rannamäe, A., Kruuv-Käo, K., Vunder, K., Mäkinen, S., Ilves, M., Antonov, K., Kruus, P., Paat-Ahi, G., Vinni, R., Carrol, N. (2015). Feasibility study for the development of business cooperation, management organisation and evaluation methodology for personalised medicine pilot project. *Tallinn University of Technology.* Available at: https://www.sm.ee/sites/default/files/content-editors/Personaalmeditsiin/final-report_management_business_evaluation_30_06_15.pdf, [Accessed 22. March 2016].

Prince, A. E., Conley, J. M., Davis, A. M., Lázaro-Muñoz, G., Cadigan, R. J. (2015). Automatic Placement of Genomic Research Results in Medical Records: Do Researchers Have a Duty? Should Participants Have a Choice? *The Journal of Law Medicine & Ethics* 43(4):827-42.

Quretec. Bioinformatics. (2016). Available at: http://www.quretec.com/bioinformatics.html, [Accessed 22. March 2016].

Reisberg, S., Talvik, H-A., Koppel, K., Laur, S., Vilo, J. (2015). Description of the current status and future needs of the Information Architecture and Data Management solutions for the national personalized medicine pilot project. Available at: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Personaalmeditsiin/description_of_the_current_status_and_future_needs_of_the_information_architecture_and_data_management_solutions_for_the_national_personalised_medicine_pilot_project.pdf, [Accessed 21 March 2016].

Rogowski, W. H., Grosse, S. D., Khoury, M. J. (2009). Challenges of translating genetic tests into clinical and public health practice. *Nature Reviews Genetics* 10, 489–495.

Ross, P., Läänelaid, S., Vanker, E., Pruks, I., Viigimaa, M., Padrik, P., Jakovlev, Ü., Metsallik, J., Mõtus, M., Kaevats, J., Taveter, K., Kuusik, R., Leego, E., Korm, K., Parv, L., Laus, R., Salminen-Mankonen, H. J., Wilcox, A. B., Lam, S. H., Haikonen, T., Kunnamo, I. (2015a).

Feasibility study for the development of digital decision support systems for personalised medicine. *Tallinn University of Technology*. Available at: https://www.sm.ee/sites/default/files/content-editors/eesmargid_ja_tegevused/Personaalmeditsiin/feasibility_study_for_the_development_of_digital_decision_support_systems_for_personalised_medicine.pdf, 22 March 2016

Ross, P. (2015b). Access rights. *Tallinn University of Technology.*

Ruotsalainen, P., Blobel, B., Nykänen, P., Seppälä, A., Sorvari, H. (2011). Framework model and principles for trusted information sharing in pervasive health. *Studies in Health Technology and Informatics* 169, 497–501.

Samuels, D. C., Han, L., Li, J., Quanghu, S., Clark, T. A., Shyr, Y., Guo, Y. (2013). Finding the lost treasures in exome sequencing data. *Trends in Genetics* 29(10), 593-599.

Shendure, J., Mitra, R. D., Varma, C., Church, G. M. (2004). Advanced sequencing technologies: methods and goals. *Nature Reviews Genetics* 5(5), 335-44.

Shirts, B. H., Salama, J. S., Aronson, S. J., Chung, W. K., Gray, S. W., Hindorff, L. A., Jarvik, G. P., Plon, S. E., Stoffel, E. M., Tarczy-Homoch, P. Z., Van Allen, E. M., Weck, K. E., Chute, C. G., Freimuth, R. R., Grundmeier, R. W., Hartzler, A. L., Li, R., Peissig, P. L., Peterson, J. F., Rasmussen, L. V., Starren, J. B., Williams, M. S., Overby, C. L. (2015). CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record. *Journal of the American Medical Information Association* 22, 1231-1242.

Shoenbill, K., Fost, N., Tachinardi, U., Mendonca, E. A. (2014). Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *Journal of the American Medical Informatics Association* 21(1), 171-80.

Soosaar, A. (2016). Meditsiinieetika põhiküsimused tulevikus. Visioonikonverents tervise tuhat tahku. Tallinn. Available at: http://www.eays25.ee/visioonikonverents/ulekanne/#, [Accessed 22 March 2016].

Spjuth, O., Bongcam-Rudloff, E., Hernández, G. C., Forer, L., Giovacchini, Guimera, R. V., Kallio, Korpelainen, E, Kandula, M. M., Krachunov, M., Kreil, D. P., Kulev, O., Labaj, P. P., Lampa, S., Pireddu, L., Schönherr, S., Siretskiy, A., Vassilev, D. (2015). Experiences with workflows for automating data-intensive bioinformatics. *Biology Direct* 10, 43.

Sutrop, M., Simm, K. (2004). The Estonian healthcare system and the genetic database project: From limited resources to big hopes. *Cambridge Quarterly of Healthcare Ethics* 13 (3), 254−262.

Sutrop, M, Simm, K (2008). *Public and Private Interests in the Genomic Era: a pluralist approach*. In: Gunning, J.; Holm, S. (Ed.). Ethics, Law and Society (205−216). Ashgate Publishing Ltd.

Sutrop, M. (2011). Changing ethical frameworks: from individual rights to the common good? *Cambridge Quarterly of Healthcare Ethics* 20(4), 533-45.

Tallinn University of Technology. HPC. Available at: http://ttu.ee/support-structure/it-services/services-4/high-performance-computing-2/, [Accessed 26 April 2016].

The European Parliament & The Council of the European Union. (2016). Regulation (EU) 2016/679 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. *Official Journal of the European Union.* Available at: http://eur-lex.europa.eu/eli/reg/2016/679/oj, [Accessed 14 May 2016].

The Parliament of Estonia. Estonian human genes research act. (2000). Available at: https://www.riigiteataja.ee/en/eli/531102013003/consolide, [Accessed 18. April 2015].

The Parliament of Estonia. Tervishoiuteenuste korraldamise seadus. (2001). Available at: https://www.riigiteataja.ee/akt/110032011009, [Accessed 18. April 2015].

TNS Emor. (2015a). Eesti elanike teadmised, hoiakud, kartused ja ootused personaalmeditsiini osas. Available at: https://www.sm.ee/sites/default/files/content-editors/Personaalmeditsiin/elanikkonna_personaalmeditsiini_uuring.pdf, [Accessed 22 March 2016].

TNS Emor. (2015b). Tervishoiutöötajate ootused ja valmisolek tervise- ja pärilikkuseinfo kasutamiseks patsiendi personaalseks riskipõhiseks nõustamiseks. Available at: https://www.sm.ee/sites/default/files/content-editors/Personaalmeditsiin/personaalmeditsiin_tervishoiutootajad.pdf, [Accessed 22 March 2016].

Towse, A., Garrison, L. P. Jr. (2013) Economic incentives for evidence generation: promoting an efficient path to personalized medicine. *Value in Health* 16(6 Suppl), S39-43.

Trinidad, S. B., Fullerton, S. M., Bares, J. M., Jarvik, G. P., Larson, E. B., Burke, W. (2010). Genomic research and wide data sharing: views of prospective participants. *Genetics in Medicine* 12(8), 486-95.

University of Tartu. About Us. Available at: http://www.ut.ee/en/andres-soosaar, [Accessed 30 March 2016].

High Performance Computing Center. University of Tartu. Available at: http://www.hpc.ut.ee, [Accessed 26 April 2016].

Ury, A. G. (2013). Storing and interpreting genomic information in widely deployed electronic health record systems. Genetics in Medicine 15(10), 779-85.

van Calcar, S. C., Gleason, L. A., Lindh, H., Hoffman, G., Rhead, W., Vockley, G., Wolff, J. A., Durkin, M. S. (2007). 2-Methylbutyryl-CoA dehydrogenase deficiency in Hmong infants identified by expanded newborn screen. *Wisconsin Medical Journal* 106, 12–15.

Vassy, J. L., Lautenbach, D. M., McLaughlin, H. M., Kong, S. W., Christensen, K. D., Krier, J., Kohane, I. S., Feuerman, L. Z., Blumenthal-Barby, J., Roberts, J. S., Lehmann, L. S., Ho, C. Y., Ubel, P. A., MacRae, C. A., Seidman, C. E., Murray, M. F., McGuire, A. L., Rehm, H. L.,

Green, R. C. (2014). The MedSeq Project: a randomized trial of integrating whole genome sequencing into clinical medicine. *Trials* 15, 85.

Venter, J.C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J. *et al*. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Vogenberg, F. R., Isaacson Barash, C., Pursel, M. (2010). Personalized medicine. *Pharmacy and Therapeutics* 35 (11), 624-42.

Welch, B. M., Kawamoto, K. (2013). The need for clinical decision support integrated with the electronic health record for the clinical application of whole genome sequencing information. *Journal of Personalized Medicine* 3, 306-325.

Worthey, E. A., Mayer, A. N., Syverson, G. D., Helbling, D., Bonacci, B. B., Decker, B., Serpe, J. M., Dasu, T., Tschannen, M. R., Veith, R. L., Basehore, M. J., Broeckel, U., Tomita-Mitchell, A., Arca, M. J., Casper, J. T., Margolis, D. A., Bick, D. P., Hessner, M. J., Routes, J. M., Verbsky, J. W., Jacob, H. J., Dimmock, D. P. (2011). Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in Medicine* 13, 255–262.