TALLINN UNIVERSITY OF TECHNOLOGY

School of Information Technology

Department of Software Science

ITC70LT

Romet Saaliste 153695

ESTONIAN GOVERNMENT RELATED CHALLENGES IN PROTECTION OF PERSONAL DATA

Master Thesis

Supervisor: Olaf Manuel Maennel

Professor

Tallinn 2018

Author’s declaration of originality

I hereby certify that I am the sole author of this research. All the used materials, references to the literature and the work of others have been referred to. This research has not been presented for examination anywhere else.

Author: Romet Saaliste

03.05.2018

Annotatsioon

Andmed on tänapäeva infotehnoloogia ajastul kõige alus. Andmeid töötlevatel asutustel või isikutel lasub kohustus neid andmeid kaitsta viisil mis on sätestatud erinevates seadustes, regulatsioonides, eeskirjades jms. Isikuandmete kaitse on antud töös läbivaks jooneks. Töö keskendub teadlikkuse probleemile Eesti riigiasutuste seas, kus registrites legaalselt avaldatakse isiku kohta käivaid granulaarseid andmete komplekte, mida kombineerides võib tekkida oht andmelekkele ja isikute profileerimisele.

Käesolev töö annab ülevaate Euroopa Parlamendi poolt heakskiidetud isikuandmete kaitse üldmäärusest (GDPR), millega asendatakse senine andmekaitse direktiiv 95/46/EÜ. Samuti vaatleb antud töö GDPRi täiendavaid Eesti seadusandlusi, mille alusel registripidajad isikuandmeid avalikult väljastavad ning analüüsib Eesti avalike asutuste väljakutseid seoses GDPRi rakendumisega. Erinevate andmehulkade kokku viimine ja isikustamine võib tihtipeale anda tulemusi, mis võib vastuollu minna konkreetse registri pidamise määruse ja andmete kaitse eesmärgiga. Töö teine pool keskendubki Eesti avalike asutuste registritest päringute tegemisele ja OSINT tööriista loomisele, mille käigus üritatakse identifitseerida Eesti kodanikke. OSINT tööriista loomise eesmärk ei ole pakkuda avalikkusele võimalust isikuandmete pärimiseks, vaid tõendamaks, et erinevate registrite kombineerimisel on üsna lihtsate vahenditega võimalik seda teha. Kokkuleppel registripidajatega kasutatakse antud tööriista üksnes eesmärgi saavutamiseks.

Töö analüüsi põhjal järeldati, et peamiselt liigituvad Eesti avalike asutuste väljakutsed registripidajatena juriidilisteks, organisatoorseteks ja tehnilisteks. Tihti on sama väljakutse seotud kõikide liigitustega, mille lahendamisele erinevate valdkondade inimesed peavad ühtselt keskenduma. Üheks suuremaks juriidiliseks väljakutseks lähiaastatel on GDPRi territoriaalse kohaldamisala printsiibi rakendamine, mille alusel GDPR rakendub ka väljaspool Euroopa Liitu registreeritud teenusepakkujatele, kui nende teenuse osutamine toimub Euroopa Liidus liikmesriigis. Organisatoorsel tasandil tuleb arvestada kasvava administratiivse tööga, töötajate koolitamisega ning sõltumatu andmekaitse ametniku määramisega. Üheks tehniliseks väljakutseteks on kindlasti

säilitamise piirangu printsiip, mille järgimine võib osutuda keeruliseks ja kulukaks. Lisaks esitati üldised soovitused GDPRi rakendamise elutsükli järgmiseks.

Töö raames loodud stsenaariumiga ja OSINT tööriistaga oli võimalik Eesti avalike asutuste registritest pärida isikuandmeid. OSINT tööriista toimimise meetodi valideerimise käigus tuvastati, et suuremal osal juhtudel on võimalik teostada andmesubjekti identifitseerimist.

Töö on kirjutatud inglise keeles ning sisaldab teksti 76 leheküljel, viite peatükki, neljateistkümmet joonist ja kaheksat tabelit.

Abstract

In the era of information technology, the data is a basis for everything. Organizations or people processing the data have an obligation to protect the data in a way that it is regulated in different acts, regulations, directives etc. Protecting personal data is the cornerstone of this thesis concentrating on awareness problem among Estonian

This thesis will give an overview about the General Data Protection Regulation (GDPR), that repeals the previous directive 95/46/EC. Additionally, this thesis reviews Estonian legislative acts that are extending the GDPR, being the reason registrars publish personal data. This thesis will also analyse the challenges the Estonian government institution may face in order to stay compliant with the GDPR. Merging different data sets into one may result in a way that contradicts with the statutes of single registry and the purpose of data protection. The second part of the thesis concentrates on making queries from Estonian government registers and building the OSINT tool in order to try identifying Estonian citizens. The purpose of building an OSINT tool is not to distribute private data for public use, rather to prove that it is possible when combining different datasets. As agreed with the registrants, the tool is used only to achieve the purpose mentioned above only.

Based on the analysis of the work, it is concluded that the challenges of Estonian government institutions are categorized as legal, organizational and technical. Often, the same challenge is associated to all categorizations where people with different expertise need to concentrate as whole in order to fulfil the goal. As GDPR is an ongoing process some general suggestions were given for the preparation of GDPR lifecycle.

With the scenario and OSINT tool, that was built while writing this thesis, it was possible to query personal data from publicly available Estonian registries. While validating the method of OSINT tool, it was concluded that in majority of the cases it is possible to identify the data subject.

The thesis is in English and contains 76 pages of text, five chapters, fourteen figures and eight tables.

Table of abbreviations and terms

API	Application Programming Interface
CERT	Computer Emergency Response Team
CSS	Cascading Style Sheets
DPA	Data Protection Authorities
DPO	Data Protection Officer
EU	European Union
GCI	Global Cybersecurity Index
GDPR	General Data Protection Regulation
GZIP	GNU zip (Open Source algorithm for file compression)
HDD	Hard Disk Drive
ISKE	Three-level IT baseline security system
JSON	JavaScript Object Notation
PDF	Portable Document Format
PPII	Protection of Personal Identifiable Information
OSINT	Open Source Intelligence
PDF	Portable Document Format
REST	Representational State Transfer
US	United States
CSV	Comma-separated Values

X-ROAD	Estonian government secure data exchange platform

Table of contents

List of figures

List of tables

1. Introduction

In the last decade the evolution of technology has grown rapidly. It surrounds us in every step and in lots of forms. Mobile phones in our pockets nowadays have twice as computing power than computers that were used in Apollo 11 mission in 1969 [1]. Every day, 2.5 quintillion bytes of data is created — so much that 90% of the data in the world today has been created in the last two years alone [2]. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. A lot of this kind of information is used for analysing, to make a decision that will make our lives more comfortable and safe.

With great amount of data comes great amount of responsibility. One can possess personal information against other and substantially increase his/her influence in certain circles in a way that could affect the data subject in a negative manner. To prevent this scenario from happening both EU and Estonia have adapted laws and regulations to reasonably limit the publicly available personal data in order to protect the rights of natural person.

This thesis will give an overview about the nature of GDPR and how it relates to the Estonian jurisdictional framework. What are the main goals, roles and principles of GDPR and how to achieve it. This thesis is a starting point to familiarize with the topic of GDPR and how does it relate to Estonian jurisdictional acts. Additionally, an analyse is done to describe the challenges the Estonian government institutions may face while adapting GDPR.

Personal data can be found online (see *Table 3*) whether it is uploaded by the data subject, a friend or a government institution. Knowing the right techniques and sources could potentially lead to profiling people and misuse of their personal data. There are different OSINT tools available online that can be used to collect such data and if used in a proper manner it may be a valuable source of intel information. Additionally, to the juridical analyse, a new OSINT tool is built that is able to collect publicly available personal information from Estonian government registers and portals to showcase that

more efficient safeguards are needed to protect against potential automated queries extracting Estonian citizens personal data.

The main contributions of this thesis are:

- the analyse of the combined effects of GDPR and Estonian registry acts to understand the challenges of Estonian governments may face.
- the development of OSINT tool that extracts personal data from Estonian governments registers.

This thesis concentrates on awareness problem among Estonian government institutions who are legally publishing granular sets of personal data. Combining information from different datasets could potentially lead to a wide scale data leakage incidents and profiling people through their personal data. It is essential to understand that safeguards that were effective in the past, needs to be re-evaluated to mitigate the risk of cybercrime. Describing and analysing the GDPR while building an OSINT tool showcases the nature of personal data and increases the awareness level about the impact combined data sets and safeguards against automated tools may have.

2. Related work

This chapter looks at existing work related to GDPR combined analyses of Estonian government registers and OSINT tools that is able to extract data from Estonian government registers.

The purposes and different analyses about GDPR are covered in variety of sources with different angels. S. Sillaots has written a paper [3] that analyses GDPR harmonization and impacts on Estonia discussing the effectiveness of the means proposed in the GDPR. The purpose of this paper is to determine whether the regulations will reach the desired main purposes that is set by the Commission of European Union. The paper addresses 4 goals:

1. strengthen the protection of individuals by increasing transparency when processing personal data;

2. the explicit regulation of the duties and responsibilities of the controller and the processor;

3. the right to erasure hoping to improve the control of the data subject over one's data;

4. strengthen the control over one's data by introducing the right to data portability.

It is concluded that 3 measures out of 4 are effective in order to achieve a higher level of protection. According to the author of the referred paper the principles introducing increasing transparency, responsibilities of data processor and controller and data portability are all welcomed changes in EU and Estonian law. The principle relating to the right to erasure will not be considered effective measure and further specification is needed. Although it seems that the principle serves the best interest of data subject, the reality of this measure is difficult to achieve.

Although there are a great number of analyses available about the GDPR in different forms, it is quite difficult to find a paper describing and analyzing the GDPR related challenges about Estonian government registers.

Another goal of the thesis concentrates on building an OSINT tool that extracts data from Estonian government portals and registers. OSINT tool is highly valuable for gathering intel information about something or someone. There are various OSINT tools available ([4], [5]) with different usage and performance capabilities. The tool called Maltego [5] is developed by Paterva for various operating systems. It has both graphic and command line interface and can be used to enumerate various information such as IP addresses, domain names, people e-mail addresses, phone numbers and social groups associated to the data subject. Another OSINT tool [4] developed by IntelTechniques allows user to query data subject's social media related information. For example, Facebook and Instagram account related information such as email addresses, telephone numbers, relationships with other users and many more. It has a capability to perform social media related custom searches that are powered by Google search engine. The selection of tools is almost limitless and choosing the right one can be a challenge. A report [6] called "Web Scraping: Application and Tools" introduces different web scraping tools and how to decide the most suitable one for particular task. The report divides the mentioned tools in two categories as partial and complete tools. Partial tools are typically plugins focusing on specific scraping technique. The complete tools offer more general scraping possibilities with graphic user interface and API-s.

J.K. Mikli concentrates on on personal data web scraping techniques based on social media platform Facebook [7]. The paper involves a study of automated data gathering methods that utilizes social networking websites as a source. The basis for the web scraping was python script that had to be modified in order to adapt with the changes that Facebook has made in their template and security policies. The goal of this paper was:

1. to study the scripts that gather personal data from different Facebook accounts;
2. to find out if it is possible and in what extent to access personal data via Facebook without special access privileges to Facebook user profile.

It was concluded that it is possible to modify outdated web scraping script in order to scrape data from Facebook. The data gathered varies depending on the privacy settings of user profile. In most of the cases it was possible to determine some amount of

personal data and the relationships with other users. These relationships can be used to determine the profile of a data subject in order to prepare an attack vector for future.

There is a various number of papers and websites on the topic of scraping personal data, but it is quite challenging to find a paper concentrating on an Estonian governments registers.

3. Data Protection Regulations

This chapter gives an overview of the GDPR and Estonian registry act's objectives, definitions, responsibilities and principals of different parties. This overview gives a basic understanding of the regulations and not every detail nor exception is covered in the following paragraphs. A more detailed understanding can be find in GDPR [8] and Estonian legislation [9]. This chapter also concentrates on the challenges different parties may face and possible threats they may encounter.

3.1. Overview of the GDPR

On April 27, 2016 the European Parliament and the Counsel of the European Union adopted the EU Regulation 2016/679 (GDPR) which will be directly applicable in all Member states and the enforcement date will be May 25, 2018. This regulation repeals the previous directive 95/46/EC [10]. The aim of GDPR is to harmonize the degree of data protection across EU nations protecting citizens from privacy and data breaches. It will provide nations standardized data protection laws within EU, allowing people to better understand their rights and how their data is being used within the EU states. While new regulation brings more clarity to citizens the challenge of standardization of the laws remains to be fulfilled by the governments and companies of each Member state.

GDPR objectives [11]:

1. "Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data";
2. "Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data";
3. "The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data".

The main purpose of GDPR is to help data subjects better understand the use of their personal data and if necessary to give regulatory measures of protecting it.

3.1.1. Terminology

To understand the fundamentals of GDPR it is essential to clarify what is personal data, what is data processing, who are data controllers and processors and what are their responsibilities.

According to the article 4 [12] point 1 of GDPR personal data means "any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to the identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." This definition of personal data states that data connected directly to the data subject or it can be connected to him/her through other means is considered personal data. For example, dataset "Andres Tamm is non-religious" is not considered personal data to the data controller operating in Estonia as there are fair amount of people named Andres Tamm ([13], [14]) who are non-religious [15] at the same time living in Estonia. Looking at the same dataset in Malta that is relatively smaller country than Estonia and not populated by Estonians, it is possible to say that the data controller holds personal data as the possibility to find two or more people named Jaan Tamm who are non-religious is unlikely. As it can be seen from the previous example, different datasets may be interpreted differently in terms of personal data regarding the attributes it holds (i.e. geographical, physiological, cultural or social identity).

The same article [12] point 2 defines the term "processing" that means "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction." This involves any use of personal data that is not used for personal reasons only. According to the article 6 [16] paragraph 1 of GDPR processing is lawful to the extent that at least one of the following applies:

1. "the data subject has given consent to the processing of his or her personal data for one or more specific purposes";

2. “processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract”;

3. “processing is necessary for compliance with a legal obligation to which the controller is subject”;

4. “processing is necessary in order to protect the vital interests of the data subject or of another natural person”;

5. “processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller”;

6. “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child”.

The term “processing” is quite broad and applies to almost anything that is done or to be done related to personal data. This also includes collection, storage and erasure of data. The Regulation for data processing does not apply if it is performed by a natural person for domestic or personal purpose, processing the data manually or non-structured way, processing the data of a deceased or processing anonymous information.

According to the article 4 [12] point 7 of GDPR the term “controller” means “the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data. The purposes and means of such processing are determined by Union or Member state law, the controller or the specific criteria for its nomination may be provided for by Union or Member state law.” This means that the organization or authority who acts as a controller has to appoint means of processing the personal data. The controller remains responsible regardless whether they gather the data from the subjects directly or not. For example, if an E-shop selling calendars is using an e-mail newsletter service provider for sending out weekly newsletters, the E-shop is considered to be a data controller in an extent that is related to the weekly newsletter process.

According to the article 4 [12] point 8 of GDPR the term “processor” means “a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.” Using the previous E-shop and e-mail newsletter service provider example, the service provider is considered as a data processor in that context. Nevertheless, data controller and data processor can be the same organization, if the E-shop handles the newsletter service by themselves. Extending this example where the E-shop is using a payroll service from accountant company, who uses the data they have gathered from the E-shop client to provide salary benchmark analysis. In this example the accountant company is a processor in terms of payroll service provider and at the same time a controller in terms of a salary benchmark service provider for their own purposes. From those examples the role of the controller and processor directly depend on the purpose of data processing.

3.1.2. Responsibilities of different parties

Regarding the responsibilities the controller is liable for, and should be able to demonstrate, the compliance with the regulations principles stated in the article 5 [17] paragraph 1. The controller should be able to follow these principles when processing personal data. The principles are as follows:

1. lawfulness, fairness, transparency;
2. purpose limitation;
3. data minimization;
4. accuracy;
5. integrity and confidentiality;
6. storage limitation;
7. accountability.

The first principle means that data should be treated in terms with the GDPR in fair manner that is transparent to the data subject (i.e. communicating personal data breach to data subject). The purpose limitation principle states that the data should be gathered only with the legitimate and specified purposes and is processed in a way that does not

conflict with the initial purposes. According to data minimization principle the personal data should be gathered in a way that is adequate and limited to a point that is necessary for the data processing purposes. At the same time personal data should be accurate (possibly kept up to date) and processed in a way that takes measures against data loss while ensures the appropriate security against unauthorized access. The principle of storage limitation represents keeping the data in a way that is identifiable no longer than necessary until the purpose of processing the personal data has ended. The accountability principle means that the controller shall be responsible for and be able to demonstrate compliance with GDPR. Additionally, if the data subject is revoking the access to his/her data, the contact point will be data controller. Receiving the request, the controller passes information to the data processor who takes necessary agreed actions in order to satisfy data subjects request.

According to GDPR article 28 [18] "where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this regulation and ensure the protection of rights of the data subject." This means that the data processor is responsible for implementing necessary controls that comply with GDPR. Furthermore, the controller must appoint a processor in a binding form (written agreement) who must act only under the documented instructions of controller. The processor must establish confidentiality agreements on all personnel processing the personal data and takes all measures required to respect the article 32 [19] "Security processing" which will be covered in the paragraphs to come. Article 28 [18] paragraph 3 point d states that at the same time the processor must follow the rules when engaging another processor (written authorisation from controller), assist the controller in receiving compliance from DPA-s related to the processor activities, provide information to the controller to demonstrate the compliance with the GDPR and return or destroy the personal data after the processing service agreement has ended (except if required by the EU Member state law).

All the records about processing activities shall be collected and preserved by the controller or the controller's representative in a registry. According to article 30 [20] the records should contain:

1. “name and contact details of the controller and, where applicable, the joint controller, the controller’s representative and the data protection officer”;

2. “purposes of the processing”;

3. “a description of the categories of data subjects and the categories of personal data”;

4. “the categories of recipients to whom the personal data has been or will be disclosed to including recipients in third countries or international organisations”;

5. “where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers referred to in the second subparagraph of article 49(1), the documentation of suitable safeguards”;

6. “the envisaged time limits for erasure of the different categories of data, where possible”;

7. “a general description of the technical and organisational security measures referred to in article 32(1), where possible”.

3.1.3. Security of personal data

To get a better overview of the regulation it is important to understand the GDPR [19] section 2 “Security of personal data”. As processing personal data in a safe manner is the primary purpose of GDPR, the security of processing and appropriate notifications are one of the focus points of this section. Considering the fast-developing nature of science and technology, the context and purpose of processing the controller and processor are obligated to implement the proper security level with measures that are technically and organisationally necessary. As the article 32 [19] states, “these measures include inter alia as appropriate pseudonymising and encryption of personal data, ensuring confidentiality, availability and resilience of processing services, restoration of availability during an incident and a process for regular testing to ensure the effectiveness of secure processing.” GDPR recommends using the approved code of conduct as referred to in article 40 [21] or approved certification mechanism as referred to in article 42 [22]. At the same time the article states that in order to assess the necessary level of security involved in data processing, the risks must be taken into

consideration. Risks that need to be assessed are unlawful or accidental destruction, loss, alternation, unauthorized access to personal data, transmission, storage or other processing activities.

Regarding the appropriate notifications during the incident article 33 [23] concentrates on notifying the DPA-s and article 34 [24] concentrates on notifying the data subject. In case of personal data breach controller must notify appropriate DPA-s without undue delay, but no later than 72 hours after the incident has occurred. The data processor is obligated to notify the data controller as soon as possible after personal data breach. The notification must contain at least information about the scope of personal data breach, contacts of additional information providers, description of possible outcome of the breach and measures that are taken or will be taken into action to mitigate the risks. If all the information is not possible to provide at the same time, it is acceptable to provide information in phases without undue further delay. Communicating the personal data breach to the data subject depends on whether the compromised information is likely to result in a high risk to the rights and freedom of natural person. Article 34 paragraph 3 [24] states that communication is not necessary when at least one condition is met:

1. "the controller has implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the personal data breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption";

2. "the controller has taken subsequent measures which ensure that the high risk to the rights and freedoms of data subjects referred to in paragraph 1 [24] is no longer likely to materialise";

3. "it would involve disproportionate effort. In such a case, there shall instead be a public communication or similar measure whereby the data subjects are informed in an equally effective manner".

3.2. Overview of the Estonian legislation regarding electronical registries

Estonian Public Information Act [25] regulates all the relevant regarding management of databases/registries. Paragraph 43^1 point 1 states that database is a structured body of

data processed within an information system of the state, local government or other person in public law or person in private law performing public duties which is established and used for the performance of functions provided in an Act, legislation issued on the basis thereof or an international agreement. Meaning that any registry related to the government is established by a legal act or statutes. This paragraph gives an overview of the acts of Estonian registries mentioned in paragraph "4.2 Manual data gathering method", where a scenario is described to manually scrape data from different Estonian registries. This scenario was built during the research with an aim to gather as much data about the data subject as possible and create a pathway for automated queries. This paragraph will concentrate on acts and statutes only where publicly accessible personal data or relevant data relating to the data gathering scenario is mentioned.

3.2.1. The Marital Property Register Act

The Marital Property Register is an Estonian state registry for the registration of property rights. In case of two people are getting a divorce, the property rights between participants are settled according to the agreement written prior marriage in marital property registry. The Marital Property Register Act [26] article 6 section 1 states that everyone can access the information entered on a registry card of the marital property register and obtain printouts thereof. The registry includes registry card, the registry file and the registry journal. The registry card has a unique numeration and entries about property rights. The registry file holds application entry, marital property contract and court decisions related information. In the registry journal the applications for entries are registered. According to the Marital Property Register Act [26] paragraph 17 section 2, the following personal data is entered on a registry card:

1. "given names and surnames of spouses";
2. "personal identification codes of both spouses or, in the absence thereof, dates of birth".

3.2.2. The statutes of Ametlikud Teadaanded

Ametlikud Teadaanded is an official online publication portal of the Republic of Estonia, where announcements and invitations are published. The official publication portal is an information system of the state. The statutes of Ametlikud Teadaanded [27]

paragraph 2 section 1 states that the controller of the portal is the Ministry of Justice. According to the same statute's [27] paragraph 8 section 1 the notice must contain at least the following personal data related information:

1. "if the notice specifies a natural person, the given name and surname of the person in the nominative case and the personal identification code verified with the population register; if the person has no personal identification code, the date of birth will be indicated in the notice, provided that it is known";

2. "The contact details of the data provider and, if necessary, the contact details of the publisher".

3.2.3. The statutes of Motor Register

Motor register is established by the government of the Republic of Estonia with a purpose of keeping records of vehicles, ships and jetties under 12 meters in length, driving licenses and other documents proving the right to drive, digital tachograph cards and registry pledges. The Traffic act [28] paragraph 173 section 3 states that the controller of motor registry is Estonian Road Administration and according to the paragraph 184 section 3 information entered in the motor register is public, except for:

1. "data of natural and legal persons";

2. "numbers of registration certificates";

3. "registration numbers";

4. "identification numbers (VIN, TIN, HIN and CIN)";

5. "information relating to health certificates and examinations".

The Traffic Act [28] section 4 states that in the event of a justified legitimate interest, the information with restricted access specified in subsection (3) of this section may be given to third parties. A justified legitimate interest is verified and the release decided by the Road Administration. At the same time Estonian Road Administration has right to release data to the EU Member states according to the rules of GDPR and to the states whose level of protection of personal data has been assessed as sufficient by the European Commission.

3.2.4. Commercial Code Act

The Commercial Code Act [29] paragraph 22 states that the controller of the portal is the Ministry of Justice. The composition of commercial register includes the registry card, business files and registry files. A separate registry card will be opened for every entry and for every registry entry a business file will be opened which contains documents that are presented by the undertaking, court or trustee in the event of bankruptcy. A registry file contains proof of payment about the state fees and other documents about undertakings that are not stored in the business files. According to Commercial Code Act [29] paragraph 64 the following personal data shall be entered on a registry card:

1. "the name or business name, personal identification code or registry code of the contact person and the Estonian address for delivery of the declarations of intent addressed to the undertaking and the procedural documents of the undertaking, and also the e-mail address of the contact person".

The Commercial Code Act [29] paragraph 28 states that entries in the commercial register are public and everyone has the right to examine the registry cards and the business files.

3.2.5. Land Cadastre Act

The purpose of Land Cadastre is to reflect the value of land, the natural status of land and the use of land. Also, it is important to ensure the quality of that information and that it is preserved and made available to the public. The controller of land cadastre is the Land Board. According to the Land Cadastre Act [30] paragraph 9 section 2 the cadastre register contains the following personal data pertaining to a cadastral unit:

1. "name, address and personal identification code of the owner or superficiary; in case of a legal person, the seat, postal address and registration number".

3.2.6. Maritime Property Act

According to the Maritime Property Act [31] article 5 and article 4 section 2 the ship registry is public and it is maintained pursuant to this Act and to the Law of Ship Flag and Ship Registers Act [32]. Additionally, the Law of Ship Flag and Ship Registers Act paragraph 47 states that the following personal data is entered to the register:

1. “the person’s name and Estonian personal identification code, or in the absence of the Estonian personal identification code, the foreign personal identification code or other identification replacing it and the date of birth (day, month, year) shall be entered in the ship registry. The name of a legal person shall be entered in the ship registry together with the registry code or registration number if the person is subject to entry in the register”.

3.2.7. The statutes of the Database of Artistic Associations

The purpose of artistic associations is to promote a single creative area and support the creative activities of its members. The objective of keeping the database of artistic associations is to keep records of artists. The controller of the database of artistic associations is the Ministry of Culture. The database is regulated by the statutes of the Database of Artistic Associations [33] where paragraph 6 section 2 states that regarding the personal data, the artistic association database must contain following personal data related information:

1. “given name, surname and personal identification code”.

It has to be stressed out that the statutes of the Database of Artistic Associations [33] does not regulate that personal data is published. Nevertheless, the database of artistic associations publishes personal data.

3.3. Combined effects of the regulations for Estonian registers

Adapting the requirements of GDPR, that is repealing Directive 95/46/EC, is a challenge to the EU Member states and companies based outside of EU. Organizational and institutional parties must merge into the context of GDPR in the period of 2 years, which will end in May 25, 2018. The new regulation is made to strengthen the protection of people through increased transparency of processing when dealing with personal data. This allows people to get a better understanding of what kind of data is collected by the institution, what are the roles and responsibilities of different organizations and which privacy policies are implemented. GDPR adds an additional detailed description explaining who are data controllers and processors and what does data processing involve. For the perspective of data subject, more rights have been introduced that gives better control over personal data (i.e. data portability and right to

erasure). This expected and more specified approach in the form of GDPR is challenging for different organizations in various fields (i.e. from the information technology giants to law firms). Additionally, to the obligation of adapting GDPR, they are forced to adapt the new regulation considering the Member state law to which the controller is subject to. In these following sections 3.3.1, 3.3.2 and 3.3.3 the combined effects of GDPR and Estonian registry acts of personal data is analysed. The definition of personal data is analysed in order to get a better understanding of the context. Followed by the discussion of the effects GDPR may have to Estonian government institutions and which juridical, organizational and technical challenges are most likely encountered to understand what kind of changes the GDPR will bring. The analyse about registers and portals are done emphasising the challenges governments may face while adapting new regulations. Finally, the possible theoretical threats of combining GDPR and Estonian register acts are discussed.

3.3.1. Juridical challenges of protecting personal data

As already mentioned in the section "3.1.1 Terminology", the definition "personal data" means any information that relates to identified or identifiable persons is personal data. The identifiable natural person can be identified directly or indirectly. These identifiable attributes could be name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

In order to understand what exactly personal data is, it is necessary to analyse the definition individually. According to MSc Jim Seamans´s post [34] in LinkedIn environment the broad term "any information" means that it is neutral from the technical format of presentation and if the type of information is objective or subjective. For example, there is a USB flash drive with the content of persons passport picture (including identification code, passport number etc) and PDF file with a content of the Holy Bible. The USB flash drive is a neutral technical format of presentation of personal data, while the passport picture represents most likely objective information. Subjective information is the PDF with the content of the Holy Bible as it is an assumption that the person on that picture is Christian. Looking into the term "identified/identifiable" which refers to a result, where it is possible to point to one particular person after analysing the data attributes that were available. The data

connected directly to the subject of interest or it can be connected to through other means to the data subject is considered identified. See the example about the dataset "Andres Tamm is non-religious" described in section 3.1.1 "Terminology". The term "natural person" refers to a living person that is not deceased, unborn nor a legal person. For example, dataset "Andres Tamm is the owner of SpaceTech AS with the registry code of 10000000" is not considered as a natural person. However, the dataset "Andres Tamm with the identification code 31111111111" represents a natural person.

According to the GDPR article 6 [16] paragraph 3 point (b) states that "the basis for the processing referred to in point (c) and (e) of paragraph 1 shall be laid down by Member state law. Paragraph 1 point (c) and (e) are stating that it is allowed by GDPR, if the controller or processor has a legal obligation to process personal data." The paragraph 3 point (b) gives Member state right to process personal data according to the rules of Member state law. In addition, the GDPR article 86 [35] states that "personal data in official documents held by a public authority or a public body or a private body for the performance of a task carried out in the public interest may be disclosed by the authority or body in accordance with Union or Member state law to which the public authority or body is subject in order to reconcile public access to official documents with the right to the protection of personal data pursuant to this Regulation." For example, the Member state of Estonia Marital Property law states that everyone has a right to access the information that is on a registry card. Registry card contains information such as card number, names and surnames of spouses, personal identification code etc. This example illustrates that although dataset presented with the name, surname and identification code combination is in conflict with the GDPR principle of protecting personal data, the Estonian Marital Property Act extends the GDPR. In legal perspective it is acceptable solution. *Table 1* demonstrates the Estonian domestic law extending the GDPR among the registers mentioned in paragraph 4.2 Manual data gathering method.

Table 1. Estonian domestic law extending GDPR.

Register name	**Contains personal data**	**Publishes personal data**	**Publishing personal data allowed**
Marital property	Yes	Yes	Yes
Official publication	Yes	Yes	Yes
Motor register	Yes	No	No

Register name	Contains personal data	Publishes personal data	Publishing personal data allowed
Commercial register	Yes	Yes	Yes
Land cadastre	Yes	No	Yes
Ship register	Yes	Yes	Yes
Artistic associations	Yes	Yes	No

According to information displayed in *Table 1* motor and land cadastre registers are not publishing personal data. All other registers mentioned (except for artistic associations) are using the GDPR article 6 paragraph 3 point (b) and article 86 to process personal data according to the rules of Member state law. The database of Artistic Association does not have legal right to publish person's name, date of birth or id code combination as the statutes of the Database of Artistic Associations [33] does not regulate that personal data is published. This means that the Ministry of Culture has to evaluate the necessity of publishing person's name, date of birth and identification code and determine whether to update the statutes of the Database of Artistic Associations. When contacting the Ministry of Culture, it was stated that they have an oral acceptance from the data subject to publish this data. Unfortunately, the ministry does not have a way of representing that agreement. Considering that, it is definitely a risk that needs their attention and mitigation.

To demonstrate the challenges the lawyers may face, it is interesting to point out from the same table that marital property register contains information about vehicle registration number, vehicle identification number (VIN) that is illegal by the Traffic Act [28] paragraph 184 section 3. This is probably the "grey" area of jurisdiction as the Act of Marital Property [26] paragraph 17 section 2 point 5 allows declare objects as joint or separate property and paragraph 7 section 1 allows everyone to access that information. The only unique identifier of the vehicle would be vehicle identification number. This is probably why marital property register publishes data that is forbidden to publish by the Traffic Act. As seen from the examples, different datasets may be interpreted differently in terms of personal data regarding the attributes it holds (i.e. geographical, physiological, cultural or social identity). This may result of many difficult challenges to the lawyers working with GDPR in EU Member states.

Another juridical challenge will be the increased territorial scope. The extended jurisdiction of the GDPR applies to all companies who process the EU citizens personal data, even if the company is located outside of the EU. The GDPR states that the rules of processing personal data by the controllers or processors apply inside or outside the EU. This means that for some organizations it may be complicated to comply with different obligations set by the different data protection acts. For example, the US company, selling their services within the EU member states, has to take US data protection regulations, GDPR and data protection acts of all the Member states it operates in into account. This will be an enormous juridical challenge to online based companies outside of EU. As seen from previous example the companies outside of EU are facing challenges of considering and adapting different kind of regulations. In a situation where no universally accepted international data protection law exists, the principles of GDPR and United State of America PPII [36] holds the leading position when personal data processing of EU citizens are concerned.

In the United States similar legal protection mechanism called PPII is in place that is concentrated on protection of the personal information. The challenge here is that GDPR is more focused on protecting person's right to operate with his/her data. At the same time PPII is more focused on how governments and organizations should operate with the data. Leaving the data subjects rights aside. This is going to be a juridical challenge to interpret the GDPR and PPII in cases where data processing involves both EU and US parties.

Additionally to the PPII a more "domestic" challenge where government institutions and all other organizations have to adapt is the unclarity of different situations. As GDPR is a general regulation, it does not focus on the details with specific cases. This means that related to GDPR some various unclear cases will merge that have to be settled in the court rooms to get an understanding how the judges are interpreting specific cases in a context of GDPR. For example, the processing of underaged children's data is going to be a challenge where the arguable question is wheatear the consent was given by the parent or not.

Estonian Ministry of Justice has conducted a draft of data protection implementation act [37] which gathers all the changes needed to be done in Estonian public registers regarding GDPR. This act comes into force on May 25, 2018. According to the act's

paragraph 1 section 1 the marital property register has to collect authentication information before enabling the access to marital property data. It is the authors experience that authentication was implemented on March 1, 2018 before the GDPR comes into force. Mentioned implementation made querying with automated means, discussed in paragraph 4.3 “Building an open source intelligence (OSINT) tool”, a lot more difficult for the author of this thesis. This action shows clearly how an implemented safeguard can decrease the risk of data exposure.

3.3.2. Organisational challenges of protecting personal data

GDPR means additional new obligations and challenges to the controller and processor as well to protect the personal data. For example, previous directive 95/46/EC article 4 [10] section 1 imposes legal compliance obligations to controllers when GDPR imposes the same obligations to controllers and processor. This means that processors, with previous directive had liabilities to the controller related to the contract, now have legal compliance obligations to GDPR. This also includes obligation to maintain the records of processing and when necessary to interact with the DPA-s. The importance of having measures for data protection and the processors responsibility have increased. This involves significant investments to the controller and processor, but this results in more transparent environment and strict processes to protect personal data.

Looking into the GDPR article 28 [18] the controller has to document all the instructions given to processor while in previous directive, the form of instructions are not regulated. The same article states that processors confidentiality obligations must be agreed with the personnel dealing with processing. This also means that all existing agreements are affected and renegotiation may be needed. For government institution this may be a huge challenge as not only it is needed to revise all the contracts with the processor, it is necessary to make detailed instruction how the data processing should take place. Another challenge the controller may face is when processor realises that instructions that are already documented, are conflicting with the applicable EU law. The processor must indicate this issue as soon as possible to the controller whose responsibility is to revise the instructions compliant to EU law. It is important to stress out that article 28 [18] section 10 states that if processor does not follow the documented instructions given by the controller and determines the means and purpose of data processing on its own, they are automatically considered as a controller by

GDPR. This means that the previous processor is a responsible to full compliance obligations of a controller. The article 82 [38] sections 1 states that if the data subject has suffered material or non-material damage as a result of an infringement of this regulation, he/she shall have the right to receive compensation from the controller or processor for the damage suffered. This implies that if processor has not acted according to the lawful instructions of controller, the processor has to compensate the damaged suffered to the data subject. This GDPR article makes the controllers and processors put an effort into thoughts and actions for reorganizing and improving the process of processing personal data. In the end the result is more transparent and co-operation between different parties help to pinpoint the existing problems that results with personal data processing.

According to the GDPR article 37 [39] paragraph 1 every public authority processing personal data must appoint a data protection officer. In Estonia this includes state and local authorities (i.e. ministries, administrations, inspectorates). This requirement does not involve courts that act under their own juridical capacity. The data protection officer has to be familiar with data protection principles and local related laws. The person has to know existing relevant technologies and development directions in the field of ICT and is able to assess the possible impacts of the directions to organization processes. Finding that person with these wide range of skill set is a challenge to most of the public authorities considering the rising demand for those people in a short time frame. Also training the employees will be a challenge for the government institutions as they have to make procedural and structural changes in order to adapt to GDPR. Those people have to overlook or redesign their processes and explain it to the rest of the organization employees.

Another challenge the government institutions have to deal with is the obligation to report to the DPA-s. As mentioned in section 3.1.3 “Security of personal data” in case of personal data breach controller must notify appropriate DPA without undue delay but no later than 72 hours after the incident has occurred. This means that incident handling while in time of crises must be in place. The different roles have to be divided before the incident occurs in order to comply with the regulation and meet the 72-hour time frame. Additionally, to the DPA, it is a necessity [40] in Estonian government organizations to notify Estonian CERT immediately when the incident occurs.

Reporting to multiple different institutions can cause a quite a lot of additional administrative work as it is needed to give periodical updates about the issue and answer additional questions the institutions may have.

3.3.3. Technical challenges of protecting personal data

The GDPR article 25 [41] implies that the controller must implement appropriate technical and organisational measures to ensure the compliance with GDPR. This applies to planning phase or to implementation phase in services or products. While GDPR is more focused on the juridical part of data gathering, the data protection processes and solutions are intentionally left unnoticed. For data protection, every member state should implement information security framework that helps them to achieve the goals required in order to comply with GDPR. In Estonia for government institution the mandatory framework is called ISKE, that is developed for Estonian public sector. Estonians adapted the framework from IT Baseline Protection Manual [42]. The goal of ISKE is to ensure the sufficient security level protecting the data processed in information systems. ISKE is a three-level baseline system that is compulsory since August 12, 2004 to all Estonian state and government organisations who handle databases or registers. ISKE [43] is regulated by the system of security measures for information systems.

ISKE framework is a continuous cycle of security implementations and improvements. This means that depending on the level of security the changeable situation and risk reassessment should be done periodically. Although most of the ISKE requirements overlap the requirements of GDPR there are some new obligation introduced by GDPR that expect technical solution. For example, the principle of data erasure (also known as right to be forgotten) and data portability. The first principle allows the data subject to erase his/her personal data. This can be the case when data processing purpose is not relevant anymore or the data subject has withdrawn the consent for data processing. Technically speaking this could be a demanding task, when the data is scattered among different databases without one unique identifier. Another example could be the archived personal data that is stored on an external HDD (i.e. locked away in a safe). Implementing the GDPR, the data erasure principle means that the HDD needs to be scanned for personal data as well. If personal data about data subject would be found, it has to be erased. This requires a strict organizational process or technical capability to

comply to the fullest. Depending on the architecture of the system, technical solution may be quite complex and challenging to achieve that guarantees the erasure of personal data. The second principle (data portability) allows the data subject to transfer their data from one organisation to another. For example, if one insurance company provides more attractive conditions than current insurance company, then the data subject has the right to ask to transport his/her personal data in machine readable way directly to the new insurance company. This means that two insurance companies (controllers) have to implement a technical and secure solution to exchange personal data. These new principles are allowing data subject to make administrative part more convenient and have more control over their data, resulting therefore in better protection of personal data. Technically speaking there is not much of a challenge to exchange information (X-road, REST API), but there definitely are some organisational challenges to tackle with. It must be pointed out that according to GDPR article 20 [44] paragraph 3 government institutions do not have to follow the data portability principle as long the processing is carried out in the public interest or in the exercise of official authority vested in the controller.

A new challenge the government institutions face, that could be categorized as technical and in some cases as organizational, is maintaining a record of processing activities. This regulation (covered in 3.1.2 Responsibilities of different parties) state that personal data processing has to be recorded for a later inspection to determine the purpose of that act. Depending on the institution it can be a challenge to implement as technical solution are relatively expensive and manual record keeping will increase the administrative work.

Another principle is about a storage limitation that represents keeping the data in a way that it is identifiable no longer than necessary, until the purpose of processing the personal data has ended. As the older directive did not emphasise the time of data retention, the GDPR takes the purpose of processing into account. This means that the data has to be retained as long as the purpose of processing remains. The challenge is to adapt the necessary technical controls which allows to anonymize the data in a way it is no longer identifiable. This principle also adapts to the older data processing actions that’s needed to be in accordance to the GDPR. For example, in some institutions there is a great number of documents containing personal data from earlier periods when the

emphasis was not on personal data security. The technical controls adapted have to be able to scan through the documents from earlier period as well in order comply with the principle. In such cases software like Azure Information Protection [45] could be used to scan, identify, label and protect the document. The problem with this tool is that it is developed as a cloud based solutions and Estonian government institutions are forbidden to process personal data in a public cloud environment.

3.3.4. Roadmap for preparation

It may be difficult to understand all the details of GDPR and how does this affect different organizations, but it is important to familiarize with the measures and start developing a plan. In *Figure 1* 5 steps are shown to help visualize the roadmap for preparation.

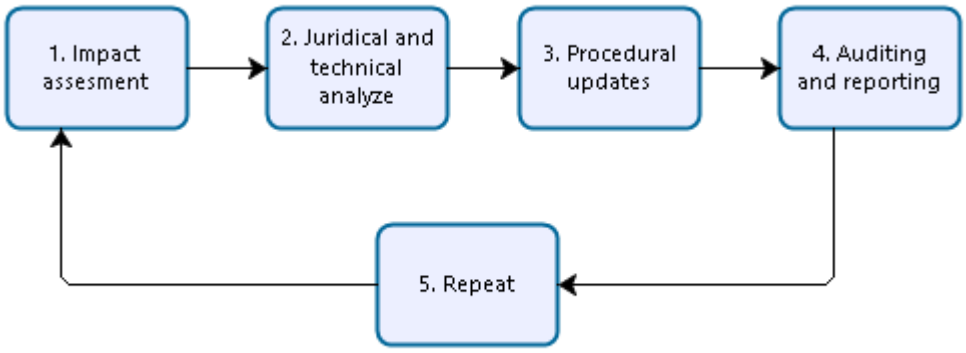


Figure 1. The cycle of GDPR.

In step 1 it is essential to understand which technical and procedural updates the GDPR impacts. In step 2 all the issues related with the juridical and technical challenges need to be overseen. Personal data gathering and storage processes must be evaluated in order to meet the requirements of GDPR. The step 3 concentrates on the creation of privacy requirements, formulation of the tasks of data protection officer, mapping of data flows, the renewal of consent agreements and the incident handling system. For public sector the Data Protection Inspectorate has prepared recommendations that helps to comply with the regulation [46]. This step may be the most time consuming and complex as remodelling the internal processes may need the most effort. In step 4 the auditing and reporting plan has to be generated. The auditing of GDPR compliancy can be done on a slower pace the reporting must be done accordance to the regulation. For example, the

incident reporting must be done in 3 days which may be quite short period when no recovery plan exists. In the step number 5 go through all the steps periodically in order to stay compliant. Within time the aspects of GDPR have become clearer and the technical solutions will develop in a way that is easier for the organizations to stay compliant with the regulation.

3.3.5. Possible threats

In every change, there may be a possibility that something important has been missed that creates an opportunity for the type of people who are considered to be poisonous. This kind of people are willing take advantage of the situation to combine a plan to manipulate or blackmail their victims. The GDPR, being the last and most voluminous regulation concerning the personal data privacy, it has the leading position and influences almost entire world. This creates a lot of pressure mainly on the EU member states and organizations operating within the EU that have to comply with the GDPR.

The main goal of GDPR is to assure natural persons right to the protection of personal data. Therefore, every EU member state have to make sure that the personal data they are controlling or processing is in accordance with the regulation. As already discussed in paragraph ”3.3.1 Juridical challenges of protecting personal data”, member state has the right to process personal data according to the rules of member state law which means that even if the GDPR lays protective measures to the personal data, the member state can extend it by its local jurisdiction. In order to obtain the natural person’s right to the protection of personal data the member state legal acts must be overseen and adapted as well. According to the article “Why anonymous data sometimes isn’t” published in Wired [47] found that 87 percent of the population in the United States, 216 million of 248 million, could likely be uniquely identified by their five-digit ZIP code. This identification was possible because of the gender, identification code and zip code. As seen from the *Table 3*, in Estonia there are quite many registers to extract person’s name and id code combination. It is possible to extract various information out of Estonian identification code itself (gender, date of birth, place of birth hospital etc). This means that it is time to oversee and adapt Estonian legal acts in order to comply the objective of GDPR. Although at the moment Estonian registers legally publish personal data, the real focus should be concentrated to serve the purpose of data privacy.

According to the year 2017 GCI report [48], Estonia is ranked as number one country of cybersecurity in Europe (5th of entire world) with the GCI normalized score of 0,84. The actions and the lead Estonia takes are noticed and followed throughout Europe to reach the greater level of cyber protection. This is a position that Estonians should be proud of and make sure that every register or system, already exists and to be built, has to stand up to the purpose of GDPR. It is important to understand that protecting data is an ongoing cycle and it will never be finished. Therefore, adapting appropriate safeguards against automated queries is crucial at this point. For example, the marital property register publishes data (person name, id code, vehicle license plate, vehicle identification number etc) that may be used against the data subject and be enough to build a platform against this person when trying to execute a social engineering or spear phishing attack. Adding information from powerful social datasets like Facebook and Google, the list of attributes for his scenario would be even greater. With this information available on the web the only obstacle will be finding a right tool or method for gathering it. One implemented safeguard can make the difference in success and a failure as already discussed in subsection 3.3.1 "Juridical challenges of protecting personal data" where authentication measure was implemented making the scenario for automated query a lot more difficult.

Another threat to Estonia would be the non-compliancy with GDPR which may result in remarkable fines. According to GDPR article 83 paragraph 5 [49] which states that depending on the circumstances of each individual case, the non-compliance shall be a subject to administrative fines up to 20 000 000 EUR, or in case of an undertaking, up to 4 % of the total worldwide annual turnover of the preceding financial year. In case of maximum penalty whichever amount is higher will has to be paid. At this moment no-one really knows how strictly this aspect is followed and if there will be a case of martyr. The DPA-s are required to give due, wheatear to impose the fines and in what amount. In Estonia the legal system does not allow administrative fines, therefore the GDPR recital 151 [50] states that the fine is imposed by the supervisory authority in the framework of a misdemeanour procedure, provided that such an application of the rules in those Member states has an equivalent effect to administrative fines imposed by supervisory authorities. Therefore, the competent national courts should take into account the recommendation by the supervisory authority initiating the fine. In any event, the fines imposed should be effective, proportionate and dissuasive. Additionally

to the threats Estonian controllers and processors may face, the GDPR places a responsible decision making and punitive role to the Estonian courts. This means that Estonian courts will have to face even increasingly higher workload in order to protect the personal data of data subjects.

4. Government data about Estonian citizens

Estonian government has a various type of data relating to their citizens. There is a considerable amount of data that categorizes under the definition of personal data, which means it has to be protected in a way where no harm could be done to the data subject by the person holding this information. As GDPR is focused on protecting the data subject, it is possible to "loosen" the terms of the regulation with Estonian domestic law. This means that it is legal to publish personal data when it is regulated by local acts of statutes. Living in a society where legally multiple sets of personal data are published could eventually backfire and harm the data subject and Estonia cybersecurity high reputation. In this chapter and overview what is an OSINT tool will be given and a manual data gathering scenario will be built. Additionally, an OSINT tool building process is described, analysed and validated with Estonian citizens names gathered from Facebook users born in Estonia.

4.1. Overview of OSINT

Open source intelligence (OSINT) is synthesized using publicly available data [51]. OSINT tools are making it easier to find scattered information on publicly available sources using only certain channels to fulfil the purpose of the tool. The sources may vary depending of the nature of the purpose (i.e. social media, API-s etc). For example, the purpose may be to collect all the email addresses related to a university or to collect all the relevant information that could be found about data subject. This all comes down to the knowledge where to look for desired information and a scenario that is written by people in order to maximize the effectiveness of the purpose.

To showcase an example of OSINT tool the website called Intel Techniques [4] is used as an example. Intel Techniques website has a great number of sources to gather information. It is scenario free meaning that everyone visiting this website can create their own scenario and gather different information they are interest in. For example, it is possible to find out information on Facebook, Twitter or Instagram about particular persons with a great ease and no technical background. The beauty of it is that everything can be done with a single OSINT tool. In this website (see *Figure 2*) social

media platform Facebook is used. It has numerous possibilities available to build personal scenarios. One could find out what kind of places the data subject has visited, what pages/photos liked, photos the data subject is tagged on and which past and future event invitations has been sent to the person.

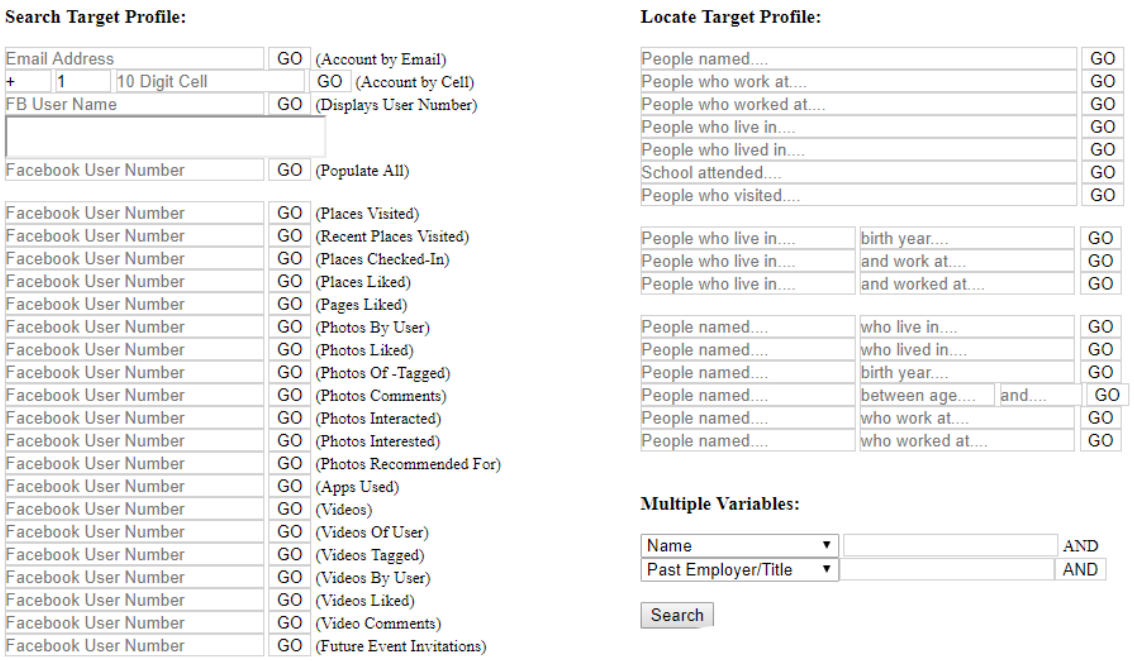


Figure 2. OSINT Facebook tool by Intel Techniques.

It has the ability to use multiple variables in order to conduct personal syntaxes to search for. For the purpose of the confidentiality the figure is partially blurred or edited by the author of this thesis. The only input data that was used to query via Intel Techniques Facebook tool was Facebook user's id number. With the tool itself seven different queries were made and they all returned a match where each result is indicated by the number from 1 to 7 in representing different results for each query (see *Figure 3*). Result number 1 is "The Yoga Barn" located in Bali, which is the last place data subject has checked in. Result number 2 is "Kesselhaus" furniture company and "Rosenvald photography's" Facebook page, which are the last 2 pages the data subject has liked. Result number 3 represent the pictures the person of interest is tagged onto. Results number 4 and 7 are referring to his current working position in Swedbank and his colleagues at that place. Result number 5 and 6 are stating that the data subject is going to the concert of Guns N´ Roses on July 16, 2018 taking place in Tallinn, Estonia and he/she has 2 siblings.

Figure 3. The result of OSINT Facebook tool by Intel Techniques.

It is remarkable that the query was conducted from single platform only, without technical knowledge and the effort that was put into the search was minimal. It is worth to note that all this information is uploaded or linked by the user himself/herself without

probably even realising how much data it is possible to gathered thanks to different kind of OSINT tools.

Understandably this does not mean that using or building an OSINT tool is necessarily a negative thing. If people with no good intentions are using the tools to gather background information, then good-natured people can also use it to their benefit. This means that the better he/she understands what information is available, the easier it is to protect himself/herself.

4.2. Manual data gathering method

This section concentrates on publicly available information only, that is shared by Estonian government ministries and administrations. This section will include only manual queries to databases via ministries websites or publicly available information on administrative institution websites to develop a scenario that is used as a pathway in the section 4.3 “Building an open source intelligence (OSINT) tool”. The section will try to manually find publicly available personal data or related information about data subject “John Doe”. According to the Data Protection Inspectorate article [52] and the ruling of Estonian Supreme Court ([53], [54]) personal data published in Estonian government registers may be re-published only if permitted by law or when data is available in Estonian open data portal [55]. As re-publishing the personal data processed in this thesis is not permitted by law and neither is the personal data available in Estonian open data portal, all the data subjects names are changed to “John Doe” or “Jane Doe” and other personal data attributes (identification code, address, cadastre number, phone number etc) are tampered or blurred.

Estonia has 11 government ministries and all of them are in the scope of this thesis. Ministries have administrative institutions, which this thesis will analyse selectively considering the actuality and interest of information that the administrative institution may have. Starting point for this research is name “John Doe” and possible data sources are presented in the *Table 2*.

Table 2. Government ministries of Estonia.

Name of the ministry	**Website of the ministry**
Ministry of Culture	www.kul.ee/en
Ministry of Defence	www.kmin.ee/en
Ministry of Social Affairs	www.sm.ee/en
Ministry of Economic Affairs and Communications	www.mkm.ee/en
Ministry of Education and Research	www.hm.ee/en
Ministry of the Environment	www.envir.ee/en
Ministry of Finance	www.fin.ee/en
Ministry of Foreign Affairs	www.vm.ee/en
Ministry of the Interior	www.siseministeerium.ee/en
Ministry of Justice	www.just.ee/en
Ministry of Rural Affairs	www.agri.ee/en

All crucial information gathered is categorized as “Accurate” and “Plausible”. Accurate means that information is confirmed by at least two different sources or there could be no misinterpretations. Plausible means that information is gathered from one source only. This method will use any information that is accessible via government in order to link the data to one particular subject of interest. After gathering information about the subject of interest, a similar scenario will be built for the next section 4.3” Building an open source intelligence (OSINT) tool”.

4.2.1. Ministry of Culture

Two main fields administered by minister of Culture are theatres and sports. Out of those two fields, Estonian open data portal has listed Estonian registry of sports [56]. Register of sports will give me 39 coaches in different areas of sports with surname “Doe” but no “John Doe” was listed there. The register of sports does not publish personal data either. Ministry of Culture is also administrating artistic associations database [57] where personal data information about person’s name and date of birth could be found. The same register has a dataset available in the open data portal, where additionally to the name and date of birth combination, id code is published in a format

of comma separated value (csv). Nonetheless, no records were found in the artistic associations database belonging to the data subject “John Doe”.

4.2.2. Ministry of Defence

Expectedly, no public sites were found to query potentially useful information in Minister of Defence website nor the institutions administered by it. Estonian open data portal had nothing relevant to show.

4.2.3. Ministry of Social Affairs

Ministry of Social Affairs deals mainly with healthcare, work, social protection, families and gender equality. Ministry has a register called marital property register [58] managed by Centre of Registers and Information Systems. Marital property register has an opportunity to search by name. Name “John Doe” returns two results (see *Figure 4*).

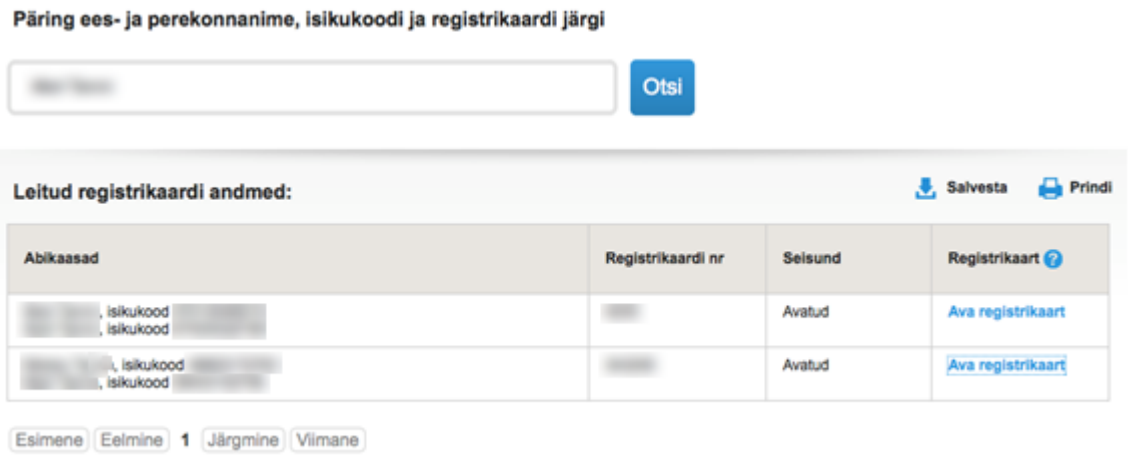


Figure 4. Query by name from marital property register.

It is interesting to see that this registry not only returns the name “John Doe” but also a person to whom our subject of interest is married to. Additionally, register gives an id code of people the query returned. This thesis targets first returned record as a subject of interest who is having an id code of 31111111111. Estonian id code has a certain structure [59] where first digit stands for male (odd number) or female (even number), two following digits stands for year of birth, two following digits month of birth and another two following digits a day of birth. If born before the year 2013 the following three digits are a hospital code the person has been born in. According to that structure it is clear that the subject of interest is “John Doe”, who is a male, born on 1911 11th of November in Võru or Põlva hospital. He is married to “Jane Doe” with an id code

41111111111 who is a female, born on 1911 11th of November in Maarjamõisa Kliinikum (Tartu). Opening the register card of Marital property register reveals additional interesting information such as some of the data subjects interest and his wife's properties. It is possible to understand that she has a real a state with the id number of 1111111 and a bank loan against that property from Nordea bank Finland Plc. According to the data of land cadastre registry this property is located in Võru, Hummuli parish, Vidriku village and the property name is Püssi. Property is 12,84ha large and cadastre number is 72508:002:0716.

The data subject owns a vehicle Mazda 323 with the license plate number 111AAA and Infi trailer with a license plate number 111AA (see *Figure 5*). Entry remark to the Marital property register integrity is made in 13 of October 2012. At this point it is possible to assume that subject of interest is an owner of the vehicles Mazda 323, Infi trailer, Opel Commodore and Opel Frontera. This information is currently plausible as he could have exchange the ownership after the entry remark. Registry card also reveals plausible information that "John Doe" has a company called Johncompany OÜ, but the actuality of this company needs further research. As the entry in registry card about the expiry date is marked as "Valid" it is safe to say that the data subject is still married to "Jane Doe".

Registrikaardi number **Õigsuse märge**	Elektroonilisel registrikaardil toodud andmed on tervikliku:		
ABIKAASAD			
Abikaasade andmed	**Kande alus**		**Kande kehtivus**
isikukood , isikukood	kandeavalduse alusel sisse kantud . Ümber kirjutatud Kohtunikuabi		Kehtiv
KANDED			
Kande number	**Kande sisu**	**Kannete muudatused ja kustutamised ning kande alus**	**Kande kehtivus**
1	Abikaasade poolt valitud varasuhteks on varalahususe varasuhe. 1. Kinnistu nr on lahusvaraks koos kõigi võlakohustustega Nordea Bank Finland Plc ees; 2. OÜ , registrikood jääb lahusvaraks; 3. Sõiduauto MAZDA, tehasetähisega JMZBA rikliku registreerimismärgiga jääb lahusvaraks; 4. INFI TRAILERS, tehasetähisega rikliku registreerimismärgiga jääb lahusvaraks; 5. Sõiduauto OPEL COMMODORE, tehasetähisega , rikliku registreerimismärgiga jääb lahusvaraks; 6. Sõiduauto OPEL FRONTERA, tehasetähisega SED5JMWL , rikliku registreerimismärgiga jääb lahusvaraks. 7. IGASUGUNE abikaasade poolt abielu kestel, pärast käesoleva lepingu sõlmimist, tulevikus soetatav vara, on abikaasade LAHUSVARAKS ja on selle abikaasa omandiks, kelle nimele vara soetatakse.	kandeavalduse alusel sisse kantud Ümber sõnastatud ja ümber kirjutatud Kohtunikuabi	Kehtiv

Figure 5. Register card of marital property register.

Additionally to the martial property register the Ministry of Social Affairs with Centre of Registers and Information Systems are responsible for the official publication portal where additional knowledge about the subject of interest company Johncompany OÜ is gained. According to the portal the companies address was registered into John Doe´s wife Jane´s property. The announcement contains information about the reason the company was shut down. As it occurs it ended due to the compulsory dissolution for not reporting the yearly financial records.

Querying from the social services and benefits data register (STAR), Property portal, Criminal record portal and E-ship registry did not reveal any additional information about the subject of interest. Although the E-ship registry publishes personal data about the person's name and id code (see *Figure 6*), this time no information about "John Doe" is returned.

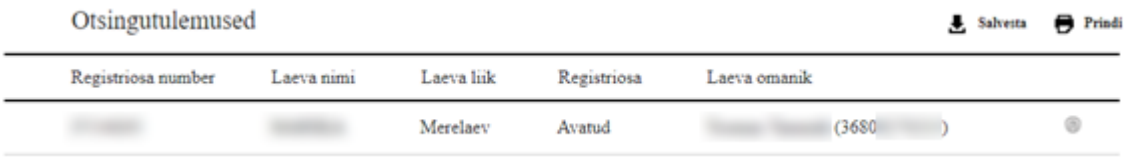


Figure 6. E-ship registry.

4.2.4. Ministry of Economic Affairs and Communication

The main strategic objectives of the Ministry involve governance that encourages entrepreneurship and innovation, an efficient and safe transport system, constantly developing information society and environmentally friendly energy supply at a justified price.

Making a query with vehicle license plate number and vehicle identification number via Estonian Road Administration E-service portal [60] it is possible to learn the main data of the vehicle with insurance and technical inspection data. Information about possible restrictions, possible traffic damages, mileage, technical details and transactions involving the vehicle can be made (see *Figure 7*).

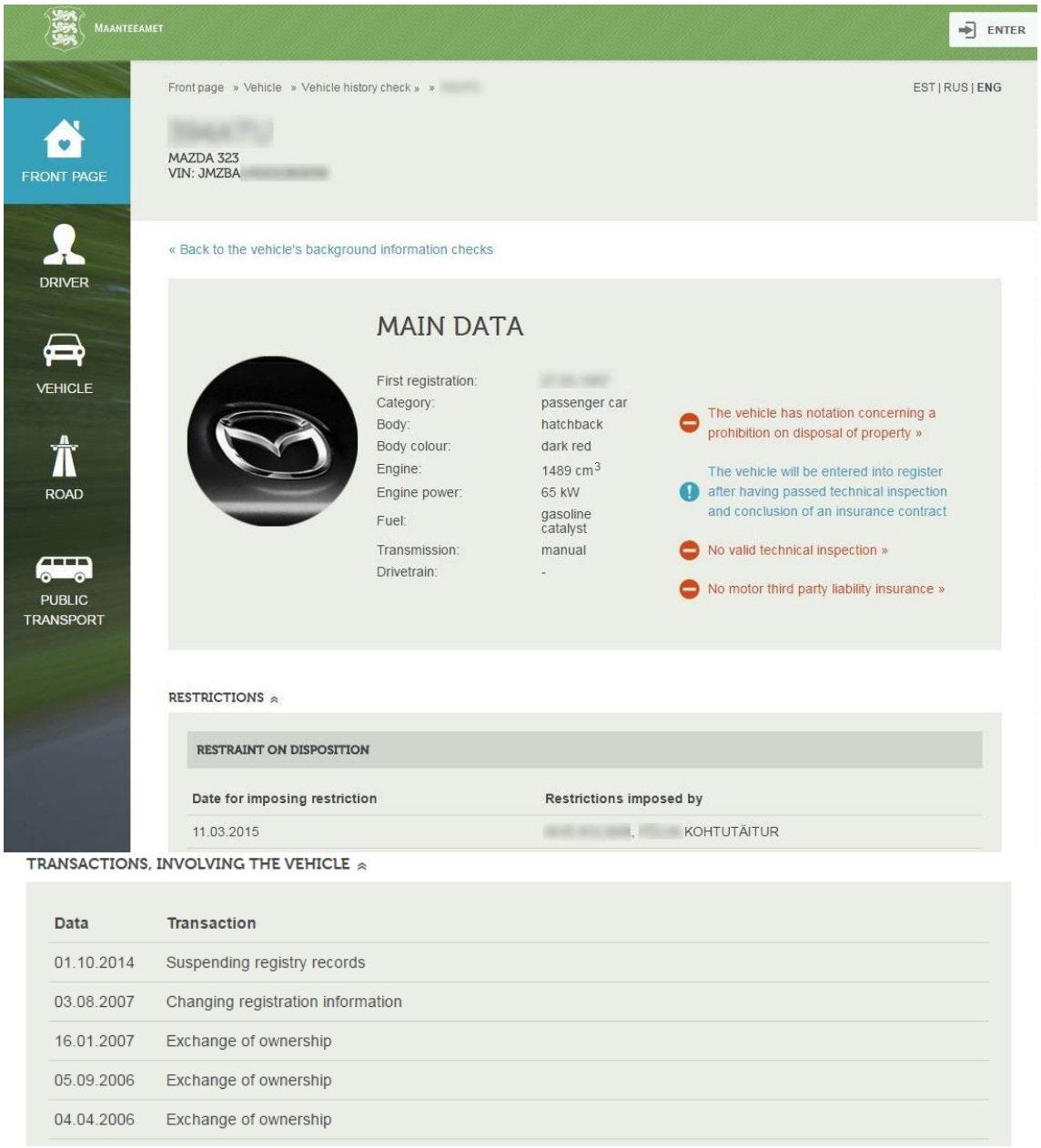


Figure 7. Vehicle information in traffic register.

Query via E-service portal confirms that the owner of that vehicle Mazada 323 is the data subject as there are no exchange transactions made after the January 13, 2010 which was the date of last integrity entry in marital portal register. Restrictions imposed by the officer of justice is made by Jane Doe who is located in Võru. It could be a hint that as data subject was born possibly in hospital of Põlva he lived or lives in the area. Taking into account that his wife is also from the same southern-region of Estonia, this information presented may be plausible. As it is clear from the *Figure 7* that vehicle (111AAA) has no insurance it may be useful to investigate further information from the

Estonian Traffic Insurance Fund. The Fund will retrieve data that this vehicle has been in 2 accidents where the latest took place in 12th of January 2008.

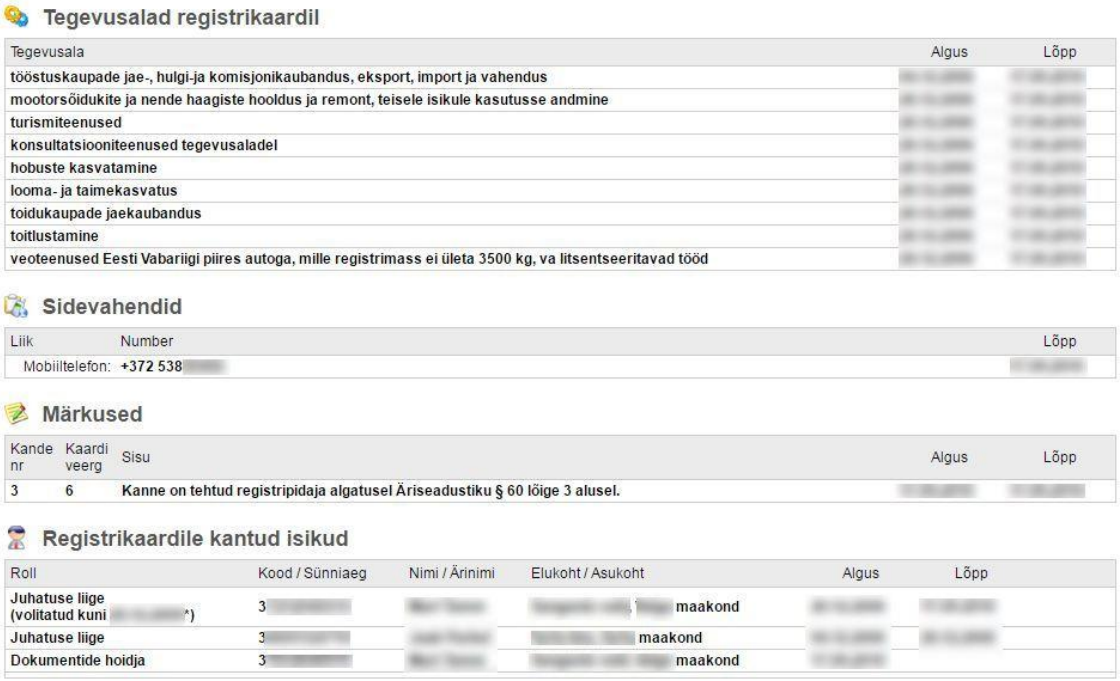
Applying the same scenario to the registered trailer (registration number 111AA) it can be seen that the owner of the trailer is also “John Doe”. Regarding vehicles Opel Commodore and Opel Frontera, the ownership information is false positive. According to the Estonian Road Administration Traffic and Vehicle Registry those vehicles are no longer registered.

Using the driver’s license query from E-service portal of Estonian Road Administration, it returns information that his driver’s license is valid until 03.11.2020.

Although information regarding the vehicles is not considered personal data, it is important to understand that it could be used to gather background information to conduct a social engineering attack.

4.2.5. Ministry of Justice

Commercial register will give additional valuable information such as the potential phone number of the subject of interest, company’s field of actions and business partners (see *Figure 8*). This contains enough information to investigate further. Making a search via Number query register [61] it is seen that the number is in use and is operated by the Telia Eesti AS. According to the commercial register managed by Centre of Registers and Information Systems the company OÜ Johncompany (registry code 11111111) was founded in December 12, 2003 (birthday of subject of interest) and shut down on September 28, 2009.



Tegevusalad registrikaardil

Tegevusala	Algus	Lõpp
tööstuskaupade jae-, hulgi-ja komisjonikaubandus, eksport, import ja vahendus		
mootorsõidukite ja nende haagiste hooldus ja remont, teisele isikule kasutusse andmine		
turismiteenused		
konsultatsiooniteenused tegevusaladel		
hobuste kasvatamine		
looma- ja taimekasvatus		
toidukaupade jaekaubandus		
toitlustamine		
veoteenused Eesti Vabariigi piires autoga, mille registrimass ei ületa 3500 kg, va litsentseeritavad tööd		

Sidevahendid

Liik	Number	Lõpp
Mobiiltelefon:	+372 538	

Märkused

Kande nr	Kaardi veerg	Sisu	Algus	Lõpp
3	6	Kanne on tehtud registripidaja algatusel Äriseadustiku § 60 lõige 3 alusel.		

Registrikaardile kantud isikud

Roll	Kood / Sünniaeg	Nimi / Ärinimi	Elukoht / Asukoht	Algus	Lõpp
Juhatuse liige (volitatud kuni *)	3		maakond		
Juhatuse liige	3		maakond		
Dokumentide hoidja	3		maakond		

Figure 8. Information from commercial register.

4.2.6. Ministry of the Environment

The Ministry of the Environment is responsible for the Land Board register from where it is possible to see the details related to the cadastre id where data subject's wife is the owner and where his former company was registered in (see *Figure 9*). Information about the size and registry time of the property is included but no additional valuable or personal data related information can be found.

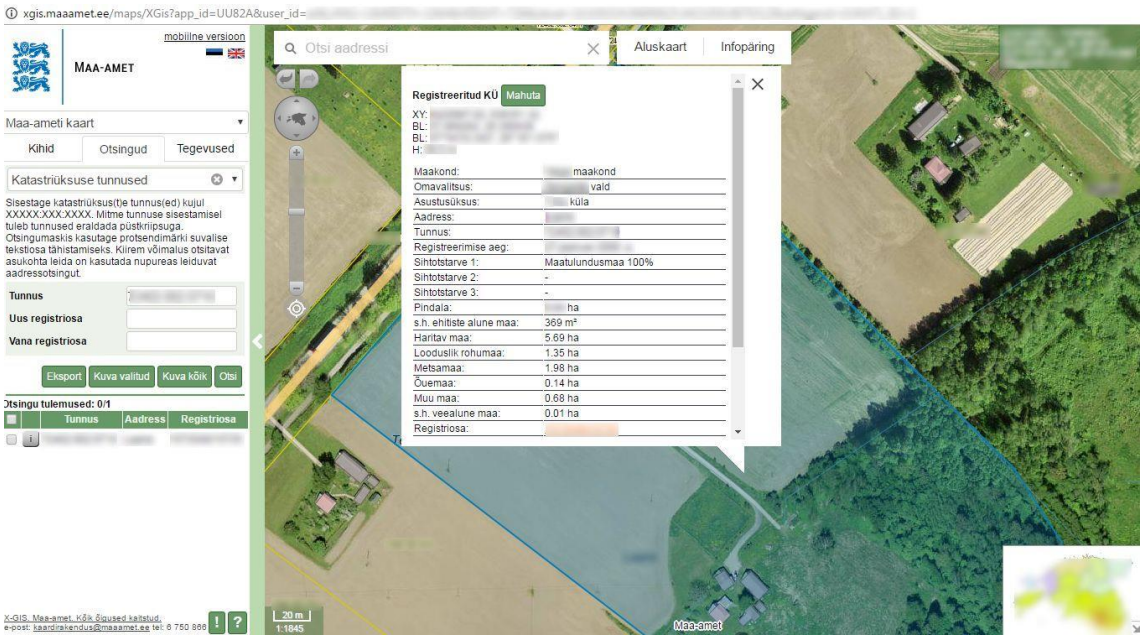


Figure 9. Property details from the land board register.

4.2.7. Ministry of Education and Research

No valuable information could be found despite of quite a few interesting registries. In case of different subject of interest, these registries may have return data that is worth further investigation.

4.2.8. Ministry of Finance

The Ministry of Finance is the government's expert in the implementation of tax, financial and fiscal policies, and setting economic goals [62]. Similarly, to Ministry of Education and Research, Ministry of Finance has valuable databases that in this scenario did not returned any valuable information.

4.2.9. Ministry of Rural Affairs

The government area of the Ministry of Rural Affairs involves the planning and implementation of rural policy, agricultural policy, fishing industry of the fisheries policy and the trade policy of agricultural products, the organisation of ensuring food safety and compliance, the coordination of the activities related to animal health and protection and plant health and protection, the organisation of agricultural research and development and agricultural education and the preparation of corresponding draft legislation [63]. In terms of this thesis no additional valuable information was found.

4.2.10. Ministry of the Interior

Minister of the Interior is mainly responsible for Estonian Police and Boarded Guard administration, Fire Department and Estonian Internal Security Service. Estonian document registry managed by Estonian Police and Boarded Guard issuing information about the person's id card and passport was found. Additional information such as confirmation about the Estonian citizenship was returned by the right of residence query.

4.2.11. Minister of Foreign Affairs

The Minister of Foreign Affairs did not return any valuable information or databases to search for information. Although minister has various datasets in the opendata.riik.ee portal, none gave additional connection points to the subject of interest.

4.2.12. Summary of Estonian government data about citizens

To understand what kind of information Estonian government is publishing, the summary of legal available attributes is analysed (see *Table 3*).

Table 3. Data attributes found in the Estonian registers publishing personal data.

Register name	**Publishes personal data**	**Available data attributes**
Marital property	Yes	Person´s name, surname, identification code, vehicle make, vehicle model, vehicle registration number, vehicle identification number (VIN)
Official publication	Yes	Person's name, surname, identification code
Motor register	No	Driving license validation date, vehicle technical information, transaction involving vehicles, restrictions
Commercial register	Yes	Legal person's name, surname, identification code, contact address
Land cadastre	No	Land cadastre id, area size
Ship register	Yes	Person's name, surname, identification code, ship name, ship number
Artistic associations	Yes	Person's name, surname, identification code

According to that information, a data subject called “John Doe” was chosen and government websites and registers were manually scanned and information gathered in order to see what kind of data is available to collect. As it reveals it is possible to find quite a bit of information if the scenario is targeted precisely enough (see *Table 4*).

Table 4. Results gathered from the Estonian government registers.

Attribute	**Value**	**Result**	**Source**
Subject of interest name	John Doe	Accurate	Popular name in Estonia chosen by the author of the thesis
Identification code	31111111111	Accurate	Marital property register
Place of birth	Võru or Põlva hospital	Plausible	Structure of identification code
Date of birth	11th of November 1911	Accurate	Structure of identification code
Subject of interest wife´s name	Jane Doe	Accurate	Marital property register
Subject of interest wife´s identification code	41111111111	Accurate	Marital property register
Owner of the vehicles	Mazda 323 (111AAA, VIN: JMZBA1111111111), Infi trailer (111AA, VIN: 1111111)	Accurate	Estonian Road Administration e-service portal
Accidents with vehicles	4 times	Accurate	Fund of traffic insurance
Contact address	Võru, Hummuli parish, Vidriku village, Püssi	Plausible	Marital property register
Owned companies	Johncompany OÜ (registration code: 11111111)	Accurate	Official publication portal
Company founded / deleted	December 12, 2006 / September 28, 2009	Accurate	Commercial register
Reason for company liquidation	not reporting the yearly financial records	Accurate	Official publication portal, Commercial register, Marital property register
Driving license	Until 03.11.2020	Accurate	Estonian Road

Attribute	Value	Result	Source
			Administration e-service portal
Phone number	+372 53811111 (Telia AS)	Plausible	Commercial register and Number query register
Documents issued date	ID card: 23/12/2009 Passport: 03/06/2011	Accurate	Document registry

As the manual data gathering scenario returned quite a bit of significant data, the method itself is quite slow in a situation where it is necessary to gather information about numerous people. In that case OSINT is a useful tool to take advantage of, as it is automated and focused to find open data according to particular scenario. Additionally, to the great features of the tool to collect relevant information, it may also be used to defend information that is published about the data subject. Conformably to the manually generated path, a relatively similar automated scenario will be built in order to automate the search. The author of this thesis took the properties of each register into consideration and excluded some of them while building the scenario for OSINT tool. It is reasonable to start the scenario with the source that returns the most data.

4.3. Building an open source intelligence (OSINT) tool

In this paragraph an OSINT tool building process is described that is able to find personal data through automated queries. The tool uses only government registers to find information. It is important to note that the author of this thesis has a consent from the government institutions to conduct automated queries on registers described in the scenario.

The main goal of built OSINT tool is to gather personal data from Estonian government registries with automated scenario. Registers used in the scenario by the OSINT tool are Marital Property Register, Official publication portal, Commercial register, E-ship register, Political Parties register, E-land register and Artistic association database. Estonian Road Administration E-service portal and Land board register that were used in manual scenario are excluded from automated scenario as they do not return personal

data. E-land register and Political parties register are included to the automated scenario to increase the possibility of registers returning personal data.

In order to achieve a goal to gather personal data, a data scraping (often referred to as web-scraping) technique is used. Data scraping is a technique in which a computer program extracts data from human-readable output coming from another program [64]. The challenge of using this technique is that websites often display poorly structured data. This means that no straightforward form will be found to teach the structure to the program. To achieve the goal of a built OSINT tool the data scraping tool Parsehub [65] is used. Parsehub is a Firefox browser extension that supports complex sites with JavaScript use. The tool is able to determine relations among elements on the webpage. It is important that the tool is relatively easy and free to use with a descent support, able to extract data in machine-readable format (JSON, CSV), supports regular expressions and has an API capability. Additional web scraping tool Data Miner was also considered as a candidate to extract data from websites, but was discarded as the tool does not have any API support to execute the scenario script. Additionally, to web scraping tool a web HTML and PHP scripts were used in order to fulfil the goal. Each Parsehub project created has an API returning CSV or JSON format outputs. One Parsehub project can contain different templates to execute various tasks (i.e. login, CAPTCHA resolving, navigating through sites, structuring and extracting the data).

According to the scenario described in sub-paragraph 4.2.12 Summary of Estonian government data about citizens a data scraper for Marital Property register [66] was needed. A project named “1_Marital_Property” with 3 templates was created. First template was created to enter the search term into the input field and trigger the submit button. The initial search term used is data subjects first name and last name in all Parsehub projects. If a query returns more than one record, only the first record is used. This rule is applied to avoid cases where multiple results for different subject of interest are returned. Second template was created to extract the data with people’s names, identification codes and to open the registry card link. Regular expression was used for extracting names and identification codes and other relevant information. Third template was created to extract necessary valuable information from registry card. In this template a regular expression is used in order to identify strings like vehicle licence plate, vehicle identification number, business registry code and property number. The

result is displayed in JSON format in Figure 10. As it can be seen, it is possible to extract personal data (identification code) about the data subject and his wife using regular expression.


```
{
 "var_input": [
  {
   "searchterm": "John Doe",
   "var_isikukood": [
    "31111111111",
    "41111111111"
   ],
   "var_name": [
    "John Doe",
    "Jane Doe"
   ],
   "var_entry_num": "1",
   "var_entry_status": "Kehtiv",
   "var_entry_content": "Abikaasade poolt valitud varasuhteks on varalahususe varasuhe. 1. Kinnistu nr 1111111 on Jane Doe
lahusvaraks koos kõigi võlakohustustega Nordea Bank Finland Plc ees; 2. OÜ JohnCompany, registrikood 11111111 jääb John Doe
lahusvaraks; 3. Sõiduauto MAZDA, tehasetähisega JMZBA111111111111, riikliku registreerimismärgiga 111AAA, jääb John Doe
lahusvaraks; 4. INFI TRAILERS, tehasetähisega 8811111, riikliku registreerimismärgiga 111AA jääb John Doe lahusvaraks; 5.
Sõiduauto OPEL COMMODORE, tehasetähisega 1111111111, riikliku registreerimismärgiga 711AAA jääb John Doe Iahusvaraks; 6.
Sõiduauto OPEL FRONTERA, tehasetähisega SED5JMWL4PV111111, riikliku registreerimismärgiga 311AAA jääb John Doe lahusvaraks. 7.
IGASUGUNE abikaasade poolt abielu kestel, pärast käesoleva lepingu sõlmimist, tulevikus soetatav vara, on abikaasade LAHUSVARAKS
ja on selle abikaasa omandiks, kelle nimele vara soetatakse.",
   "var_veh_plate": [
    "111AAA",
    "111AA",
    "311AAA"
   ],
   "var_veh_vin": [
    "JMZBA111111111111",
    "SED5JMWL4PV111111"
   ],
   "var_regcode": [
    "11111111"
   ],
   "var_property_num": [
    "1111111"
   ]
  }
 ]
}
```


Figure 10. Extracted data from marital property register in JSON format.

After creating the first project about extracting data from Marital Property Register, a second project called 2_CommercialRegister was created. This project returns data from commercial register [67]. This query returns only information about sole proprietorship. This project consists of 5 templates. First template was created to enter the search term into the input field and trigger the submit button. The second template gathers information about the juridical entity name, code, time of first entry, status, address. The third template extracts information about the possible tax arrears the company may have and information about the juridical form of the company. The fourth template is responsible for resolving the CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) in order to see information in registry card. The fifth template extracts data about juridical entities email address, the identification codes of the board members. The result is displayed in JSON format in *Figure 11*.

```
{
   "var_input":[
      {
         "searchterm":"[redacted]",
         "Arireg_list":[
            {
               "arireg_name":"[redacted]",
               "arireg_code":"[redacted]",
               "arireg_entry":"[redacted]",
               "arireg_status":"Entered into the register",
               "arireg_address1":"Harju maakond, Tallinn, Pirita linnaosa, Kaarli pst 1a-20, 12543"
            }
         ],
         "Arireg_details":[
            {
               "var_arireg_debt":"The operator Jane Company (81111111) has no tax debt as at 09.04.2018.",
               "var_arireg_type2":"Private limited company",
               "var_arireg_address2":"Harju maakond, Tallinn, Pirita linnaosa, Kaarli pst 1a-20, 12543"
            }
         ],
         "captcha":"22593",
         "Arireg_Micro":[
            {
               "var_arireg_address":"Aadress on Harju maakond, Tallinn, Pirita linnaosa, Kaarli pst 1a-20, 12543"
            }
         ],
         "var_arireg_idcode":[
            "41111111111",
            "32222222222",
            "33333333333"
         ],
         "var_arireg_email":[
            "jane@jcompany.ee"
         ]
      }
   ]
}
```


Figure 11. Extracted data from commercial register in JSON format.

The third project is named 3_Eship that consists of 3 templates. This project returns data from e-ship register [68]. First template was created to enter the search term into the input field and trigger the submit button. The second template gathers data about the identification code of the ship owner, ship type, ship name and ship registry number. In this template a name check between search term and extracted name is done. This is needed because the E-ship register returns information with "%search term%" principle. The third template is extracting different additional information for background data (co-owners, home port, length, width of the ship etc). This template does not gather data that will be used in further projects as an input parameter.

The fourth project created is named 4_Artitic_Associations and consists of 3 templates. This project gathers information from database of artistic associations [69]. The first template is created to insert the search term into specific fields and trigger the search button. The second template is checking the queried name against the returned name list as the database of artistic association returns information "%search term%" principle. In case there is a positive match the second template continues the data gathering which is done in the third template. Result of the third template is shown in *Figure 12*.

.


```
{
   "keywords":[
      "Jane Doe"
   ],
   "var_artistic_input":[
      {
         "searchterm":"Jane Doe",
         "actual_artistic_name":[
            {
               "var_artistic_name":"Jane Doe",
               "var_firstname":"Jane",
               "var_lastname":"Doe",
               "var_date_of_birth":"01.08.1978",
               "var_association":"NÄITLEJATE LIIT",
               "var_time_of_join":"03.03.1998"
            }
         ]
      }
   ]
}
```


Figure 12. Extracted data from database of artistic associations in JSON format.

The fifth project is called 5_Parties and consists of 3 templates. This project extracts information from political parties database [70]. This is a database where all members of any sort of political parties are listed (name and date of birth). The first template is responsible for entering the search term into the input field and submit the query. The second template is similar to project 5_Artitic_Associations second template where the search term is compared to the data resulted and decided whether there is a positive match or not. In case of positive match, the project will continue with the third template where data extraction is done.

The sixth project is called 6_LandRegister that consists of 3 templates. This project extracts data from Land Register [71] where information about cadastral data, owners and restrictions are stored. The first template is responsible for inputting the data subject name and executing the query. The second template is responsible for resolving the CAPTCHA and the third template will collect the available data (address, cadastre number, property type) about the data subject.

In previously described scenarios, each project represents one Estonian government portal where the data is being extracted. Using Parsehub for defining the data extraction scenario, it will be possible to use an API for each project to extract data. This means

that combining different API-s it is possible to automate seemingly one query into multiple queries and gather information over the Estonian government registers and portals and present the data in one landing page. To achieve described functionality a webpage was built with PHP and HTML to call out the Parsehub project API-s according to the input defined by a webpage user (see *Figure 13*).

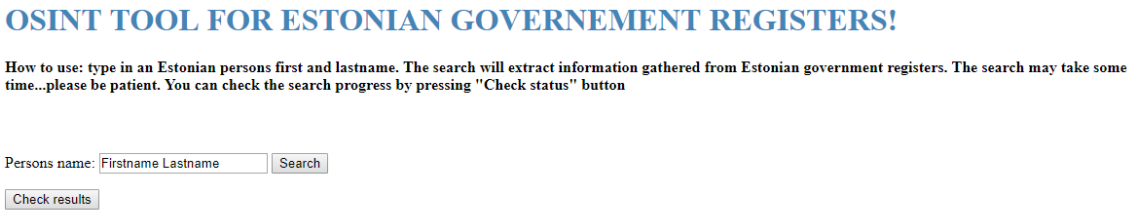


Figure 13. OSINT tool for Estonian government registers.

The webpage consists of 3 PHP files where first one calls out the API with right parameters. Second file stores data into the database and is called out by a Parsehub webhook. The webhook is triggered every time a Parsehub status event changes (i.e. initialized, queued, progress, complete) The third file calls out Parsehub function to view extracted data and present it to the user. As Parsehub API content-encoding is in gzip format the third files is responsible for decoding the results. The results are presented in *Figure 14*.

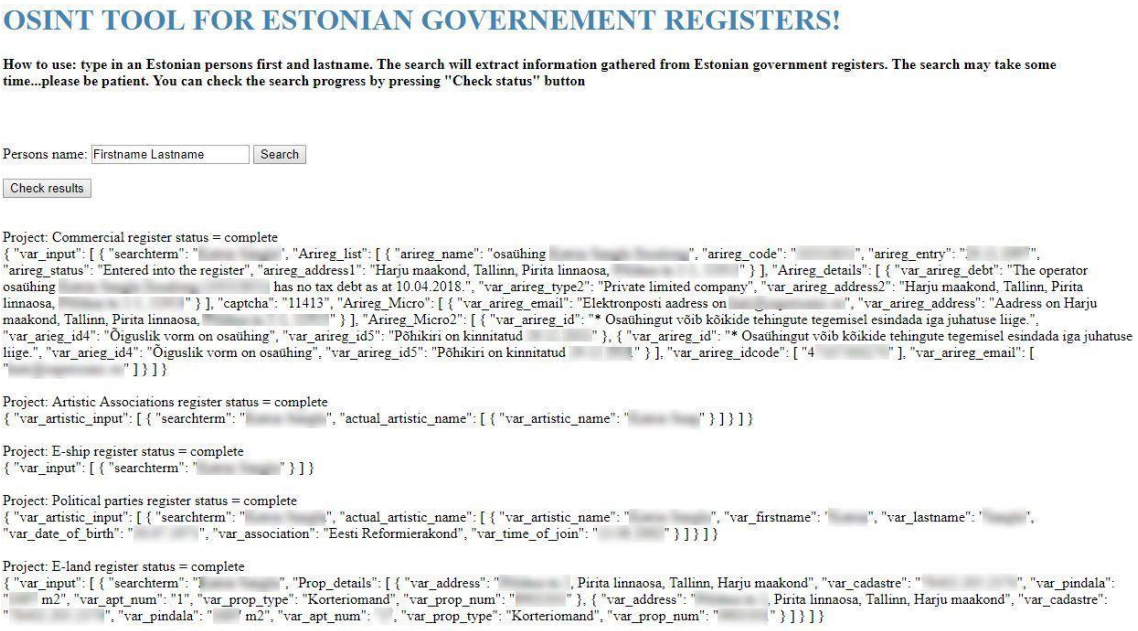


Figure 14. The example of OSINT tool output.

On March 1, 2018 the Centre of Registers and Information Systems applied an authorization functionality (Estonian ID card and Mobile ID) to the Marital Property register. Most likely the safeguard was applied due to a draft of data protection implementation act [37], which have to be implemented no later than on May 25, 2018. Due to applied safeguard the author of this thesis cannot use Parsehub API on the OSINT tool, but the Parsehub project can still be executed and validated in a debug mode.

4.4. Validation of the OSINT tool

In this paragraph a validation process and the results for built OSINT tool is described. Based on the management research textbook [72], the research uses a positivists approach within a realist's ontology meaning the author of this thesis gathers facts through experiments, analyses the facts and finishes by representing the results.

As the main purpose of GDPR is to protect data subject's personal data, the validation method concentrates on finding out the percentage (success rate) of cases the personal data is returned by OSINT tool that was built. Bearing in mind that validation of the OSINT tool will generate some load onto the registers, the agreed number of queries with the government institutions is 300 per register. This allows to validate the data gathering method success rate and analyse the positive results returned by the OSINT tool. The author initially wanted to use the input data from population registry (first name, last name and older than 17 years), but unfortunately the request was tackled into complicated procedures. To avoid complexity and save time, the author used Facebook users first and last name (born in Estonia) as a primary source of initial search term. It is considered a positive result when the OSINT tool will match the definition of personal data from the list described by Estonian Data Protection Inspectorate [73]:

1. any data concerning an identified or identifiable natural person;
2. data revealing political opinions or religious or philosophical beliefs;
3. data revealing ethnic or racial origin;
4. data on the state of health or disability;
5. data on genetic information;

6. biometric data;
7. information on sex life;
8. information on trade union membership;
9. information concerning commission of an offence or falling victim to an offence before a public court hearing, or making of a decision in the matter of the offence or termination of the court proceeding in the matter.

If the tool finds from different registers at least two different data subjects with the same input (i.e. 2 persons with the same name), the result is considered positive (one positive result per input). This serves the purpose of this tool to find personal data in any case possible as long as the search term matches the data subject. In case there is no match from the mentioned list, the result is considered negative. False positives are cases when register(s) return personal data with the similar or different data subjects name that is queried (i.e. input is "Külli Paat" and output is "Külliki Paat").

Every Parsehub project was executed separately with the Parsehub client and the results were saved for further analyse. The results for 300 test cases per register are described in the *Table 5* where the most number of positive results per register was in e-land register and political parties registry followed by the marital property registry, commercial registry, artistic associations registry and e-ship registry.

Table 5. OSINT validation results by registries.

Register name	**Number of positive results**	**Number of false positives**	**Percentage of positive results**
Marital property	31	0	10.3
Commercial register	17	4	5.6
E-Ship register	1	0	0.3
Artistic associations	21	0	7
Political parties	90	0	30
E-Land register	206	0	68.6

It must be pointed out that in Marital property all and Commercial registries some of the queries resulted in more than one name and identification code combination. For Marital property register whenever there was a positive result the subject of interest personal data and his/her wife's/husband's personal data were retrieved. This means that the number of identification codes returned is 62 when taking into account only the number of personal data OSINT tool returns. The same applies to the Commercial registries where the additional personal data was returned (6 results) about subjects that were not in the interest of the scope (i.e. more than one board member). According to the validation method when the input parameter does not match with the outputted data, the subject name is considered a false positive. Therefore, described cases are not merged into the overall success rate calculations.

In some situations, different registries returned a positive result for the same data subject. Within this validation process 104 cases occurred (see *Table 6*) where personal data from different registries returned a positive result for the same data subject.

Table 6. Number of cases OSINT tool returns positive result.

Number of registers return positive result for the same data subject	**Number of cases occurred**
6 registers	0
5 registers	1
4 registers	4
3 registers	20
2 registers	79
1 register	127

To find out the percentage (success rate) of the built OSINT tool (see equation 1) the total number of cases occurred (231) from *Table 6* needs to be multiplied by 100 and divided with number of total test cases (300).

$$Success\ rate = \frac{(Number\ of\ cases\ returned\ positvie\ result * 100)}{Total\ number\ of\ test\ cases} = 77\% \qquad (1)$$

The success rate of built OSINT tool method to find personal data from Estonian government registers is 77 %. This success rate shows that it is possible for most of the cases to collect personal data and build some type of social engineering vector to manipulate the potential victim. By this OSINT tool total 6 registers were included and with every registry added through this method can increase the success rate.

4.5. Results and implications

Various data gathering scenarios have a different purpose. In previous paragraphs the goal was to prove that it is possible to build and OSINT tool that is able to extract data from different Estonian government registers using automated scenarios. To achieve this goal a manual scenario was built that is able to extract personal data. This scenario showed a pathway to automated scenarios and building the OSINT tool. The built OSINT tool is able to extract personal data (see *Table 7*) through various data attributes.

Table 7. OSINT tool finding different data attributes.

OSINT tool scenario	**Personal data attribute**
1_Marital_Property	Person´s name, surname, identification code, vehicle make, vehicle model, vehicle registration number, vehicle identification number (VIN), business registry code, property number.
2_CommercialRegister	Identification code, business name, business registry code, business address, business debt records, business email address
3_Eship	Identification code, ship name, ship registration number, ship owners name
4_Artitic_Associations	Persons name, date of birth, artistic association, time of joining the association.
5_Parties	Persons name, date of birth, political parties, time of joining the parties.
6_LandRegister	Address, cadastre number,

As it turns out from previous paragraphs, it is possible to find personal data in Estonian government register. Although the location of the tool is not revealed as the author of this thesis does not have right to re-publish personal data, it is possible to get the idea of

the tool through the *Figure 13* and *Figure 14*. In *Table 8* OSINT tool result about a single data subject is presented.

Table 8. OSINT tool result about a single data subject.

Attribute	**Value**
Subject of interest name	Jane Doe
Identification code	41111111111
Address	Harju maakond, Tallinn, Pirita linnaosa, Kaarli pst 1a-20, 12543
Cadastre number	78402:205:2316
Date of birth	11th of November 1911
E-mail	Jane@jcompany.ee
Business	Jane Company OÜ
Business associates identification codes	32222222222, 33333333333
Business registration number	81111111
Political Parties	Eesti Reformierakond

In addition, this OSINT tool will present unstructured data as a structured data. This opens up a different possibility for automated queries. This is most likely alarming while combining different registers data sets, a great number of intel information will be gathered. Although in legal terms all those registers have right to publish personal data, combining this information is a powerful tool in a context of data protection and GDPR. In situations where every registry owner acts within the borders of their own registry, it is easy not to realize the data protection problem as a whole. Combining datasets could lead to wide scale data leakage incidents and profiling data subjects. Demonstrating the possibilities through OSINT tool or some other method is one way to increase the awareness of the registry owners.

This also shows that it is time to revise the current domestic acts regulating the government registers and only publish personal data when it is crucial. While publishing personal data, new safeguards are needed to prevent different automated queries to extract the content of the registers. As seen from previous paragraph, CAPTCHA is not

sufficient method to protect against modern web scraping tools. Estonian domestic law needs to be supplemented by mandatory requirements for public registers and services. Authorisation methods like Estonian id-card or mobile-id should be minimum requirements to increase the level of security.

One alternative usage for the built OSINT tool is validating how well Estonian governments are applying safeguards in order to follow the purpose of GDPR. When no modifications are made to the tool itself and it has similar success rate in the future finding personal data, it is possible to say that no extra safeguards were implemented by the Estonian government. If the success rate is significantly lower it is possible to say that actions has been taken towards protecting Estonian citizens personal data. The same method described in the subsection 4.3 can be applied in any country with electronic registers to understand the actual improvements the registry owners are making. The tool itself needs modifications as it does not work out of the box due to the peculiarities and properties of various registers, but when applying the same method this option for alternative usage is available.

Building an OSINT tool can be beneficial for gathering various intelligence information about someone, but there is an ethical side for this as well. OSINT tools may gather publicly available information in a way that magnifies the significance of data. The author of this thesis does not publish the location of the OSINT tool, partly because this is not the purpose of building the tool and partly due to ethical reasons as this is not appropriate. It is important and challengeable to keep in mind that building an OSINT tool may be sensitive subject and it is fair to look this process from the data subject´s point of view. M. Jakobsson and J. Ratkiewicz study about ethical phishing experiments [74] states that “ethical experiment must not expose the participants to any risk“. The same principle can be adapted while describing the OSINT building scenario, as it is important not to reveal too much information about the data subject. While writing this thesis, it was a challenge to tamper the collected data in such way that the data subject remains anonymous, but still publish enough information to make it observable and understandable to the reader. One has to keep in mind the fact that even publicly available data can still be sensitive and thereof harm the data subject or the ones building the tool or making the query.

While writing this thesis, permissions from registry owners were needed to execute the automated queries. Applying the same method described in subsection 4.3 an OSINT tool can be built to gather information about a large number of people without asking for anyone's permission. This is possible, because not enough effective safeguards are implemented. For the data subject's perspective this kind of situation can be alarming as someone can theoretically hold information with the "victim's" identification code, location, phone number, family members and relatives etc. This can lead to different forms of manipulation of the data subject. On the other hand, it can be useful for the potential "victim" to try some of the OSINT tools available to make him/her realize the effects of publishing data by the government institutions, commercial organizations or by the data subject himself/herself. Realizing the significance of the published data, the data subject has an opportunity to avoid publishing seemingly meaningless information in the future and thanks to the GDPR data erasure principle some information even from the past.

5. Summary

Personal data is shared worldwide constantly and on a daily basis. Whether it is data subject himself/herself that publishes it, organizations or public institutions, there have to be safeguards in place in order to protect the data. The baseline of protecting personal data is to know what to protect and how this information can be accessed and used against the data subject. The first part of the thesis concentrates on GDPR and Estonian legislative acts in order to understand what to protect and present the challenges Estonian government registrants are facing. The second part concentrates on how to find personal data and combine the different datasets to build an OSINT tool about data subject.

There are great number of written papers about GDPR and different OSINT tools. While papers about the effects of GDPR on Estonia have been written, this thesis focuses on the challenges the GDPR and Estonian legislative acts will bring to Estonian government registrants. At the same time while there are papers written about web scraping of personal data, most of them are based on social media platforms. This thesis focuses on extracting personal data from Estonian government websites.

General overview about GDPR and Estonian legislation regarding electronical registries was given. In the overview Estonian public registries that are able to search and output data about the data subject were covered. After the overview an analyse was done regarding the challenges the registrants may face while implementing the GDPR into practise. The challenges were categorized as juridical, organizational and technical. It was concluded that one of the big juridical challenge is the GDPR principle about the territorial scope, where non-EU members operating and processing personal data in EU member states need to be compliant with the GDPR as well. At the organizational level, it is needed to take into account the increasing administrative workload, training of the employees and appointing the independent data protection officer (DPO). The GDPR principle about the storage limitation is one technical challenge that is needed to be considered as it may become costly and complicated to implement. Additionally, GDPR is seen as an ongoing process and because of those general suggestions about adapting the GDPR were given.

According to GDPR article 86 [35], the publication of personal data can be allowed by a member state law. *Table 1* demonstrates the registries acting under the GDPR article 86. Mentioned registers were used in order to build a scenario to manually collect data about the data subject. The attributes gathered about the data subject are presented in *Table 3*. The scenario was somewhat altered to perform automated queries via OSINT tool that was built while writing this thesis. The purpose of the OSINT tool was to find personal data from Estonian government registers. To the automated data-gathering scenario, 6 registers were included and a OSINT tool webpage was built. The input parameter was name of the data subject and the output were the results of extracted registers data attributes. According to the validation process (see section 4.4) the tool collected 77% of cases personal data about the data subject.

When applying the GDPR, the main focus must remain to the data subject. Various challenges and threats must be evaluated and appropriate safeguards implemented. People must realize what are the impacts of their taken and non-taken actions in order to protect personal data. During the process of writing this thesis, the Centre of Registers and Information Systems applied an online authorization functionality (Estonian id card and mobile id) to the Marital Property register that resulted in excluding the register from OSINT tool. Relaying on a fact that Estonia is ranked number one country on cybersecurity in Europe in 2017 [48] the appropriate safeguards must be implemented to keep that position while increasing the awareness level of personal data security.

References

[1] Vcloudnews, "Every day big data statistics - 2.5 quintillion bytes of data creadted daily," Vcloudnews, [Online]. Available: http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily. [Accessed 01 02 2018].

[2] IBM, "Bringing big data to the Enterprise," IBM, [Online]. Available: https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html. [Accessed 01 02 2018].

[3] S. Sillaots, “Harmonizing personal data protection regulation in the General Data Protection,” 2014.

[4] M. Bazzell, “IntelTechniques,” IntelTechniques, [Online]. Available: https://inteltechniques.com/menu.html. [Accessed 13 02 2018].

[5] Paterva PTY Ltd, “Paterva,” [Online]. Available: https://www.paterva.com. [Accessed 28 03 2018].

[6] O. Castrillo-Fernández, “Web Scraping: Applications and Tools,” 2015.

[7] J. K. Mikli, “Sotsiaalmeediast isikuandmete masskogumise meetodid,” 2016.

[8] The European Parliament and the Council Of The European Union, "Access to European Union Law," Eur-Lex, [Online]. Available: http://eur-lex.europa.eu/legal-content/EN. [Accessed 05 02 2018].

[9] The Ministry of Justice, "Riigi Teataja," Riigi Teataja, [Online]. Available: https://www.riigiteataja.ee/en/. [Accessed 05 02 2018].

[10] The European Parliament and the Council Of The European Union, "EUR-Lex - National law applicable," Publications Office, [Online]. Available: http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A31995L0046. [Accessed 09 02 2018].

[11] The European Parliament and the Council Of The European Union, "Subject-matter and objectives," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-1-gdpr/. [Accessed 05 02 2018].

[12] The European Parliament and the Council Of The European Union, "Definitions," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-4-gdpr/. [Accessed 05 02 2018].

[13] Statistics Estonia, “Populaarsed eesnimed,” Statistics Estonia, [Online]. Available: https://www.stat.ee/public/apps/nimed/TOP. [Accessed 22 04 2018].

[14] Statistic Estonia, “Levinumad perenimed,” Statistic Estonia, [Online]. Available: https://www.stat.ee/public/apps/nimed/pere/TOP. [Accessed 22 04 2018].

[15] Wikipedia, “Religion in Estonia,” Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Religion_in_Estonia. [Accessed 22 04 2018].

[16] The European Parliament and the Council Of The European Union, "Lawfulness of processing," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-6-gdpr/. [Accessed 05 02 2018].

[17] The European Parliament and the Council Of The European Union, “Official Journal of the European Union,” EUR-Lex, [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1517820519816&uri=CELEX:32016R0679. [Accessed 06 02 2018].

[18] The European Parliament and the Council Of The European Union, "Processor," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-28-gdpr/. [Accessed 06 02 2018].

[19] The European Parliament and the Council Of The European Union, "Security of processing," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-32-gdpr/. [Accessed 06 02 2018].

[20] The European Parliament and the Council Of The European Union, "Records of processing activities," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-30-gdpr/. [Accessed 06 02 2018].

[21] The European Parliament and the Council Of The European Union, "Codes of conduct," Intersoft Consultin, [Online]. Available: https://gdpr-info.eu/art-40-gdpr/. [Accessed 06 02 2018].

[22] The European Parliament and the Council Of The European Union, "Certification," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-42-gdpr/. [Accessed 06 02 2018].

[23] The European Parliament and the Council Of The European Union, "Notification of a personal data breach to the supervisory authority," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-33-gdpr/. [Accessed 06 02 2018].

[24] The European Parliament and the Council Of The European Union, "Communication of a personal data breach to the data subject," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-34-gdpr/. [Accessed 06 02 2018].

[25] Parliament of Estonia, “Estonian Public Information Act. - RT I 2000, 92, 597; 04.07.2017, 1,” Riigiteataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/516102017007/consolide. [Accessed 16 02 2018].

[26] Parliament of Estonia, “Marital Property Register Act. - RT I 1995, 87, 1540; 09.05.2017, 1,” Riigi Teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/519062017004/consolide. [Accessed 06 02 2018].

[27] Ministry of Social Affairs, "The statutes of Ametlikud Teadaanded. - RT I, 01.04.2015, 9.," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/509122016001/consolide#. [Accessed 07 02 2018].

[28] Parliament of Estonia, "Traffic Act. - RT I 2010, 44, 261; 20.12.2017, 1," Riigi Teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/516022016004/consolide. [Accessed 07 02 2018].

[29] Parliament of Estonia, "Commercial Code Act. - RT I 1995, 26, 355; 17.11.2017, 2," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/519122017001/consolide. [Accessed 07 02 2018].

[30] Parliament of Estonia, "Land Cadastre Act. - RT I 1994, 74, 1324; 05.01.2018, 1," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/529012018004/consolide. [Accessed 07 02 2018].

[31] Parliament of Estonia, "Law of Maritime Property Act. - RT I 1998, 30, 409; 29.06.2014, 109," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/528032014003/consolide. [Accessed 08 02 2018].

[32] Parliament of Estonia, "Law of Ship Flag and Ship Registers Act. - RT I 1998, 23, 321; 09.05.2017, 1," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/en/eli/519062017011/consolide. [Accessed 08 02 2018].

[33] Ministry of Culture, "The statutes of Artistic Associations database. - RT I, 06.01.2015, 13; 06.01.2016, 3," Riigi teataja, [Online]. Available: https://www.riigiteataja.ee/akt/106012016023. [Accessed 08 02 2018].

[34] J. Seaman, “GDPR: The difference between Personally Identifiable Information (PII) and Personal Data,” LinkedIn, [Online]. Available: https://www.linkedin.com/pulse/gdprthe-difference-between-personally-identifiable-jim-seaman/. [Accessed 16 02 2018].

[35] The European Parliament and the Council Of The European Union, “Processing and public access to official documents,” Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-86-gdpr/. [Accessed 13 02 2018].

[36] United States Department of Labor, “Guidance on the Protection of Personal Identifiable Information,” United States Department of Labor, [Online]. Available: https://www.dol.gov/general/ppii. [Accessed 30 03 2018].

[37] Ministry of Justice, “Estonian Data Protection Inspectorate,” Estonian Data Protection Inspectorate, [Online]. Available: http://www.aki.ee/sites/www.aki.ee/files/elfinder/article_files/iks_rs_en.pdf. [Accessed 04 04 2018].

[38] The European Parliament and the Council Of The European Union, "Right to compensation and liability," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-82-gdpr/. [Accessed 09 02 2018].

[39] The European Parliament and the Council Of The European Union , “Designation of the data protection officer,” Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-37-gdpr/. [Accessed 04 03 2018].

[40] Information System Authority, “Information System Authority,” [Online]. Available: https://www.ria.ee/public/ISKE/naidisdokumendid/LISA1.15.Turvaintsidentide_kasitlemise_kord.doc. [Accessed 29 03 2018].

[41] The European Parliament and the Council Of The European Union, "Data protection by design and by default," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-25-gdpr/. [Accessed 09 02 2018].

[42] Bundesamt für Sicherheit in der Informationstechnik, "IT Grundschutz - Catalogues," [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/International/GSK_15_EL_EN_Draft.pdf?__blob=publicationFile&v=2. [Accessed 09 02 2018].

[43] Ministry of Economic Affairs and Communication, "The system of security measures for information systems. - RT I 2007, 71, 440; 2009, 6, 39," [Online]. Available: https://www.riigiteataja.ee/akt/12901110?leiaKehtiv. [Accessed 09 02 2018].

[44] The European Parliament and the Council Of The European Union, “Right to data portability,” Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-20-gdpr/. [Accessed 29 03 2018].

[45] Microsoft, “What is Azure Information Protection,” Microsoft, [Online]. Available: https://docs.microsoft.com/en-us/information-protection/understand-explore/what-is-information-protection. [Accessed 28 03 2018].

[46] Data Protection Inspectorate, “Kuidas valmistuda andmekaitse üldmääruseks,” Data Protection Inspectorate, [Online]. Available: http://www.aki.ee/et/soovitused_maaruseks_valmistumisel. [Accessed 04 03 2018].

[47] B. Schneier, “Why anonymous data sometimes isn´t,” Wired, [Online]. Available: https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/. [Accessed 12 02 2018].

[48] International Telecommunication Union, "Global Cybersecurity Index 2017," 2017.

[49] The European Parliament and the Council Of The European Union, "General conditions for imposing administrative fines," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/art-83-gdpr/. [Accessed 12 02 2018].

[50] The European Parliament and the Council Of The European Union, "Administrative fines in Denmark and Estonia," Intersoft Consulting, [Online]. Available: https://gdpr-info.eu/recitals/no-151/. [Accessed 12 02 2018].

[51] C. Hobbs, M. Moran and D. Salisbury, Open Source Intelligence in the Twenty-First Century, Palgrave Macmillan UK, 2014.

[52] Data Protection Inspectorate , "Kas avalikust registrist pärit andmeid võib internetis taasavalikustada?," Data Protection Inspectorate , [Online]. Available: http://www.aki.ee/et/kas-avalikust-registrist-parit-andmeid-voib-internetis-taasavalikustada. [Accessed 22 04 2018].

[53] Supreme Court of Estonia, "Lahendid," Supreme Court of Estonia, [Online]. Available: https://www.riigikohus.ee/lahendid?asjaNr=3-3-1-3-12. [Accessed 22 04 2018].

[54] Supreme Court of Estonia, "Lahendid," Supreme Court of Estonia, [Online]. Available: https://www.riigikohus.ee/lahendid?asjaNr=3-2-1-159-14. [Accessed 22 04 2018].

[55] Open Data Portal of Estonia, "Open Data Portal of Estonia," Open Data Portal of Estonia, [Online]. Available: https://opendata.riik.ee/en. [Accessed 22 04 2018].

[56] Ministry of Culture, "Eesti spordiregister," Foundation of Sports Education and Information, [Online]. Available: https://www.spordiregister.ee/en. [Accessed 15 02 2018].

[57] Ministry of Culture, "Artistic Associations," Centre of Registers and Information Systems, [Online]. Available: https://ariregister.rik.ee/loomeliidud/liikmete_nimekiri?ll_id=8&lang=eng. [Accessed 14 02 2018].

[58] Ministry of Social Affairs, “Marital Property register,” Centre of Registers and Information Systems, [Online]. Available: https://abieluvararegister.rik.ee/. [Accessed 14 02 2018].

[59] “Isikukood,” Wikipedia, [Online]. Available: https://et.wikipedia.org/wiki/Isikukood. [Accessed 15 02 2018].

[60] The Ministry of Economic Affairs and Communications, “E-service of the Estonian Road Administration,” Estonian Road Administration, [Online]. Available: https://eteenindus.mnt.ee/public/soidukTaustakontroll.jsf. [Accessed 13 02 2018].

[61] Technical Inspection of Estonia, “NBA,” Centre of Registers and Information Systems, [Online]. Available: https://nba.tja.ee/numbriparing.aspx. [Accessed 15 02 2018].

[62] The Ministry of Finance, “Rahandusministeerium,” The Ministry of Finance, [Online]. Available: http://www.fin.ee/introduction. [Accessed 15 02 2018].

[63] The Ministry of Rural Affairs, “Overview and Structure,” The Ministry of Rural Affairs, [Online]. Available: https://www.agri.ee/en/ministry-contacts/overview-and-structure. [Accessed 15 02 2018].

[64] “Data Scraping,” Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Data_scraping. [Accessed 25 02 2018].

[65] “Parsehub,” [Online]. Available: https://www.parsehub.com. [Accessed 25 02 2018].

[66] Center of register and information Systems, “Marital Property register,” Center of register and information Systems, [Online]. Available: https://abieluvararegister.rik.ee/. [Accessed 23 04 2018].

[67] Center of register and information Systems, “E-business Register,” Center of register and information Systems, [Online]. Available: https://ariregister.rik.ee/index?lang=eng. [Accessed 23 04 2018].

[68] Centre of Register and Infromatin Systems, “E-laevakinnistusraamat,” Centre of Register and Infromatin Systems, [Online]. Available:

https://laevakinnistusraamat.rik.ee/. [Accessed 23 04 2018].

[69] Centre of Register and Infromatin Systems, “Artistic Associations,” Centre of Register and Infromatin Systems, [Online]. Available: https://ariregister.rik.ee/loomeliidud?lang=eng. [Accessed 23 04 2018].

[70] Centre of Registers and Information Systems, “Political Parties,” Centre of Registers and Information Systems, [Online]. Available: https://ariregister.rik.ee/erakonnad?lang=eng. [Accessed 23 04 2018].

[71] Centre of Registers and Information Systems, “Kinnistusraamat,” Centre of Registers and Information Systems, [Online]. Available: https://kinnistusraamat.rik.ee/detailparing/Avaleht.aspx. [Accessed 23 04 2018].

[72] M. Easterby-Smith, R. Thorpe and P. R. Jackson, Management Research, 2012.

[73] Estonian Data Protection Inspectorate, “Mis on isikuandmed,” Estonian Data Protection Inspectorate, [Online]. Available: http://www.aki.ee/et/mis-isikuandmed. [Accessed 23 04 2018].

[74] M. Jakobsson and J. Ratkiewicz, “Designing Ethical Phishing Experiments: A study of (ROT13) rOnl query features,” [Online]. Available: http://www2006.wwwconference.org/programme/files/pdf/3533.pdf. [Accessed 03 05 2018].