

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Tiina Käsukond 176841IAAM

**ANDMELAO ÄRIVAJADUSTE
KIRJELDAMINE JA ANDMEMUDELI
VÄLJATÖÖTAMINE SISEMINISTEERIUMI
NÄITEL**

Magistritöö

Juhendaja: Priit Rospel
MSc

Tallinn 2019

Kokkuvõte

Käesoleva magistr töö eesmärgiks oli luua rahvastikuregistri järelvalve teostamiseks vajalike andmete analüüsimist võimaldav andmeladu, mis võimaldaks saada operatiivsemat ülevaadet rahvastikuregistris teostatavatest päringutest ja tõsta seeläbi rahvastikuregistri järelvalve teostamise efektiivsust. Rahvastikuregistri järelvalve teostamiseks ei ole organisatsioonis kasutusel eraldiseisvat analüüsikeskkonda, kuid huvi ja vajadus selle järgi on olemas.

Eesmärgi täitmiseks kaardistas autor olemasoleva rahvastikuregistri järelvalve teostamise protsessi ja tõi välja protsessi eelised ja puudused, kirjeldas rahvastikuregistri järelvalve teostamise ärivajadused, pakkus välja sobivaima andmelao arhitektuuri, kirjeldas andmete eraldamise, teisendamise ja laadimise protsessi ning disainis andmelao kontseptuaalse andmemudeli.

Protsessi kaardistamiseks kogus autor informatsiooni rahvastikuregistri järelvalvet teostava isikuga läbiviidud intervjuude abil ning kasutades BPMN modelleerimise keelt, modelleeris autor rahvastikuregistri järelvalve üldise protsessi ning ühe alamprotsessi. Kaardistatud hetkeolukorra analüüsi tulemusena selgus, et olemasolevate vahenditega on kasutajal piiratud ligipääs informatsioonile, käsitsi analüüsiks vajalike andmete kogumine on aeganõudev ning olemasolev rahvastikuregistri päringu andmete hoiustamise ja välja võtmise tehniline lahendus avaldab negatiivset mõju andmebaasi jõudlusele.

Kirjeldatud puuduste põhjal analüüsis autor andmelao loomise otstarbekust. Selleks modelleeris autor TO-BE versiooni järelvalve teostamise alamprotsessist, kui loodaks andmeladu ning analüüsis, kas andmelao loomisega on võimalik olemasolevaid puudusi kõrvaldada. Analüüsi tulemusena selgus, et andmelao loomisega on võimalik olemasolevaid puudusi kõrvaldada ning vähendada oluliselt, tegevuste arvu poolest, rahvastikuregistri järelvalve teostamiseks vajalike andmete kogumisele kuluvat aega.

Intervjuu käigus kogus autor rahvastikuregistri järelevalve teostamise ärivajadused ja määras koos kasutajaga nende prioriteedid, seostas need andmeallikates olevate andmetega ning analüüsis nende teostatavust. Analüüsi tulemusena selgus, et kõiki kirjeldatud ärivajadusi on andmeteallikatest saadavate andmete põhjal võimalik realiseerida. Samuti hindas autor andmeallikate andmete kvaliteeti, mis on hea.

Andmelao arhitektuuri valikul saab arvesse võtta mitmeid kriteeriume ning maailmapraktikaid. Analüüsidest erinevaid arhitektuure, sealhulgas SMIT-is kasutusel olevat hübriidarhitektuuri, tegi autor ettepaneku luua rahvastikuregistri järelevalve teostamise andmeladu SMIT-is realiseeritud hübriidarhitektuuri näitel. Valitud andmelao arhitektuuri põhjal disainis autor kontseptuaalse andmelao andmemudeli, ning kirjeldas andmete eraldamise, teisendamise ja laadimis protsesse, mis toetavad andmelao eesmärgipärast toimimist.

Töö eesmärgid said täidetud ning magistritöö tulemusena valmis kirjeldus ärivajadustest ja andmelao kontseptuaalne andmemudel. Töö tulemust saab kasutada sisendina andmelao projekti nõuete kirjeldamiseks riigihankeks valmistumisel ning andmelao arendamise etapis tehnilise andmemudeli disainimisel ja ärivajaduste realiseerimisel.

Summary

The aim of this master's thesis was to create a data warehouse that allows analyzing the data necessary for the supervision of the population register, which would enable to obtain a more efficient overview of the requests carried out in the population register and thereby increase the efficiency of the supervision of the population register. There is no separate analysis environment in the organization for supervision of the population register, but there is interest and need for it.

To achieve this goal, the author mapped out the existing supervision process of population register and highlighted the advantages and disadvantages of the process, described the business needs for supervising the population register, proposed the most suitable data warehouse architecture, described the process of data extraction, transformation and loading, and designed a data warehouse conceptual data model.

To map the process, the author gathered information through interviews with the related party, and using the BPMN modeling language, the author modeled the general process of supervising the population register and one sub-process. As a result of the mapped current situation analysis, it was found that access to information by the user is limited by the available tools, the collection of data necessary for manual analysis is time-consuming and the existing technical solution for storing and extracting data from the population register has a negative impact on the performance of the database.

Based on the shortcomings described, the author analyzed the feasibility of creating a data warehouse. To do this, the author modeled the TO-BE version of the subprocess to create a data warehouse and in analyzing whether the data warehouse to be created solves the existing problems. As a result of the analysis it was found that by creating a data warehouse it is possible to eliminate the existing problems and to reduce considerably, in terms of the number of activities, the time required for the collection of data necessary for supervising the population register

During the interview, the author collected the business needs for supervising the population register and set their priorities with the user, linked them to data from data sources, and analyzed their feasibility. As a result of the analysis, it was found that all the described business needs can be realized based on data from data sources. The author also evaluated the data quality of data sources, which is good.

Various criteria and world practices can be taken into account when choosing a data warehouse architecture. Analyzing different architectures including the hybrid architecture used in SMIT, the author suggested creating a data warehouse for supervising the population register on the example of hybrid architecture implemented in SMIT. Based on the selected data warehouse architecture, the author designed a conceptual data warehouse data model and described data extraction, transformation and loading processes that support the purposeful operation of the data warehouse.

The goals of the work were fulfilled and the master's thesis resulted in a description of the business needs and the data warehouse conceptual data model. The result of the work can be used as an input to describe the requirements of the data warehouse project in preparation for the public procurement and in the data warehouse development stage in the design of the technical data model and implementation of the business needs.