

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Jorgen Ööpik 164345IABB

**ÜLIÕPILASORGANISATSIOONI
BEST-ESTONIA UUDE ANDMEBAASI
ANDMETE ÜLEKANDMISE
ETTEVALMISTAMINE**

bakalaureusetöö

Juhendaja: Erki Eessaar
PhD

Tallinn 2022

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Jorgen Ööpik

30.12.2021

Annotatsioon

Käesoleva töö siht on toetada üliõpilasorganisatsiooni BEST-Estonia liikmelisuse ja punktisüsteemi uue tarkvara realiseerimist. Töö sisuks on organisatsiooni valduses olevate andmete korrastamine ning nende ülekandmise ettevalmistamine uude MariaDB andmebaasisüsteemi abil loodud andmebaasi.

Eesmärgi saavutamiseks tuuakse töös välja andmesiirde ja andmekvaliteedi teoreetilised põhimõtted ning põhjendatakse valitud lahendusi. Seejärel analüüsitakse olemasolevate andmete kvaliteeti ning luuakse SQL laused andmete üleviimiseks uude loodud andmebaasi.

Töö tulemusena on valminud andmesiirdeks vajalikud SQL laused, need on BEST-Estoniale zip faili kokkupakituna üle antud ning on kasutusele võetud ajutised andmekvaliteedi parandamise meetmed kuniks uues andmebaasis olevate andmete haldamiseks on loodud toimiv tarkvara. Tegelik andmete ülekanne uude andmebaasi toimub alles siis, kui on valminud liikmelisuse ja punktide haldamise tarkvara uus versioon, mis kasutab uut andmebaasi, kuhu andmete ülekandmist käesolev töö ette valmistab.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 64 leheküljel, 6 peatükki, 28 joonist, 4 tabelit.

Abstract

Preparation for the Data Transfer to the New Database of the Student Organization BEST-Estonia

The goal of this bachelor thesis is to support the development of a new version of the student organization BEST-Estonia membership- and point management software. The subject matter of the thesis is to analyze and improve data quality and to create a script consisting of SQL sentences for the data transfer to the newly created database. The actual data transfer will happen after a new version of the membership- and point management software will be ready (creating the software is outside the scope of the thesis). However, the goal is to prepare the transfer so that it could be done right after the new software will be completed and the organization could then immediately stop the use of the old and problematic software.

The thesis starts with a short explanation of theoretical aspects of data quality and data transfer that are important for this work. In addition, it provides a description of the current state of data collection in the organization and explains the work process and tools behind writing the thesis.

Next, the results of the work are presented. Firstly, preparation for the data transfer led to the refinement of requirements to the new database. Thus, the thesis presents the changes in the new database. Secondly, summary of the results is presented in terms of different numeric values that characterize input data as well as extraction, transformation and loading of the data. For instance, in total 250 SQL statements were created. Completion of the entire prepared data transfer process took about 17.5 minutes. Executing the statements for cleansing the data took the largest part of the time (about 10 minutes).

Lastly, the analysis of the provided solution is presented. The author describes quality problems in the existing data. There are a lot of different problems. The author explains concrete SQL statements that were created to extract the data from the data sources, transform (cleaning, integrating) the data in the dedicated integration area (a separate schema in the target database), and load the data to the data structures of the new database. The cleaning of data takes place in the integration area. Data is not changed in the source systems (an old MariaDB database; Google Sheets) to preserve the original data. Moreover, the author explains temporary (until the new system version will be launched) prevention measures that were established to reduce

quality problems in the new data that will be registered by using the old systems. In addition, the author envisions future steps in the system development. The SQL statements have been given to BEST-Estonia in a zip file and the author will conduct a training to explain the usage of statements and temporary data quality improvement measures. The author hopes to participate in the actual data transfer.

The thesis is in Estonian and contains 64 pages of text, 6 chapters, 28 figures and 4 tables.

Lühendite ja mõistete sõnastik

BEST-Estonia	<i>Boards of European Students of Technology</i> , tudengiorganisatsiooni haru Tallinna Tehnikaülikoolis
CSV	<i>Comma Separated Values</i> . Failivorming, kus andmeväljade eraldajaks on komad
DMTG	Disaini ning turundusega tegelev töögrupp BEST-Estonias
Ettevalmistusala	Ajutiselt loodav skeem sihtandmebaasis, milles loodud tabelitesse koondatakse andmeid ja kus korrastatakse andmeid
FRTG	Sponsorlusega tegelev töögrupp BEST-Estonias
Google Forms	Google vormid, pilves asuv tarkvara, mille abil saab luua andmete sisestamiseks mõeldud vorme
Google Sheets	Google arvutustabelid, tabelitööluse tarkvara
Juhatus	BEST-Estonia viieliikmeline juhtorgan, mille liikmed juhivad organisatsiooni tööd
Koordinaator	Füüsiline isik, kes juhib töögrupi tegevusi
MTG	Muusikaga seotud tegevustega tegelev töögrupp BEST-Estonias
MTÜ	Mittetulundusühing
Revisjon	BEST-Estonia kolmeliikmeline organ, mille liikmed kontrollivad juhatuse tegevust
RV	Rahvusvahelise BEST-i organisatsiooni harudega tegelev töögrupp BEST-Estonias
Sihtandmebaas	Uus MariaDB abil loodud andmebaas
SQL	<i>Structured Query Language</i> . Standardiseeritud andmebaasikeel
TTG	Tehnika ja IT alane tegelev töögrupp BEST-Estonias
XAMPP	Vabavaralise tarkvara kogum, mis sisaldab Apache HTTP Serveri, MariaDB andmebaasisüsteemi ning PHP ja Perli programmeerimiskeeles kirjutatud skriptide käitamise tarkvara ning on mõeldud nendel põhinevatel süsteemide arendamiseks [1]

Sisukord

1 Sissejuhatus.....	12
1.1 Töö taust ja lühikirjeldus	12
1.2 Probleem ja eesmärgid	12
1.3 Ülevaade töö struktuurist	14
2 Teoreetiline taust.....	16
2.1 Andmekvaliteet	16
2.1.1 Andmete profileerimine ja standardiseerimine.....	16
2.2 Andmesiire	17
2.3 Andmesiirde tööriistadest.....	19
2.3.1 Hevo Data.....	19
2.3.2 IRI NextForm	20
3 Metoodika	21
3.1 Ülevaade töö objektist.....	21
3.1.1 Lähteandmed liikmelisuse andmebaasis.....	21
3.1.2 Lähteandmed avaldustest.....	24
3.1.3 Lähteandmed punktisüsteemis.....	24
3.2 Ülevaade töö protsessist	25
3.3 Ülevaade kasutatud vahenditest	27
4 Töö tulemused.....	29
4.1 Muudatused andmebaasi ülesehituses	29
4.2 Andmesiire	30
4.2.1 Tulemused ettevalmistusalal.....	31
4.2.2 Tulemused sihtandmebaasis	33
4.2.3 Tulemused SQL laused – skript.....	36

5 Töö tulemuste analüüs	38
5.1 Andmekvaliteedi probleemid lähteandmetes ja nende parandamine	38
5.1.1 Faili kodeeringu täpsustus	38
5.1.2 Meiliaadressid ja telefoninumbrid	39
5.1.3 Täisnime lahutamine kaheks osaks.....	40
5.1.4 Isiku unikaalse identifikaatori puudumine punktisüsteemis.....	41
5.1.5 Arvuliste väärtuste andmetüübi muutmine.....	42
5.1.6 Puuduvad liikmelisuse andmed	43
5.1.7 Liikmelisuse seisundi kuupäevade otsimine.....	45
5.2 Ennetusmeetmed	47
5.3 Andmete ülekanne	48
5.3.1 Eriala, liikmelisuse seisundi liigi ning isikute andmete ülekanne	49
5.3.2 Liikmelisuse andmete ülekanne.....	51
5.3.3 Punktide andmete ülekanne	52
5.4 Järgmised tööd.....	55
5.5 Hinnang tulemusele.....	55
6 Kokkuvõte.....	57
Kasutatud kirjandus	58
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	60
Lisa 2 – Osapoolte ja isikute register BEST-Estonia infosüsteemi uues versioonis.....	61
Lisa 3 – Liikmelisuste register BEST-Estonia infosüsteemi uues versioonis.....	62
Lisa 4 – Punktide saamiste register BEST Estonia infosüsteemi uues versioonis.....	63
Lisa 5 – Klassifikaatorite register BEST Estonia infosüsteemi uues versioonis	64

Jooniste loetelu

Joonis 1. Osa tabelist Liikmed.....	21
Joonis 2. Osa tabelist Staatus.....	22
Joonis 3. Osa tabelist Eriala.....	22
Joonis 4. Näide avalduse vastustest.....	24
Joonis 5. Näide ühest punktisüsteemi failist Google Sheets-is.....	25
Joonis 6. Näide dispetšer lehest [T] Tehniline.....	25
Joonis 7. Andmesiirde protsess.....	27
Joonis 8. Tabel Eriala uuendatud versioon.....	30
Joonis 9. Tabel Liikmelisuse_seisundi_liik uuendatud versioon.....	30
Joonis 10. CSV failidest tulenevate tabelite kodeeringu täpsustused.....	39
Joonis 11. Asendus meiliaadressi loomine isikute andmetes.....	39
Joonis 12. Telefoninumbri redigeerimine õigesse formaati.....	40
Joonis 13. Eesnime ning perenime lahutamine 2021. aasta avalduste näitel.....	41
Joonis 14. Punktisumma formaatimine ning teisendamine kohaliku taseme punktitali näitel.....	43
Joonis 15. Isiku ühendamine liikmelisuse tabeli identifikaatoriga ees- ja perenime alusel.....	43
Joonis 16. Abitabeli puudulikud_liikmed loomine.....	44
Joonis 17. Puuduvate isikute otsimine punktide tabelis ning nende lisamine tabelisse puudulikud_liikmed.....	44
Joonis 18. Puuduolevate isikute id-ga seostamine ning liikmelisuse andmebaasi lisamine 2019 avalduse näitel.....	45
Joonis 19. Abitabeli uued_kpvd loomine.....	46
Joonis 20. Isikute nimede lisamine abitabelisse uued_kpvd.....	46
Joonis 21. Projekti kuupäeva lisamine ning järgneva koosoleku kuupäevaga asendamine. ...	47
Joonis 22. Liikmelisuse staatuse kuupäeva ülekanne liikmelisuse tabelisse.....	47

Joonis 23. Ennetusmeede – praeguses tarkvaras täpsustavad lausedid kuupäevadele.....	48
Joonis 24. Eriala ning liikmelisuse seisundi liigi andmete ülekanne.....	50
Joonis 25. Isikuandmete ülekanne sihtandmebaasi.....	51
Joonis 26. Isikute ülekanne liikmelisuse tabelisse.....	52
Joonis 27. Isikute liikmelisuse kuupäevade ülekanne täisliikme kuupäeva näitel.....	52
Joonis 28. Punkti andmete ülekanne MTG töögrupi näitel.....	54

Tabelite loetelu

Tabel 1. Lähteandmete tabelite veergude kirjeldused.....	23
Tabel 2. Ettevalmistusallas olevad tabelid ning nende kirjeldused ja neid iseloomustavad arvulised väärtused.....	32
Tabel 3. Andmesiirde vastavustabel ettevalmistusala veergudest sihtandmebaasi veergudesse.	34
Tabel 4. Sihtandmebaasi veergude vaikumisi väärtused.	36

1 Sissejuhatus

Sissejuhatuses antakse ülevaade töö taustast ning struktuurist ning kirjeldatakse käsitlevat probleemi ja töö eesmärki.

1.1 Töö taust ja lühikirjeldus

Üliõpilasorganisatsiooni BEST-Estonia kahe liikme – Karl Henri Kaljus ja Stella Leego – lõputööde tulemusel on valminud uus andmebaas, mis peab parandama ja kiirendama organisatsiooni infohaldust [2] ja punktide haldust [3]. Nende lõputööde raames leiti detailsed nõuded süsteemi tööks vajalikele andmetele, planeeriti süsteemi funktsionaalsust ning prototüübiti kasutajaliidest.

Hetkeseisuna on olemas ainult punktisüsteemi halduse uus kasutajaliides. Paraku ei ole see kasutusel ning vajab kõigile personalijuhil soovidele ja vajadustele vastamiseks uuendamist. Liikmelisuse halduse allsüsteemi jaoks pole samuti veel loodud uut toimivat kasutajaliidest. Sellest tulenevalt jätkub organisatsioonis andmete kogumine hetkel vanal viisil. See tähendab, et punktide sisestamine toimub tabelitötluse rakenduse Google Sheets vahendil põhinevas süsteemis ja liikmelisuse andmed on BEST-Estonia poolt renditud *zone.ee* virtuaalserveris paiknevas MariaDB andmebaasis.

Töö autori hinnangul on hetkel organisatsioonile kõige vajalikum valmistada ette olemasolevate andmete ülekandmine e andmesiire uude andmebaasi. Kui valmib uue struktuuriga andmebaasil põhinev liikmelisuse ja punktihalduse allsüsteemide tarkvara, siis saab koheselt läbi viia ettevalmistatud andmete ülekande ning lõpetada andmete kogumise vanal ja ebaefektiivsel viisil.

1.2 Probleem ja eesmärgid

Hetkel toimub liikmete ja punktide haldamine ebaefektiivselt ning nimetatud süsteemid ei ole omavahel ühenduses. Tarkvarad ei jõusta andmetele kehtivaid piiranguid ning ei teki ülevaadet liikmelisuse seisundimuudatuste ega punktide redigeerimise ajaloo kohta. Seetõttu ei ole võimalik liikmelisuse tarkvaral kasutajatele kuvada täpseid liikmete arve mingi ajahetke

seisuga, liikmete organisatsioonis tegutsemise aega ning eelnevaid ajutisi juhirolle (juhatusi, koordinaatoreid, revisjoni läbiviijaid). Selle info leidmine eeldab mahukat ja aeganõudvat käsitööd Google Sheets failidega ning virtuaalserveris olevas andmebaasis, mis ei pruugi kokkuvõttes anda täpseid tulemusi.

Organisatsiooni käsutuses olevate andmete kvaliteet on halb. Selle põhjusteks on:

- erinevate süsteemide kasutamisest tulenev andmete dubleeritus, kontrollimatu andmete liiasus (andmete sünkroniseerimine ei toimu automaatselt) ning selle tagajärjel tuleb kooskõla luua käsitsi,
- andmetele kehtivate piirangute mittejärgustamine tarkvara poolt,
- puuduv ühtne reeglistik selle kohta, millised peavad olema sisestatavad andmed.

Eelneva tõttu ei teki tervikpilti liikmetest ja nende tegevusest. Kuna ei ole tarkvaralist ühendust liikme ja tema kohta sisestatud punktide vahel ning puuduvad süsteemisesed piirangud ja kontrollid, siis on olemas suur inimlike vigade tekkimise oht, mille tagajärjel kuvatakse süsteemi kasutajatele lõpuks ebaõiget informatsiooni.

Selleks, et realiseerida ja kasutusele võtta eelnevates lõputöodes projekteeritud tarkvara, on tarvis eelnevalt kogutud andmed uude andmebaasi üle kanda. Kuna kuni uue tarkvara valmimiseni jätkatakse andmete registreerimist kasutusel olevatesse andmebaasidesse, siis ei saa lõplikku andmete ülekannet kahjuks kohe (lõputöö tulemusena) ära teha. Küll aga saab selle ette valmistada, et uue tarkvara valmides saaks andmed kohe üle kanda ja vanade süsteemide kasutamisest koheselt loobuda.

Lõputöö eesmärgiks on luua tarkvara (skript – SQL lausete hulk), mille abil andmed olemasolevatest süsteemidest uue süsteemi andmebaasi üle kanda. Kuigi andmeid hetkel üle ei kanta on eesmärgiks luua skript, mille saaks tulevikus muutmata või minimaalsete muudatustega käivitada. Eesmärgiks on üle kanda kogu andmete ajalugu, st mitte jätta midagi vanadesse süsteemidesse maha. Kuna praegu on andmed registreeritud viisil, mis jätab sellesse ajalukku auke, siis on ühtlasi eesmärgiks andmete ülekandmise käigus neid auke täita, üritades tuletada puuduvaid andmeväärtuseid olemasolevatest andmeväärtustest. Sisuliselt on vaja andmed olemasolevatest andmebaasidest välja lugeda, neid andmeid töödelda ja lõpuks kanda töödeldud andmed uude andmebaasi. Siin on tegemist andmeaitadest/andmevakkadest tuntud

ETL (*Extract-Transform-Load*) protsessi analoogiga [4], ainult et andmeid ei laadita mitte andmeaita/andmevakka, vaid uue struktuuriga operatiivandmete andmebaasi.

Eesmärgi saavutamiseks on vaja olemasolevate andmete kvaliteeti analüüsida, leida üles puudused andmetes ning andmeid võimalusel täiendada ja korrastada, et viia need vastavusse uue süsteemi nõuetega. Andmete ülekandmiseks võetakse kasutusele eraldi andmebaasi skeem – ettevalmistusala – kus töödeldakse andmeid. Seal parandatakse nii vigu andmetes kui ka ühendatakse liikmelisuse süsteemi ja punktisüsteemi andmed. Andmete ülekandmise käigus peab töötlemine toimuma võimalikult palju tarkvara abiga, et vältida võimalikke inimlikke vigu (näiteks trükivigu) ja kiirendada ülekande protsessi. Samuti on vaja kasutusele võtta ajutised ennetusmeetmed, et kuni andmete lõpliku ülekandmiseni ja vanade süsteemide kasutamisest loobumiseni ei tekiks vanadesse andmebaasidesse uusi andmekvaliteedi probleeme.

1.3 Ülevaade töö struktuurist

Töö käigus analüüsitakse olemasolevaid andmeid, korrastakse neid SQL lausetega ning luuakse skript andmete üle kandmiseks uude andmebaasi. Lõputöö dokument on jaotatud neljaks põhiosaks.

Teoreetilise tausta all (peatükk 2) kirjutatakse andmekvaliteedi mõistest ning kvaliteedi parandamisest. Samuti kirjeldatakse andmesiirde protsessi ning selle erinevaid tehnilisi lahendusi ja põhjendatakse autori valikuid.

Metoodika osas (peatükk 3) antakse ülevaade süsteemi hetkeolukorrast, selgitatakse andmesiirde protsessi ja tuuakse välja kasutatud tehnilised vahendid.

Töö tulemuste osas (peatükk 4) on välja toodud töö resultaadi kirjeldused. Kõigepealt antakse ülevaade andmesiirde ettevalmistamise tulemusena täpsustunud nõuetest tulenevatest muudatustest uue andmebaasi struktuuris. Seejärel esitatakse erinevad arvulised väärtused (nagu näiteks lähtefailide arv, ridade arv uues andmebaasis, andmete ülekandeks kuluv aeg, loodud SQL lausete arv, koodiridade arv), mis iseloomustavad loodud lahendust. Lahenduse täpsem analüüs koos koodinäidetega on järgmises peatükis.

Viimases, töö tulemuste analüüsi osas (peatükk 5), tuuakse detailsemalt välja lähteandmete probleemid ja kirjeldatakse täpsemalt andmesirde tehnilist lahendust koos kordinäidetega. Samuti kirjutatakse, mis võiksid olla järgmised süsteemi loomise sammud ning antakse hinnang töö tulemusele.

2 Teoreetiline taust

Koostatud töö praktiline osa koosneb kahest tegevuste komplektist – andmekvaliteedi tõstmisest ning andmesiirdest. Vanades süsteemides ei olnud kasutusel andmete kontrollimeetmeid, mis takistaksid teatud (ilmselgelt ebaõigete) väärtuste registreerimist või nõuaksid teatud atribuutide väärtuste registreerimist. Seega personalijuhid, kes lisasid uusi liikmeid andmebaasi, jätsid näiteks registreerimata erinevaid kuupäevi või tegid vigu meiliaadresside ja telefoninumbrate sisestamisel. Selle tagajärjel ei ole organisatsioonil oma liikmetest terviklikku ja kvaliteetset ülevaadet.

Selles peatükis selgitatakse veidi erinevaid võimalusi, kuidas tagada paremat andmekvaliteeti ning tuuakse välja, millised lahendusi on olemas andmesiirdeks, et saavutada andmete võimalikult täielik ülekanne uude andmebaasi.

2.1 Andmekvaliteet

Andmed on märkide hulgad mingis keskkonnas. Digitaalsel kujul andmed võivad olla nii andmebaasisüsteemi abil loodud andmebaasides kui väljapool andmebaasisüsteemi hoold olevates failides. Ka käesolevas töös on osad lähteandmed MariaDB andmebaasis ja osad tabelarvutusprogrammi failides. Andmekvaliteet väljendub selles kui korrektselt on andmeid sisestatud, säilitatud ning töödeldud ning kui laialdast kasutust need võimaldavad. Andmekvaliteedi tagamine pole lihtsalt ühekordne tegevus, vaid see on jätkuv protsess, millele tuleb koguaeg tähelepanu pöörata [5]. Artiklis [5] on välja toodud andmekvaliteedi neli dimensiooni: täpsus, täielikkus, koosõlalisus ning olulisus. Antud jaotises selgitatakse, kuidas võiks toimuda andmete profileerimine, mis faktorid mõjutavad andmete kvaliteeti ning selgitatakse mõningaid võimalusi andmete parendamiseks.

2.1.1 Andmete profileerimine ja standardiseerimine

Andmete profileerimine on erinevatele meetoditele toetuv andmete analüüs, millest tulenevalt on võimalik tuvastada võimalikke puuduseid andmetes. Artiklis [6] on välja toodud andmekvaliteedi vead, mida andmete profileerimine võib aidata avastada.

- Puuduvad andmed.

- Valed andmed.
- Korduvad andmed.
- Mittekooskõlalised andmed, sh andmed mis on loogiliselt võimatud (nt isik viibib samal ajal kahes erinevas kohas).

Samas artiklis nimetatakse ka meetodeid, mis annavad informatsiooni andmete struktuuri, sisu ning kvaliteedi praegusest seisust [6]. Järgnevalt nimetatakse neid meetodeid, mida oli vaja käesolevas töös kasutada. Järgnev kirjeldus eeldab, et andmed on SQL-andmebaasis, kuid põhimõtteliselt samasuguseid meetodeid saab kasutada ka teiste andmemudelite korral.

- Atribuudi nime analüüs – Nimi peab selgelt kirjeldama, milliseid andmeid kogutakse ning nimest peab olema eeldatav andmetüüp.
- Andmetüübi analüüs – Andmetüüp peab vastama sellele, milliseid andmeid veerus hoiustatakse. Samasisulised andmed peavad olema sama andmetüübiga veergudes. Õige tüüp võimaldab tõhusamat tööprotsessi (vähem tüübiteisendusi operatsioonide läbiviimiseks).
- Andmete unikaalsuse analüüs – Analüüsi eesmärgiks on leida puuduvad andmed ja korduvad andmed. Andmete haldamise süsteemis peavad olema jõustatud reeglistikud, mis välistavad korduvate andmete registreerimise ja nõutud andmete registreerimata jätmise.
- Mustrite analüüs – Analüüsitakse ühes veerus olevaid andmeid ning tuvastatakse nende võimalikke iseärasusi. Näiteks uuritakse, kas telefoninumbrid on üheselt sisestatud või millised on meiliaadressi kirjapildid.
- Viidete analüüs – kas andmete kaudu esitatavad viited ühes tabelis viitavad olemasolevatele andmetele teises tabelis või on need viited katki.

E. Kavaku lõputöös on kirjeldatud andmete standardiseerimist andmekvaliteedi tõstmiseks [7]. Näiteks käesolevas töös lähtutakse eeldusest, et telefoninumbri alguses tuleb alati registreerida suunakood kujul: +<suunakood>.

2.2 Andmesiire

Andmesiire on käsitlevate andmete valimine, ettevalmistamine ning lõplik üleviimine sihtasukohta ja andmete lähteasukoha kasutamise lõpetamine [8]. See võib toimuda näiteks erinevatest failidest andmebaasisüsteemi abil loodud andmebaasi või ühest

andmebaasisüsteemi abil loodud andmebaasist teise andmebaasi. Keerukust tõstab see, et kasutatavad andmebaasisüsteemid võivad olla erinevad (sh põhineda erinevatel andmemudelitel) ja võimalike failivormingute hulk on suur. Selle protsessi juures on oluline arvestada andmekvaliteediga. Näiteks on võimalik, et andmete lähtekohas on andmekvaliteet kasutajate hinnangul piisav, kuid andmete sihtkohas on kvaliteedinõuded rangemad ja ülekantavad andmed ei vasta nendele nõuetele. Sellisel juhul tuleb mingil viisil kvaliteeti parandada või andmete ülekandmisest loobuda. Andmesiirdel on oluline selle ajastus, et ülekandmise ajal ning järel ei tekiks süsteemides tõrkeid [8]. Andmesiirde juures võib [9] tekitada probleeme näiteks järgnevalt.

- Lähteandmete tähendus on ebaselge, kuna selle kohta puuduvad mudelid.
- Andmete erinev struktuur lähtekohas ja sihtandmebaasis.
- Sihtandmebaasi rangemad piirangud andmetele võrreldes andmete allikaga.
- Korduvad ja vastuolulised faktid andmete allikas.
- Andmete puudumine andmete allikas, kusjuures puuduvate andmeväärtuste teadasaamiseks puudub võimalus või tuleb neid andmeid tuletada kaudselt erinevatest allikatest.

Käesolevas töös esinesid andmesiirde juures kõik nimetatud raskused. Olemasoleva süsteemi kohta ei ole mudeleid. Autor pidi tuginema varasematele lõputöödele [2, 3], oma kogemusele organisatsiooni liikmena ning intervjuudele teiste organisatsiooni liikmetega.

Andmesiirde puhul tuleb otsustada, kas kanda uude andmebaasi üle kõik andmed lähteandmebaasist või ainult teatud tingimustele vastavad andmed (näiteks ajaloolisi andmeid mitte kanda) [9]. Andmesiirde puhul tuleb otsustada, kas kanda andmeid üle järk-järgult ning kasutada paralleelselt nii vana kui uut andmebaasi või ühe korraga ning loobuda peale ülekannet vana andmebaasi kasutamisest [9]. Andmesiirde puhul tuleb otsustada, kas kanda andmed üle neid käigult teisendades või teha ülekannet mitmesammuliselt – kõigepealt andmeid koondades, integreerides ja puhastades ning siis tulemust sihtandmebaasi üle kandes [9].

Käesolevas töös kantakse uude andmebaasi üle kõik (ka ajaloolised) andmed ning andmete ülekandmiseks kasutatakse ettevalmistusala. Käesolev töö valmistab andmesiirde ette, kuid

andmed kantakse lõpuks üle ühekorruga ning siis loobutakse kohe ka vana andmebaasi kasutamisest.

Käesolevas töös sooritati andme ülekanne CSV failidest ja olemasolevast MariaDB andmebaasist ettevalmistusalasse ning sealt MariaDB sihtandmebaasi. Ettevalmistusala on eraldi skeem sihtandmebaasis. See skeem luuakse ülekande alguses ja eemaldatakse ülekande lõpus. Töös ei kasutatud spetsiaalselt välja töötatud teenuseid, sest ülekantavad andmed on isikuandmed ning nende andmete kolmandate osapooltega jagamine ei ole tulenevalt isikuandmete kaitse seadusest lubatud [10]. Kuna ülekantavate andmete hulk on väike, ülekanne on ühekordne tegevus ja organisatsioonil ei ole üleliigseid rahalisi vahendeid, siis tundus tasulise tarkvara hankimise ja õppimise asemel lihtsam ning odavam luua ühekordne lahendus SQL lausete kogumi näol. Samuti soovis autor harjutada SQL-lausete koostamist ning saada ülevaade kasutatavatest süsteemidest ning lähteandmetest.

2.3 Andmesiirde tööriistadest

Leidub mitmeid tööriistu teisendamaks andmeid soovitud formaati ja/või kandmaks andmeid üle soovitud asukohta. Tööriistade toetatud protsess koosneb enamasti automatiseeritult andmete valimisest, ettevalmistamisest ja muutmisest. Tööriist võib olla võimeline teostama pidevat ülekannet sihtkohta. Andmesiirde tööriistu kasutatakse näiteks andmete konsolideerimiseks mitmest lähtekohast ja ülekandeks ühte sihtkohta või serveri/andmekeskuse uuendusest tuleneval andmete ülekandel.

Järgnevalt esitatakse mõningad näited programmidest ning nende eelistest ja puudustest.

2.3.1 Hevo Data

Hevo keskkonnas on võimalik oma andmeülekanne mõne minutiga ettevalmistada. Tasuta versioonis on piiratud võimalused, kuid spetsiifilisemad andmesiirde olukorrad vajavad tasulist versiooni [11]. Selleks, et sooritada ülekannet Exceli failidest, on vaja soetada tasuline versioon. Keskkond sobiks ideaalselt neile, kellel on vaja andmete ülekanne teha Facebook Analytics või Youtube Analytics tööriistadest, kuid keskkond pakub ka teisi variante. Ülekantavad andmed peavad juba olema struktureeritud.

Vahend võimaldab kasutada üle saja erinevat tüüpi andmeallika. Nendest mõned tuntumad on Google Sheets, Google Drive ning GitHub.

2.3.2 IRI NextForm

IRI NextForm pakub lahendust vähemstruktureeritud andmete ülekandeks soovitud sihtandmebaasi. Tööriista kasutades on võimalik oma andmeid struktureerida ning redigeerida ning seejärel teostada ülekannet [12]. Maksumus sõltub andmete lähtekoha andmetüüpidest, ning sihtkoha riistvarast.

Andmeid saab näiteks üle kanda sellistest failidest.

- „.DOC, .DOCX
- .EML, .OST, .PST
- .PDF, RTF
- .PPT., .PPTX
- .TXT, .XML
- .XLS, .XLSX“ [13]

Andmeid saab näiteks üle kanda selliste SQL-andmebaasisüsteemide abil loodud andmebaasidesse.

- „DB2 UDB
- DB2 Mainframe²
- MySQL
- Oracle
- SQL Server
- Sybase
- Teradata“ [13]

3 Metoodika

Selles peatükis antakse detailsem ülevaade algsest andmete seisust ja lähteandmete asukohast. Samuti selgitatakse tööprotsessi. Lisaks tuuakse välja millised vahendid on andmesiidiks vajalikud ning milliseid vahendeid töö raames täpsemalt kasutati.

3.1 Ülevaade töö objektist

Töö kirjutamise ajal kogutakse töö kontekstis olulisi andmeid kolmest erinevast süsteemist:

- liikmelisuse andmed on BEST-Estonia poolt renditud zone.ee virtuaalserveris MariaDB andmebaasis,
 - hiljutised liikmeavaldused on ka Google Forms failides,
- punktisüsteemi andmed on sisestatud organisatsiooni poolt loodud Google Sheets-i keskkonda.

3.1.1 Lähteandmed liikmelisuse andmebaasis

Liikmelisuse andmebaas koosneb kokku kolmest tabelist: *Liikmed*, *Staatust* ja *Eriala*. Vana andmebaas on renditud serveris ning aja jooksul on sinna loodud ka mitmeid teisi tabeleid, mis tänapäeval enam kasutust ei leia. Tabelis *Liikmed* on 18.12.2021 seisuga 1469 rida. Tabel sisaldab andmeid, mis on liitunud liikmed täitnud pabervormis või Google Forms-is loodud avalduse vormis ning on personalijuhi poolt käsitsi andmebaasi sisestatud (vt Joonis 1).

id	muudetud	eesnimi	perenimi	pildifail	synniaeg	staatus	millal_besti	avaldus	t2isliikmeks	eriala
1	2004-03-09 16:54:32	Annika			1979-08-16	5	1998-05-06	0000-00-00	0000-00-00	5
1703	2021-04-29 14:27:11	Artur			1999-11-20	4	0000-00-00	2021-03-04	0000-00-00	95
3	2004-03-09 16:55:57	Eire			1979-05-01	5	1998-05-06	0000-00-00	0000-00-00	5
4	2004-09-29 17:04:47	Helena			1979-05-03	5	2000-03-03	0000-00-00	0000-00-00	3
5	2009-12-30 11:04:20	Heleri			1978-08-27	5	1997-12-03	2009-12-30	0000-00-00	6
6	2007-07-31 14:53:31	Jüri			1980-12-17	5	2000-06-07	2007-07-31	0000-00-00	3

Joonis 1. Osa tabelist *Liikmed*.

Vanas andmebaasis on staatuse ja eriala väärtused eraldi tabelites. Tabelis *Staatust* hoiustatakse organisatsioonile vajalikke liikmestaatuse kirjeldusi (vt Joonis 2) ning tabelisse *Eriala* on aja jooksul kogutud Tallinna Tehnikaülikoolis õpetatud erialade nimesid (vt Joonis 3). Nende tabelites olevate väärtuste määramine, lisamine ja muutmine toimub praeguses kasutatavas best.ee domeenis personalijuhi vaates. Seisuga 18.12.2021 on tabelis *Staatust* 14 rida ning tabelis *Eriala* 104 rida.

id	nimetus	nimetus_en
1	Juhatus liige	Member of board
2	Täisliige	Full member
3	Beebilliige	Baby member
4	Uus	Interested
5	Alumni	Alumni
6	Ekслиige	Ex-member

Joonis 2. Osa tabelist *Staatust*.

id	eriala	eriala_en
2	Ehitus	Civil Engineering
3	Informaatika	Computer Science
4	Mehaanika	Mechanics
5	Keemia	Chemistry
6	Majandus	Economics

Joonis 3. Osa tabelist *Eriala*.

Liikmelisuse tabelis välisvõtme kitsendusi ei ole. Õnneks viidete terviklikkuse reegli rikkumisi nende andmete puhul ei tuvastatud.

Eraldi tabelina pandi kirja vajalikud veerud MariaDB liikmelisuse lähteandmebaasis olevates kolmes tabelis (vt Tabel 1). Neid veerge kasutatakse andmesiirdes.

Tabel 1. Lähteandmete tabelite veergude kirjeldused.

Tabeli nimi	Veeru nimi	Selgitus
Liikmed	id	Isiku arvuline identifikaator tabelis
Liikmed	muudetud	Kuupäev, millal isiku andmeid viimati muudeti
Liikmed	eesnimi	Isiku eesnimi
Liikmed	perenimi	Isiku perekonnanimi
Liikmed	synniaeg	Isiku sünnikuupäev
Liikmed	staatus	Isiku liikme staatus arvulise identifikaatorina
Liikmed	millal_besti	Kuupäev, millal isik on saanud beebiliikmeks
Liikmed	avaldus	Kuupäev, millal isik esitas avalduse liitumiseks organisatsiooniga
Liikmed	t2isliikmeks	Kuupäev, millal isik on saanud täisliikmeks
Liikmed	eriala	Isiku eriala ülikoolis arvulise identifikaatorina
Liikmed	astus_kooli	Aasta, millal isik astus Tallinna Tehnikaülikooli
Liikmed	mobla	Isikuga kontakti saamiseks kasutatav telefoninumber
Liikmed	email	Isikuga kontakti saamiseks kasutatav meiliaadress
Staatus	id	Liikme staatuse arvuline identifikaator
Staatus	nimetus	Liikme staatuse nimetus eesti keeles
Staatus	nimetus_en	Liikme staatuse nimetus inglise keeles
Eriala	id	Eriala arvuline identifikaator
Eriala	eriala	Eriala nimetus eesti keeles
Eriala	eriala_en	Eriala nimetus inglise keeles

3.1.2 Lähteandmed avaldustest

Viimased neli aastat on selleks, et liikme avaldusi oleks võimalik täita veebis, kasutatud tarkvara Google Forms. See kiirendab isikute andmebaasi lisamise protsessi ning vähendab paberi kasutamise vajadust. Andmesirdeks on kokku kogutud 2019–2021 aastate personalijuhtide käest Google Forms-i avalduste failid (vt Joonis 4). Faile on kokku kolm tükki ning nendes on kokku 313 kirjet.

	A	B	C	D	E	F
1	Timestamp	Täisnimi	Sünniaeg (dd/mm/yyyy)	Meiliaadress	Telefoninumber	Eriala
2	2021-08-25	Anna	2000			Ärindus
4	2021-08-25	Ander	2000			IT süsteemide administreerimine
5	2021-08-25	Johannes	2001			Mereveendus ja sadamakorraldus
6	2021-08-25	Ralf	1997			Ärindus
7	2021-08-25	Karoline	2002			Rakenduskeemia, toidu- ja geenitehnoloogia
8	2021-08-25	Hardi	2001			Rakendusfüüsika
9	2021-08-25	Liis	2001			Robotika ja tootearendus
11	2021-08-25	Britta	2002			Rakenduskeemia, toidutehnoloogia ja geenitehnoloogia
12	2021-08-25	Johanna	2001			Ärindus
13	2021-08-25	Johannes	2002			Ehitus juhtimine ja ehitiste projekteerimine
14	2021-08-25	Rene	2000			Hoonete projekteerimine ja ehitusjuhtimine
15	2021-08-25	Robert	2000			Tootearendus ja robotika

Joonis 4. Näide avalduse vastustest.

3.1.3 Lähteandmed punktisüsteemis

Punktisüsteemi jaoks on loodud Google Sheets-is kuus erinevat faili ning kokku on punktide jaoks loodud kaheksa töölehte. Failid on loodud viiele töögrupile (FRTG, MTG, DMTG, TTG ja RV), kus on punktidele mõeldud üks leht ning üks üldisem personalijuhi (HR) fail, kuhu on loodud punktide sisestamiseks kolm töölehte erinevate kategooriate jaoks.

Neis failides salvestatakse punktide näol andmeid liikmete erinevate tegevuste kohta (vt Joonis 5). Igale failile on loodud ka nii-öelda dispetšer leht, kuhu on märgitud vaikimisi punktisummad erinevate tegevuste jaoks ja on välja toodud liikmete nimekiri (vt Joonis 6).

Punktisüsteem töötab põhimõttel, kus punktide sisestamise lehel valitakse liige, tema tegevus, punkte andva tegevuse sooritamise aeg ning lisatakse vajadusel kommentaar. Läbi valemite kuvatakse punktisumma.

2	Kes	Mida	Mitu	Aeg	Kommentaar	Punkte
3	Anne	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
4	Kaspar	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
5	Maarja	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
6	Kristel	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
7	Karin	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
8	Kristel	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
9	Kristiina	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
10	Katre	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
11	Mihkel	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
12	Karl-Ivar	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
13	Ebe-Kai	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
14	Liis	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
15	Moonika	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
16	Kirsti	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
17	Indra	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
18	Mari-Liis	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
19	Karin	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
20	Ingvar	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
21	Karolin	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
22	Kalmer	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
23	Janet	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
24	Karina	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@
25	Lauri	SKY koosolek	1	2015-Jan-1	Detsembri SKY	60@

+ ☰ Kohalik tase ▾ Meeskonnatöö ▾ Projektitimid ▾ [T] Tehniline ▾

Joonis 5. Näide ühest punktisüsteemi failist Google Sheets-is.

Liikmed	Projektid	Punktid
Nimed		Kohaliku taseme punktid
Näide Näide	EBEC	800@ Koolitusel osalemine (1h) (vana) 20@
Kristiina	Võti Tulevikku	800@ Koosolekul osalemine (1h) (vana) 20@
Karl-Ivar	Suvekursus	800@ SKY koosolek 60@
Maria	GA	800@ Koolituse läbiviimine (1h) 40@
Tarvo	IPF	800@ Kohalikul üldkogul osalemine 200@
Mihkel	EBEC Baltic	800@ BESTi partneri üldkogu 200@
Anne	Enginaator	800@ Kohvipausi ettevalmistamine 10@
Kaspar	bePROUD	800@ Protomine 10@
Kristel	JLT	400@ Varia 10@
Karmen	JBT	400@ Korraldus RV üritus mis ei toimunu 300@
Rainer	Department Meeting	400@
Ingvar	TTA+DM+IPF Pre/Post	400@

Joonis 6. Näide dispetšer lehest [T] Tehniline.

3.2 Ülevaade töö protsessist

Enne andmete ülekandmist oli vaja saada ligipääs Google Drive keskkonda, kus asetsevad praeguse punktisüsteemi failid ning ligipääs BEST-Estonia virtuaalserverile zone.ee keskkonnas, kus asub liikmete andmetega MariaDB andmebaas ning kus on ka uus andmebaas,

millesse andmete ülekannet ette valmistatakse. Andmete ülekanne toimub samuti selles virtuaalserveris.

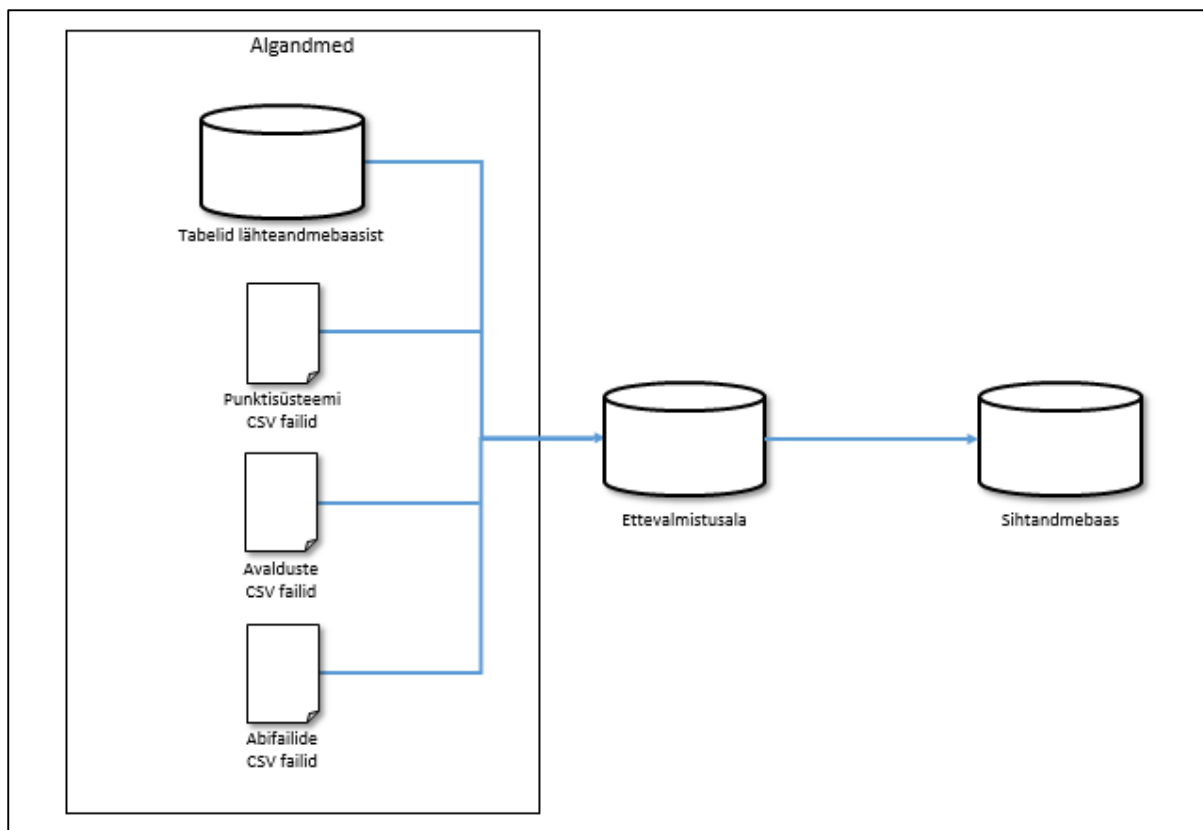
Uues andmebaasis, kuhu on vaja andmeid üle kanda, loodi eraldi skeem – ettevalmistusalala. Ettevalmistusalal toimuvad andmete muudatused/parandused ja sealt kantakse andmed lõpuks üle uue sihtandmebaasi skeemi tabelitesse. Andmete ülekandmise protsess koosneb mitmest etapist (vt Joonis 7).

Esimeses etapis laetakse alla kõik *Google Drive*-i keskkonnas olevad punktisüsteemi failid Exceli failiformaadis. Exceli failides tehakse väiksemaid parandusi. Seejärel teisendatakse need CSV formaadis failideks ning imporditakse need failid ettevalmistusalale. Ettevalmistusalale tekivad tabelid, mille struktuur oli samasugune nagu CSV failides olevate andmete struktuur. Seejärel eksporditakse liikmelisuse andmebaasist kolm tabelit ning imporditakse need tabelid ettevalmistusalale.

Teises etapis imporditakse kolme aasta avalduste andmed ning kaks abifaili ettevalmistusalale. Ka need lähteandmed olid CSV formaadis. Üks neist abifailidest sisaldab punktisüsteemi tehnilisi väärtuseid (vt Joonis 2) ja teine sisaldab ajaloolisi organisatsiooni juhtrollide ametiaegsid ning positsioone, mis on koostatud eelneva personalijuhi poolt Exceli failina.

Kolmandas etapis analüüsitakse olemasolevaid andmeid ja lisatakse abiveerge ning abitabeleid, eesmärgiga siduda ära isikute liikmelisuse ja isikutele määratud punktide andmed. Samuti parandatakse liikmelisuse andmekvaliteeti, registreerides puuduvaid kuupäevi, formaatides andmeid vastavalt nõutele ning lisades teatud isikutele juurde juhirollide andmeid.

Viimases etapis sooritatakse andmesiire sihtandmebaasi skeemi andmete vastavustabeli järgi (vt Tabel 3).



Joonis 7. Andmesiirde protsess.

Eelneva tulemusena valmis SQL lausete komplekt andmesiirde läbiviimiseks. Tegelik andmesiire toimub alles siis, kui on loodud uus toimiv tarkvara liikmelisuse ja punktide haldamiseks.

Google Sheets töölehtedes, Google Forms failides ja alguses MariaDB andmebaasis olevates tabelites andmemuudatusi ei tehta. Selle põhjuseks on soov säilitada algandmed ning mitte häirida olemasoleva vananenud süsteemi tööd. Ülekandmisel tekkivate probleemide korral saab kasutada muudatusteta algandmeid.

3.3 Ülevaade kasutatud vahenditest

BEST-Estonia renditavas virtuaalserveris on installeeritud vabavaraline XAMPP tarkvarakomplekt, kuhu kuuluvatest programmidest leidsid käesolevas töös otsest või kaudset kasutust Apache HTTP Server, MariaDB andmebaasisüsteem ning PHP.

SQL lausete kirjutamiseks oli kasutatud koodiredaktori tarkvara Brackets. Brackets on tasuta ning avatud lähtekoodiga tarkvara [14]. Selle tarkvara kasuks otsustamise tingis tarkvara kiire avamine, lihtne koodi redigeerimise võimaldamine ning autori varasem kogemus selle tarkvara kasutamisega. Kuigi Brackets on mõeldud rohkem veebiarenduste koodi kirjutamise jaoks, on see ka sobiv SQL lausete kirjutamiseks.

Valdavalt oli töökeskkonnaks MariaDB (ver 10.4.21) andmebaasisüsteem. Andmebaasi visualiseerimiseks kasutati andmebaasi haldusvahendit PhpMyAdmin. Sama vahend on andmete visualiseerimiseks kasutusel BEST-Estonia renditavas virtuaalserveris, millest on kirjutatud Stella Leego lõputöös. Autor lõi seal õppimiseks ja testimiseks eraldi andmebaasi skeemi.

Veebirakendusest Google Sheets ei kanta andmeid otse MariaDB andmebaasi ettevalmistusalale, sest lähteandmetest on kõigepealt vaja eemaldada veerud ja sümbolid, mis põhjustavad ülekandel vigaseid tulemusi. Seepärast kasutatakse andmete korrastamiseks tabelitöötlusprogrammi Excel. Selles eemaldatakse punktisüsteemi statistika kuvamiseks loodud veerud ning andmetes olevad sümbolid, mis CSV failiks konverteerimisel ja ettevalmistusalasse kandmisel tekitavad vigu. Excelis salvestatakse punktisüsteemi failid CSV formaadis failidena ning nendes olevad andmed laetakse ettevalmistusalale. Iga CSV faili kohta tekib ettevalmistusalal eraldi tabel.

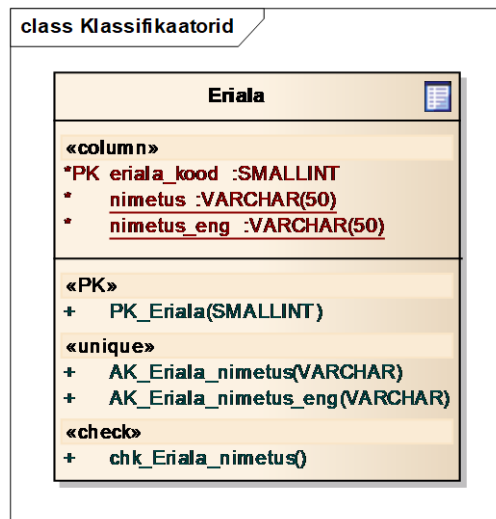
4 Töö tulemused

Selles peatükis antakse ülevaade ettevalmistamise tulemusena tehtud muudatustest sihtandmebaasis. Samuti esitatakse selles peatükis erinevad arvulised väärtused (näiteks ridade arv ettevalmistusalas ja uues andmebaasis, andmete ülekandeks kuluv aeg, loodud SQL lausete arv ja füüsiliste koodiridade arv), mis iseloomustavad loodud lahendust.

4.1 Muudatused andmebaasi ülesehituses

Analüüsidest olemasolevaid andmeid ning vaadates üle eelnevate lõputöodes välja toodud nõuded, ilmnes vajadus lisada loodud andmebaasis tabelitesse uusi veerge.

BEST-Estoniaga on läbi aastate liitunud inglise keeles kõnelevaid isikuid, kuid uus andmebaas ei sisalda andmeid, et kasutajatele teatud informatsiooni inglise keeles kuvada. Seega lisati tabelisse *Eriala* uus veerg *nimetus_en* (vt Joonis 8). See tagab, et vajadusel on võimalik tulevikus kasutajatele kuvada erialade nimetusi inglise keeles. Samasugune täiendus tehti tabelis *liikmelisuse_seisundi_liik*, kuhu lisati veerg *nimetus_en* (vt Joonis 9). Täiendus annab lisafunktsionaalsuse kuvada organisatsiooni liikmelisuse staatuste nimesid inglise keeles. Eelnev loob muuhulgas võimaluse esitada digitaalseid organisatsiooni liitumise avaldus inglise keeles. Vanas andmebaasis on lähteandmetena olemas ingliskeelsed tõlked, mida saab uude andmebaasi üle kanda. Selline mitmekeelse andmebaasi sisu hoidmiseks mõeldud disainilahendus vastab disainimustrile „tõlkeveerud“ [15].



Joonis 8. Tabel *Eriala* uuendatud versioon.



Joonis 9. Tabel *Liikmelisuse_seisundi_liik* uuendatud versioon.

4.2 Andmesiire

Andmesiire viidi läbi kasutades peatükis 3 kirjeldatud meetodikat. Selles peatükis on esitatud vastavustabel ettevalmistusala tabelite veergude ja sihtandmebaasi veergude vahel. Peatükk esitab samuti erinevaid töö mahukust ja ulatust kirjeldavaid arvulisi näitajaid andmete ettevalmistusala, sihtandmebaasi ja loodud SQL lausete kohta. Konkreetsete SQL lausete näited on peatükis 5.

4.2.1 Tulemused ettevalmistusalal

Ettevalmistusalas loodi kokku 18 tabelit. Need tabelid sisaldavad lähteandmeid, mida ettevalmistusalas omakorda töötlemata hakatakse.

- Kolm liikmelisusega seonduvaid isikuandmete tabelit (vanast MariaDB andmebaasist).
- Kaheksa punktisüsteemist tulenevat punktisüsteemide tabelit.
- Kolm Google Forms-is täidetud avalduste andmetega tabelit.
- Kaks imporditud abifaili väärtuste tabelit. Üks tabel sisaldab punktisüsteemi dispetšerfaili väärtuseid (valdkondi, tegevusi ja nende vaikimisi punktisummasid) ja teises tabelis on ajaloolised juhtrollide ametiajad ning positsioonid.
- Kaks koostatud abitabelit, millest ühe tabeli abil otsitakse liikmeks astumise kuupäevi ning teises tabelis eraldatakse välja puuduliku informatsiooniga isikud, keda ei eksisteeri vanas liikmelisuse andmebaasis, kuid neile on punktisüsteemis punkte antud. Abitabelite täpsemad selgitused on välja toodud jaotistes 5.1.6 ja 5.1.7

Ettevalmistusalasse 13 CSV faili importimine võttis aega kuus minutit. Iga faili tuli importida ükshaaval, sest PhpMyAdmin ei võimalda korraga mitut faili valida ning iga kord tuleb valida õiged sätted, et andmed loetaks sisse õiges formaadis. Andmete sisselugemiseks kuluv aeg sõltub failis olevates ridade hulgast. Õiged sätted tuleb valida vahendis PhpMyAdmin *Import* vahelehel ning need on välja toodud allpool:

- Character set: windows-1252 [16]
- Format: CSV
- Columns separated with: sümbol – ;
- Columns enclosed with: sümbol – ;
- First line of the file contains the table column names: Valitud

PhpMyAdmin vanast MariaDB andmebaasist kolme tabeli (*Liikmed*, *Eriala* ja *Staatust*) eksportimise aega ei näita, kuid see tundus väga kiire. Eksportimise tulemuseks on SQL lausetega failid. Kolme faili importimine ettevalmistusalasse võttis kokku aega 74 sekundit. Importimisel tuleb valida õiged sätted, mis on välja toodud allpool:

- Character set: windows-1252

- SQL Compatibility mode: MYSQL323

Imporditud tabelitesse lisati abiveerge andmete korrastamiseks (andmete õigesse tüüpi teisendamine, isikunime lahutamine eraldi eesnime ja perenime komponentideks, unikaalsete identifikaatorite määramine). Detailsemad abiveergude selgitused on jaotises 5.1.

Tabelis 2 on välja toodud ettevalmistusalas olevad tabelid ning nende tabelite ridade ja veergude arv. Arvulised väärtused on esitatud seisuga 11.12.2021. Andmesirde tegelikul läbiviimisel võib ridade arv suureneada, kuid veergude arv ei peaks muutuma.

Tabel 2. Ettevalmistusalas olevad tabelid ning nende kirjeldused ja neid iseloomustavad arvulised väärtused.

Tabeli nimi	Tabeli kirjeldus	Veergude arv	Ridade arv
2019_avalduis	2019 Google Forms-is esitatud avaldused	15	92
2020_avalduis	2020 Google Forms-is esitatud avaldused	15	139
2021_avalduis	2021 Google Forms-is esitatud avaldused	15	82
dmtg_punktid	DMTG töögrupi punktid	13	1437
mtg_punktid	MTG punktid	13	229
frtg_punktid	FRTG töögrupi punktid	13	1820
ttg_punktid	TTG töögrupi punktid	13	224
rv_punktid	RV töögrupi punktid	13	613
hr_kohalik_tase	Kohaliku taseme punktid	13	10630
hr_meeskonnatoo	Meeskonnatöö punktid	13	2530
hr_projektitiimid	Projektitiimide punktid	11	329
liikmed	BEST-Estonia liikmete nimekiri	41	1469
eriala	TalTech-i eriala nimetused	3	104
staatus	Liikmelisuse staatuste nimekiri	4	14

Tabeli nimi	Tabeli kirjeldus	Veergude arv	Ridade arv
juhtrollid	Juhtrolli ametiajad ja positsioonid	8	138
valdkond_tegevus	Punktisüsteemi dišpetserfaili nimekiri	5	130
uued_kpvd	Liikmelisuse kuupäevade genereerimise abitabel	8	130
puudulikud_liikmed	Puuduvate isikute nimekiri	7	113

4.2.2 Tulemused sihtandmebaasis

Andmesiirdel toimus ülekanne kokku 11 tabelisse (*Eriala*, *Isik*, *Liikmelisus*, *Liikmelisuse_seisundi_liik*, *Liikmelisuse_syndmus*, *Osapool*, *Punkti_saamine*, *Punkti_saamise_syndmus*, *Tegevus*, *Valdkond*, *Valdkonna_tegevus*) ja 45-sse tabeli veergu. Nendest 40 veergu on andmed kantud lähteandmetest ning viide veergu on vaikumisi väärtustega. Sihtandmebaasi struktuur on esitatud lisades 2–5.

Vastavustabelis kirjeldatakse, millistest ettevalmistusala tabelitest ja veergudest kanti üle andmed millistesse sihtandmebaasi tabelitesse ja veergudesse (vt Tabel 3). Välja tuuakse veerud kuhu lähteallikatest andmeid üle kantakse. Vastavustabelist on välja jäetud sihtandmebaasi veerud (vt Tabel 4), mille puhul kasutatakse vaikumisi väärtuseid .

Tabel 3 esitab ainult neid tabelleid, kust kanti andmeid sihtandmebaasi. Osad ettevalmistusala tabelid on lihtsalt abitabelid, mida kasutati ainult ettevalmistusala siseselt. Tabelis 3 ei ole mainitud kolme avalduse faili tabelit, sest nendes olevad andmed kantakse sisse ettevalmistusalas olevasse tabelisse *Liikmed*. Tabelis 3 on toodud välja ettevalmistusala tabel *Punktid*, mis väljendab punktisüsteemist tuleneva kaheksa faili alusel loodud tabelleid – kõigist nendest tuli andmed ükshaaval üle kanda.

Tabel 3. Andmesiirde vastavustabel ettevalmistusala veergudest sihtandmebaasi veergudesse.

Ettevalmistusala tabel	Ettevalmistusala veerg	Sihtandmebaasi tabel	Sihtandmebaasi veerg	Ridade arv
Eriala	<i>id</i>	Eriala	<i>eriala_kood</i>	104
Eriala	<i>eriala</i>	Eriala	<i>nimetus</i>	104
Eriala	<i>eriala_en</i>	Eriala	<i>nimetus_eng</i>	104
Juhtrollid	<i>isik_id</i>	Liikmelisuse_ syndmus	<i>liikmelisus_id</i>	138
Juhtrollid	<i>staatus_id</i>	Liikmelisuse_ syndmus	<i>seisund_syndmuse_ jarel</i>	138
Juhtrollid	<i>kes</i>	Liikmelisuse_ syndmus	<i>kommentaar</i>	138
Juhtrollid	<i>millal</i>	Liikmelisuse_ syndmus	<i>toimumise_aeg</i>	138
Liikmed	<i>id</i>	Isik	<i>isik_id</i>	1582
Liikmed	<i>email</i>	Isik	<i>e_meil</i>	1582
Liikmed	<i>eesnimi</i>	Isik	<i>eesnimi</i>	1582
Liikmed	<i>perenimi</i>	Isik	<i>perenimi</i>	1582
Liikmed	<i>hyydnimi</i>	Isik	<i>hyydnimi</i>	1582
Liikmed	<i>mobla</i>	Isik	<i>telefoni_nr</i>	1582
Liikmed	<i>synniaeg</i>	Isik	<i>synni_kp</i>	1582
Liikmed	<i>muudetud</i>	Isik	<i>reg_aeg</i>	1582
Liikmed	<i>muudetud</i>	Isik	<i>viimase_ muutmise_aeg</i>	1582
Liikmed	<i>id</i>	Liikmelisus	<i>isik_id</i>	1582
Liikmed	<i>eriala</i>	Liikmelisus	<i>eriala_kood</i>	1582
Liikmed	<i>staatus</i>	Liikmelisus	<i>Liikmelisuse_ seisundi_ liik_kood</i>	1582

Ettevalmistusala tabel	Ettevalmistusala veerg	Sihtandmebaasi tabel	Sihtandmebaasi veerg	Ridade arv
Liikmed	<i>astus_kooli</i>	Liikmelisus	<i>sisseastumise_aasta</i>	1582
Liikmed	<i>millal_besti, avaldus, t2isliikmeks</i>	Liikmelisuse_syndmus	<i>toimumise_aeg</i>	3485
Liikmed	<i>millal_besti, avaldus, t2isliikmeks</i>	Liikmelisuse_syndmus	<i>reg_aeg</i>	3485
Liikmed	<i>id</i>	Osapool	<i>osapool_id</i>	1582
Punktid	<i>kord_punkte</i>	Punkti_saamine	<i>punktisumma</i>	17806
Punktid	<i>Mitu_dec</i>	Punkti_saamine	<i>tegevuse_arv</i>	17806
Punktid	<i>Aeg</i>	Punkti_saamise_syndmus	<i>toimumise_aeg</i>	17806
Punktid	<i>Aeg</i>	Punkti_saamise_syndmus	<i>reg_aeg</i>	17806
Punktid	<i>Kommentaar</i>	Punkti_saamise_syndmus	<i>kommentaar</i>	17806
Staatus	<i>id</i>	Liikmelisuse_seisundi_liik	<i>Liikmelisuse_seisundi_liik_kood</i>	16
Staatus	<i>nimetus</i>	Liikmelisuse_seisundi_liik	<i>nimetus</i>	16
Staatus	<i>nimetus_en</i>	Liikmelisuse_seisundi_liik	<i>nimetus_eng</i>	16
Valdkond_tegevus	<i>tegevus_kood</i>	Punkti_saamine	<i>tegevus_kood</i>	17806
Valdkond_tegevus	<i>valdkond_kood</i>	Punkti_saamine	<i>valdkond_kood</i>	17806
Valdkond_tegevus	<i>tegevus_kood</i>	Tegevus	<i>tegevus_kood</i>	114
Valdkond_tegevus	<i>tegevus</i>	Tegevus	<i>nimetus</i>	114
Valdkond_tegevus	<i>punktisumma</i>	Tegevus	<i>vaikimisi_punktid</i>	114
Valdkond_tegevus	<i>valdkond_kood</i>	Valdkond	<i>valdkond_kood</i>	8
Valdkond_tegevus	<i>valdkond</i>	Valdkond	<i>nimetus</i>	8

Ettevalmistusala tabel	Ettevalmistusala veerg	Sihtandmebaasi tabel	Sihtandmebaasi veerg	Ridade arv
Valdkond_ tegevus	<i>tegevus</i>	Valdkonna_ tegevus	<i>tegevus_kood</i>	130
Valdkond_ tegevus	<i>valdkond</i>	Valdkonna_ tegevus	<i>valdkond_kood</i>	130

Sihtandmebaasi on loodud uute funktsionaalsuse toetuseks mõned veerud, mille jaoks ei ole lähteandmeid. Seega sisestati nendesse veergudesse vaikimisi väärtused. Tabelis 4 tuuakse välja sihtandmebaasi veergude vaikimisi väärtused. Vaikimisi väärtuste selgitused on jaotises 5.3.

Tabel 4. Sihtandmebaasi veergude vaikimisi väärtused.

Tabeli nimi	Veeru nimi	Vaikimisi väärtus
Isik	isiku_seisundi_liik_kood	1
Isik	on_nous_isikuandmete_kasutamine	0
Liikmelisus	on_nous_avaldamisega	0
Punkti_saamine	punkti_saamise_seisundi_liik_kood	1
Punkti_saamine_syndmus	registreerija_id	1094

4.2.3 Tulemused SQL laused – skript

Töö tulemusena loodi kaks SQL lausete skriptifaili – *siire_ettevalmistus* ja *siire_andmesiire*. Esimene fail koosneb ettevalmistusalas andmete korrastamise lausetest ja teine fail koosneb ainult andmete ülekandmiseks vajalikest lausetest. Skriptid on eraldi failides, et lihtsamalt eristada lauseid, mis on mõeldud andmete ettevalmistamiseks ning lauseid, mis on mõeldud andmete ülekandmiseks sihtandmebaasi.

Ettevalmistusala skripti füüsiliste koodiridade arv on 901 ning andmesiirde skripti koodiridade arv on 784 – kokku 1685 rida. Koodiridade arvu lugemiseks lähtuti LOC (*Lines of Code*) põhimõttest [17], mis saavutati spetsiaalselt koodiridade lugemiseks loodud programmiga *cloc* [18]. Enne koodiridade arvu leidmist formaaditi koodi kasutades programmi Instant SQL

Formatter [19] valides sealt andmebaasisüsteemiks MySQL. Formaatumise põhimõtted olid järgmised.

Suurtähtedega on:

- klauslite võtmesõnad (nt *INSERT,SELECT,DELETE FROM*).

Väiketähtedega on:

- funktsioonid (nt *concat,length,trim*),
- andmetüübid (nt *int, decimal,varchar*),
- veergude nimed.

Ridade lahutamine ning joondamine.

- Avaldised, klauslid ja funktsioonid algavad uuest reast.
- Kui on veergudele viitamised listina, siis pannakse veergude nimed eraldi ridadele, kusjuures koma on veeru nime järgi.
- SQL klauslid ja veergudele viitamised on eraldatud tühikutega ja joondatud vasakule.
- Kommentaarid on eraldi ridadel.
- SQL laused on üksteisest eraldatud kommentaari märgisega --.

Skriptides olevate SQL lausete koguarv on 250, millest

- 155 on UPDATE laused,
- 40 on INSERT laused,
- 33 on ALTER laused ja
- 22 on DELETE laused.

Skriptide täitmine võttis aega 10 minutit ja 14 sekundit, sellest andmete korrastamine 10 minutit ja 2 sekundit ning andmete ülekande sihtandmebaasi 12 sekundit.

Kogu andmesiirde protsess koos lähteandmete üleslaadimine ettevalmistusalale, andmete korrastamise ning andmete sihtandmebaasi ülekandmisega võttis aega 17 minutit ja 31 sekundit. Selles ajas ei ole arvestatud failide allalaadimist, CSV formaati teisendamist ega abifailide uuendamist.

5 Töö tulemuste analüüs

Selles peatükis analüüsitakse töö tulemusi. Peatükis kirjeldatakse, milliseid probleeme sisaldasid lähteandmed ja kuidas neid lahendati ning milliseid ennetusmeetmeid rakendati, et vanades süsteemides uute andmete sisestamisel ei tekiks uusi andmekvaliteedi puudujääke. Samuti tuuakse välja, millistest lausetest koosneb andmesiire uude andmebaasi.

5.1 Andmekvaliteedi probleemid lähteandmetes ja nende parandamine

Selles jaotises tuuakse välja lähteandmetes olevad probleemid. Iga probleemi kohta selgitatakse loodud lahendust ja tuuakse joonisena koodinäide. Lähtetabeleid on mitu, kuid näitena tuuakse välja laused vaid ühe tabeli kohta. Ülejäänud SQL laused on analoogilised – koodis muutub vaid tabeli nimetus.

Ettevalmistusalas olevates samasisulistest tabelites loodud abiveerge nimetatakse samanimeliselt, et tabelite loetavus on selgem. Näiteks ettevalmistusalas on mitu punktisüsteemist tulnud andmetega tabelit: *mtg_punktid*, *tig_punktid*, *frtg_punktid*. Kõigis nendes tabelites on veergude nimed samad, et tabelid oleksid lihtsamini loetavad. Importimisel tulnud veergude on nimed juba failides samad (seda juba Google Sheetsist allalaadimisel) ning abiveerud mis on vaja luua ettevalmistusalas (nt *Mitu_dec*, *eesnimi*, *perenimi*) on ka läbi tabelite ühesuguste nimedega.

5.1.1 Faili kodeeringu täpsustus

CSV failide importimisel ettevalmistusalasse ei ole vaikumisi arvestatud eesti keeles kasutatavate võõrtähtedega š ja ž. Selleks, et andme korrastamine annaks korrektse tulemuse ning SQL lausete käivitamisel tuvastatakse võõrtähti, on lisatud ettevalmistuse skripti algusesse tabeli kodeeringu täpsustamine *utf8_estonian_ci* (vt Joonis 10) [20].

```

ALTER TABLE 2019_avalldus CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE 2020_avalldus CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE 2021_avalldus CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE juhtrollid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE frtg_punktid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE hr_kohalik_tase CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE hr_meeskonnatoo CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE hr_projektitiimid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE mtg_punktid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE dmtg_punktid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE ttg_punktid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE rv_punktid CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;
ALTER TABLE valdkond_tegevus CONVERT TO CHARACTER SET utf8 COLLATE utf8_estonian_ci;

```

Joonis 10. CSV failidest tulenevate tabelite kodeeringu täpsustused.

5.1.2 Meiliaadressid ja telefoninumbrid

Vanas liikmelisuse tabelis puuduvad osadel liikmetel meiliaadressid. Tegemist on juba organisatsioonist lahkunud liikmetega. Seega määrati nende liikmestaatuseks *Ebaaktiivne* ning loodi neile ajutine asendus-meiliaadress lõpuga *@vajakindlasti.muuta* (vt Joonis 11). Selle taga on idee, et kui tulevikus soovivad sellise meiliaadressiga isikud organisatsiooniga taasliituda, siis on nende isikuandmed juba süsteemis olemas, kuid enne kui neid aktiivseks muuta, tuleb neilt meiliaadress üle küsida ning andmebaasis uuendada.

Sama lahendust on ka rakendatud nende isikute puhul, kelle isikuandmed puuduvad liikmelisuse andmebaasis, kuid nende isikute nimed eksisteerivad punktisüsteemis ning neid lisatakse ettevalmistusalas olevasse liikmelisuse tabelisse.

```

UPDATE liikmed
SET staatus = 7
WHERE email = "";
--
UPDATE liikmed
SET email=IF(email=" ",Concat(eesnimi,".",perenimi,"@vajakindlasti.muuta"),email);

```

Joonis 11. Asendus meiliaadressi loomine isikute andmetes.

Uues andmebaasis on isikutel kohustuslik registreerida telefoninumber, kusjuures see peab algama suunakoodiga (Eesti numbril puhul +372). Telefoninumbri väärtus võib seal sisaldada

vaid numbrimärke, sidekriipse, tühikuid ning + märke. Vanas andmebaasis ei olnud telefoninumbri veerul kitsendusi. Seega osad telefoninumbri olid ilma suunakoodideta ja sisaldasid uue andmebaasi jaoks ebasobivaid sümboleid. SQL lausetega muudeti telefoninumbri nõuetele vastavaks (vt Joonis 12).

```
UPDATE liikmed
SET `mobla` = substring(`mobla`,7)
WHERE `mobla` LIKE "(%";
--
UPDATE liikmed
SET `mobla` =replace(`mobla`,`;`,`");
WHERE id=26;
--
-- Korrastada telefoninumber formaadile: +372 xx xxx xxx
--
UPDATE liikmed
SET `mobla` = replace(`mobla`, " ", "");
--
UPDATE liikmed
SET `mobla` = concat("+372",`mobla`)
WHERE `mobla` NOT LIKE " "
AND `mobla` NOT LIKE "%+%"
AND `mobla` NOT LIKE " ";
--
UPDATE liikmed
SET `mobla` = concat(LEFT(`mobla`,4),' ',
                    substring(`mobla`,5,2),' ',substring(`mobla`,7,3),
                    ' ',RIGHT(`mobla`,length(mobla)-9));
--
UPDATE liikmed
SET `mobla` = NULL
WHERE `mobla` LIKE "%";
```

Joonis 12. Telefoninumbri redigeerimine õigesse formaati.

5.1.3 Täisnime lahutamine kaheks osaks

Uues andmebaasis on isikute eesnimed ja perenimed eraldi veergudes. Google Forms-is täidetud avaldustes ei ole isiku nimi osadeks jaotatud ning eesnimi ja perenimi on esitatud ühe väärtusena. Sama probleem on ka punktisüsteemis olevate nimedega.

Lahendusena on tekitatud ettevalmistusala tabelites, kus on täisnimi ühe väärtusena, ka eraldi veerud eesnime ja perenime jaoks. Seejärel lahutatakse SQL-lausetega eesnimi ja perenimi eraldi veergudesse (vt Joonis 13). Töö tulemusena loodud skript loeb perenimeks väärtuse

pärast viimast tühikut ning kogu eelnevat käsitletakse kui eesnime. Seega loodud lahendus ei töötle õigesti selliseid isikunimesid, kus on kaks eesnime ning kaks perenime. Käesolevas töö tulemusena ei leitud algandmetes selliseid nimesid, kuid loodi ennetusmeede, millest on räägitud jaotises 5.2.

```
ALTER TABLE 2021_avalduk
    ADD COLUMN 2021_a_id          int NOT NULL auto_increment PRIMARY KEY first,
    ADD COLUMN isik_id           int after 2021_a_id,
    ADD COLUMN perenimi         varchar (48) after täisnimi,
    ADD COLUMN eesnimi          varchar(48) after täisnimi;

--
DELETE
FROM 2021_avalduk
WHERE 2021_a_id = 1;
--
UPDATE 2021_avalduk
SET   täisnimi = rtrim(täisnimi)
WHERE täisnimi LIKE '% ';
--
UPDATE 2021_avalduk
SET   eesnimi = substr(täisnimi, 1, length(täisnimi) - length(substring_index(täisnimi, ' ', -1)) - 1);
--
UPDATE 2021_avalduk
SET   perenimi = substring_index(täisnimi, ' ', -1);
--
UPDATE 2021_avalduk
SET   eesnimi = perenimi
WHERE eesnimi = "";
--
UPDATE 2021_avalduk
SET   perenimi = NULL
WHERE eesnimi = perenimi;
--
UPDATE 2021_avalduk
SET   eesnimi = rtrim(eesnimi)
WHERE eesnimi LIKE '% ';
--
UPDATE 2021_avalduk
INNER JOIN liikmed
ON     liikmed.eesnimi = 2021_avalduk.eesnimi
AND    liikmed.perenimi = 2021_avalduk.perenimi
SET    2021_avalduk.isik_id = liikmed.id;
```

Joonis 13. Eesnime ning perenime lahutamine 2021. aasta avalduste näitel.

5.1.4 Isiku unikaalse identifikaatori puudumine punktisüsteemis

Punktisüsteemi failides on punkte saanud isiku kohta registreeritud ees- ja perekonnanimi, kuid mitte liikmelisuse andmebaasi *isik_id* väärtust ega ka mitte mingit muud isikut unikaalselt identifitseerimist võimaldavat väärtust. Ees- ja perenime kombinatsioon ei saa olla isiku unikaalne identifikaator, st mitmel erineval inimesel võib olla sama isikunimi. Samas on isikunimi ainus andmeväärtus, mille alusel on praegu võimalik erinevatest allikatest pärinevaid isikuandmeid siduda. Olemasolevate andmete analüüs ei leidnud liikmete hulgast korduvate nimedega inimesi, mis ei tähenda, et neid ei võiks tulevikus lisanduda.

Autor uuris personalijuhilt, kuidas nad käituvad sellises olukorras kui organisatsioonis on samanimelised inimesed. Selgus, et praegu ei olegi sellele probleemile lahendust. Autori eelduste kohaselt lisatakse sellisel juhul inimeste eristamiseks nimele mingi unikaalsust tagav arv ning vastavalt on välja pakutud ka ennetusmeede (vt jaotist 5.2).

5.1.5 Arvuliste väärtuste andmetüübi muutmine

Punktisüsteemist arvulised väärtused imporditi ettevalmistusalasse kui tekstiväärtused. Samuti oli importimise tulemusel arvu taga tundmatu sümbol, mis tuleneb Google Sheets-is kasutatud punkti märgisest Ⓢ punktisumma taga. Sihtandmebaasis salvestatakse punktide andmed arvutüüpi (*decimal*) veerus komakohtadega arvudena.

Ettevalmistusala punkti tabelites loodi uued *decimal* tüüpi veerud *Mitu_dec* ja *Punkte_dec*. Selleks, et andmeid saaks salvestada sihtandmebaasi arvulise väärtusena, eemaldati ettevalmistusalal SQL lausega tundmatu sümbol punkte esitava väärtuse lõpust. Kui punktisüsteemis on punktide kustutamisest jäänud alles tühjasid ridasid, siis eemaldatakse need DELETE lausega. Seejärel tõsteti UPDATE lausega tekstilised väärtused loodud veergudesse ühtlasi andmeid käigupealt teisendades. Nii toimiti kõikide ettevalmistusala tabelitega, kus asuvad liikmete punktide andmed.

Igasse punktisüsteemist imporditud failist tekkinud tabelisse loodi ka identifikaatori veerg, mis aitab punktide sisestusi üle kanda.

```

ALTER TABLE hr_kohalik_tase
  ADD COLUMN hr_k_t_id          int NOT NULL auto_increment PRIMARY KEY first,
  ADD COLUMN perenimi          varchar (48) after kes,
  ADD COLUMN eesnimi           varchar(48) after kes,
  ADD COLUMN isik_id           int after hr_k_t_id,
  ADD COLUMN mitu_dec          decimal(6,2),
  ADD COLUMN punkte_dec        decimal(6,2);

--
DELETE
FROM   hr_kohalik_tase
WHERE  kes=""
OR     punkte="";
--
UPDATE hr_kohalik_tase
SET    punkte=LEFT(punkte,Length(punkte)-1);
--
UPDATE hr_kohalik_tase
SET    mitu_dec = Replace(mitu,',','.');
--
UPDATE hr_kohalik_tase
SET    punkte_dec = Replace(punkte,',','.');

```

Joonis 14. Punktisumma formaatimine ning teisendamine *kohaliku taseme* punktitali näitel.

Punktisüsteemi tabelites ja avalduste tabelites loodi veerg *isik_id*, millele lisati liikmelisuse tabelist isikule vastav arvuline väärtus, otsides vastet eesnime ja perenime alusel (vt Joonis 15).

```

UPDATE frtg_punktid
INNER JOIN liikmed
ON      liikmed.eesnimi = frtg_punktid.eesnimi
AND     liikmed.perenimi = frtg_punktid.perenimi
SET     frtg_punktid.isik_id = liikmed.id;

```

Joonis 15. Isiku ühendamise liikmelisuse tabeli identifikaatoriga ees- ja perenime alusel.

5.1.6 Puuduvad liikmelisuse andmed

Sidudes liikmelisuse tabeli isikute identifikaatorit *isik_id* punktisüsteemi punktide tabeliga ees- ja perenime alusel, tulid ilmsiks puuduolevad liikmed, keda ei eksisteeri praeguses liikmelisuse andmebaasis. Tegemist on isikutega, kes on organisatsioonis saanud punkte üksikute tegevuste eest, kuid pole aktiivselt BEST-Estonias tegutsenud. Seetõttu on küll salvestatud nende tegemised punktisüsteemis, kuid nende andmeid pole lisatud praegusesse liikmelisuse andmebaasi. Selleks, et nende punkte mitte ära kaotada, filtreeriti nad eraldi loodud abitabelisse *puudulikud_liikmed* (vt Joonis 16). Seejärel otsiti puuduolevaid isikuid nime järgi punktide tabelitest (vt Joonis 17). Järgnevalt otsiti isikuid avalduste tabelites ning kui leiti neile vastavad

andmed, siis lisati nende andmed liikmelisuse tabelisse (vt Joonis 18). Selline lahendus hetkel töötas, sest organisatsioonis ei olnud korduvate nimedega inimesi.

```
CREATE TABLE puudulikud_liikmed
(
    id INT AUTO_INCREMENT PRIMARY KEY,
    eesnimi VARCHAR(48),
    perenimi VARCHAR(48),
    2019_avalduid INT,
    2020_avalduid INT,
    2021_avalduid INT,
    puudu_info INT DEFAULT 0
);
```

Joonis 16. Abitabeli *puudulikud_liikmed* loomine.

```
INSERT
INTO puudulikud_liikmed
(
    eesnimi,
    perenimi
)
SELECT DISTINCT eesnimi,
                perenimi
FROM            hr_kohalik_tase
WHERE          hr_kohalik_tase.isik_id IS NULL;

--
INSERT IGNORE
INTO  puudulikud_liikmed
(
    eesnimi,
    perenimi
)
SELECT DISTINCT eesnimi,
                perenimi
FROM            hr_meeskonnatoo
WHERE          hr_meeskonnatoo.isik_id IS NULL;
```

Joonis 17. Puuduvate isikute otsimine punktide tabelis ning nende lisamine tabelisse *puudulikud_liikmed*.

```

-- Lisame avalduste tabelitest neile isikutele id, kellel seda saab
UPDATE puudulikud_liikmed
INNER JOIN 2019_avalduk
ON         2019_avalduk.eesnimi = puudulikud_liikmed.eesnimi
AND       2019_avalduk.perenimi = puudulikud_liikmed.perenimi
SET       puudulikud_liikmed.2019_avalduk_id = 2019_avalduk.2019_a_id;
--
-- Avalduse lehelt liikmete lisamine
INSERT INTO liikmed
(
    muudetud,
    eesnimi,
    perenimi,
    synniaeg,
    staatus,
    avalduk,
    eriala,
    astus_kooli,
    mobla,
    email,
    t2isliikmeks
)
SELECT     Now(),
           2019_avalduk.eesnimi,
           2019_avalduk.perenimi,
           `Sunniaeg (dd/mm/yyyy)`,
           7,
           `timestamp`,
           7,
           year(`timestamp`),
           telefoninumber,
           meiliaadress,
           0000-00-00
FROM       2019_avalduk
INNER JOIN puudulikud_liikmed
ON         puudulikud_liikmed.eesnimi = 2019_avalduk.eesnimi
AND       puudulikud_liikmed.perenimi = 2019_avalduk.perenimi
WHERE     puudulikud_liikmed.2019_avalduk_id IS NOT NULL;

```

Joonis 18. Puuduolevate isikute id-ga seostamine ning liikmelisuse andmebaasi lisamine 2019 avalduse näitel.

5.1.7 Liikmelisuse seisundi kuupäevade otsimine

Vanas liikmelisuse andmebaasis olid puudu paljud isikute organisatsiooniga liitumise või erinevate teiste seisundimuudatuste kuupäevad (need olid seal tähistatud 0000-00-00). Selle põhjus on, et vanas tarkvaras ei ole automatiseeritud kuupäevade lisamist ning eelnevad personalijuhid jätsid kuupäeva väljad mõnikord täitmata. Lahendusena otsiti punktisüsteemist tabelist *Projektitiimid* liikmete staatuse muutumise kuupäevaid. BEST-Estonias järgitakse liikmestaatuse saamisel põhimõtet, kus isik peab olema sooritanud organisatsioonis projekti, et saada endale liikmestaatus. Liikmestaatus saadakse organisatsiooni igakuisel üldkoosolekul.

Sellest põhimõttest lähtuvalt loodi SQL laused, võttes välja kõigi isikute nimed tabelist *Projektitiimid* ning lisades need abitabelisse *uued_kpvd* (vt Joonis 19). Seejärel otsiti välja nende isikute projekti sooritamise kuupäevad ning leiti tabelist *hr_kohalik_tase* projektile järgnev esimene üldkoosolek ning määrati see kui liikmestaatuse saamise kuupäev (vt Joonis 21). Selliselt leitud kuupäevad ei pruugi olla täpsed. Seega kasutati sellisel viisil leitud kuupäevi vaid juhul kui isikul see kuupäev liikmelisuse andmebaasis puudus (vt Joonis 22).

```
CREATE TABLE uued_kpvd
(
    id                INT AUTO_INCREMENT PRIMARY KEY,
    isik_id           INT,
    eesnimi           VARCHAR(48),
    perenimi          VARCHAR(48),
    millal_besti      DATE,
    t2isliikmeks      DATE,
    i_projekt         VARCHAR(8),
    ii_projekt        VARCHAR(8)
);
```

Joonis 19. Abitabeli *uued_kpvd* loomine.

```
INSERT INTO uued_kpvd
(
    eesnimi,
    perenimi
)
SELECT DISTINCT eesnimi,
                perenimi
FROM            hr_projektitiimid
WHERE          lõpp != '1990-01-01';
```

Joonis 20. Isikute nimede lisamine abitabelisse *uued_kpvd*.

Tabelis *hr_projektitiimid* on sisestusi, millele on lähteandmetes määratud kuupäev 01.01.1990. Säärast lahendust on kasutatud punktisüsteemis salvestamiseks ajaloolisi punktisisestusi. SQL lause elimineerib need, sest vastasel juhul kuvatakse vigane kuupäev veerus *millal_besti* (vt Joonis 20).

```

UPDATE uued_kpvd
INNER JOIN hr_projektitiimid
ON      hr_projektitiimid.eesnimi=uued_kpvd.eesnimi
AND     hr_projektitiimid.perenimi=uued_kpvd.perenimi
SET     `millal_besti` = hr_projektitiimid.lõpp
WHERE   hr_projektitiimid.lõpp != '1990-01-01' ;
--
UPDATE uued_kpvd
INNER JOIN hr_kohalik_tase
ON      hr_kohalik_tase.eesnimi=uued_kpvd.eesnimi
AND     hr_kohalik_tase.perenimi=uued_kpvd.perenimi
SET     `millal_besti` =
      (
          SELECT DISTINCT hr_kohalik_tase.aeg
          FROM            hr_kohalik_tase
          WHERE           hr_kohalik_tase.aeg > millal_besti
          AND             hr_kohalik_tase.mida='SKY koosolek'
          ORDER BY       hr_kohalik_tase.aeg ASC limit 1);

```

Joonis 21. Projekti kuupäeva lisamine ning järgneva koosoleku kuupäevaga asendamine.

```

UPDATE liikmed
INNER JOIN uued_kpvd
ON      uued_kpvd.isik_id = liikmed.id
SET     liikmed.millal_besti = uued_kpvd.millal_besti
WHERE   liikmed.millal_besti = "0000-00-00";

```

Joonis 22. Liikmelisuse staatuse kuupäeva ülekanne liikmelisuse tabelisse.

5.2 Ennetusmeetmed

Kuna tegelik andmesiire toimub alles siis kui on loodud uus liikmelisuse ja punktide haldamise tarkvara, siis võidakse vahepeal vanasse liikmelisuse andmebaasi lisada uusi vigaseid andmeid. Suurimaks ohukohaks on sündmuste kuupäevade sisestamata jätmine või valesti sisestamine, samasuguste isikunimedega inimesed, ning isikunimed, mis koosnevad kahest või rohkemast eesnimest ning kahest või rohkemast perenimest. Seega personalijuhi ja IT koordinaatoriga peetud koosolekul lepitati kokku ennetusmeetmed kuni uue tarkvara valmimiseni.

Töö kirjutamise ajal kasutatavas liikmelisuse andmebaasiga ühendatud tarkvaras pole piisavaid kuupäeva väljade selgitusi ega täpset sisestamise juhendit. Selleks, et liikmelisuse staatuse muutumisega seotud kuupäevade lisamine oleks korrektne, lisati tarkvara lähtekoodi täpsemad

selgitused (vt Joonis 23) ning personalijuhi käsiraamatusse lisati täiendavad punktid selle kohta, kuidas toimuvad isikutele kuupäeva sisestused.

Avaldus BESTi	07.09.2016	<input type="checkbox"/>	Täpsustus: Bestiga liitumise kuupäev.
Võeti BESTi liikmeks	09.06.2017	<input type="checkbox"/>	Täpsustus: Beebiliikmeks saamise kuupäev.
Võeti täisliikmeks	pp.kk.aaaa	<input type="checkbox"/>	Täpsustus: Täisliikmeks saamise kuupäev.

Joonis 23. Ennetusmeede – praeguses tarkvaras täpsustavad lausedid kuupäevadele.

Andmesiirdeks koostatud SQL lausete üks puudujääk on mitte toimetulek olukorraga (mida hetkel ei esine), kus isikul on kaks või rohkem eesnime ja kaks või rohkem perenime ning nimest on automaatselt vaja eraldada eesnimed ja perenimed. Kuna tegelik andmesiire toimub alles tulevikus, siis tuleb selliste andmete lisandumise võimalusega arvestada. Liikmeavalduse esitamiseks kasutakse tööriista Google Forms, kus vanas versioonis kirjutati isiku nimi ühte lahtrisse.

Ennetusmeetmena on Google Forms-is nimelahter jagatud kaheks – eesnimi ja perenimi. Selle tulemusena on avalduse täitjal vastutus kirjutada nimi korrektses vormis ning need andmed saab lisada vanasse liikmelisuse andmebaasi. Sellega on olemas andmete allikas, kus nimi on omaniku enda poolt korrektselt osadeks jaotatud ning seda saab kasutada võimalike probleemide lahendamisel.

Sanimeliste isikute probleemi lahendamiseks luuakse vanas liikmelisuse andmebaasis veergude *eesnimi* ja *perenimi* kombinatsioonile UNIQUE kitsendus. See aitab kontrollida, kas sellise nimega isik on juba andmebaasis olemas. Kui on, siis tuleb lisada uue liikme perenime lõppu unikaalne arv. Samasuguse arvuga nime tuleb kasutada ka punktisüsteemis. Kui lõpliku andmesiirde läbiviimisel selgub, et on lisandunud korduvate nimedega inimesi, siis sihtandmebaasis tuleb nimedest need arvud eemaldada.

5.3 Andmete ülekanne

Selles jaotises tuuakse välja andmesiirdel kasutusel olevate SQL lausete näiteid.

5.3.1 Eriala, liikmelisuse seisundi liigi ning isikute andmete ülekanne

Joonisel 24 on eriala nimetuste ja liikmelisuse staatuste ülekanne sihtandmebaasi tabelitesse *Eriala* ja *Liikmelisuse_seisundi_liik*. Sihtandmebaasi tabelitesse lisatakse ka veerud ingliskeelsete tõlgete jaoks. Tabelisse *Liikmelisuse_seisundi_liik* lisatakse kaks uut seisundit, mis on nõuetena välja toodud Stella Leego lõputöös (ülejäänud seisundid olid selles tabelis juba registreeritud).

```

-- liikmelisuse seisundi liigile inglise keelse tõlke lisamine
ALTER TABLE liikmelisuse_seisundi_liik
    ADD COLUMN nimetus_eng          varchar(50) NOT NULL;
--
-- liikmelisuse seisundi liigi ülekanne
INSERT INTO liikmelisuse_seisundi_liik
(
    liikmelisuse_seisundi_liik_kood,
    nimetus,
    nimetus_eng
)
SELECT id,
       nimetus,
       nimetus_en
FROM   d69756_ettevalmistus.staatus;
--
INSERT INTO liikmelisuse_seisundi_liik
(
    liikmelisuse_seisundi_liik_kood,
    nimetus,
    nimetus_eng
)
VALUES
(
    15,
    "Peatatud",
    "Stopped"
),
(
    16,
    "Taotletud",
    "Requested"
);
-- erialale ENG versioon juurde
ALTER TABLE eriala
    ADD COLUMN nimetus_eng          varchar(50) NOT NULL;
-- eriala ülekanne
INSERT INTO d69756_punktidliikmed.eriala
(
    eriala_kood,
    nimetus,
    nimetus_eng
)
SELECT id,
       eriala,
       eriala_en
FROM   d69756_ettevalmistus.eriala;

```

Joonis 24. Eriala ning liikmelisuse seisundi liigi andmete ülekanne.

Joonisel 25 on isikute andmete ülekanne lähteandmebaasi liikmelisuse tabelist sihtandmebaasi tabelisse *Isik*. Ülekantavad andmed on viidud nõuetele vastavatele vormi ettevalmistusalas.

```

INSERT INTO d69756_punktidliikmed.isik
(
    isik_id,
    e_meil,
    eesnimi,
    perenimi,
    hyydnimi,
    telefoni_nr,
    synni_kp,
    reg_aeg,
    viimase_muutmise_aeg,
    isiku_seisundi_liik_kood
)
SELECT id,
    email,
    eesnimi,
    perenimi,
    hyydnimi,
    mobla,
    synniaeg,
    avaldus,
    muudetud,
    1
FROM d69756_ettevalmistus.liikmed;

```

Joonis 25. Isikuandmete ülekanne sihtandmebaasi.

5.3.2 Liikmelisuse andmete ülekanne

Sihtandmebaasis salvestatakse liikmelisust ja sellega seonduvaid sündmusi eraldi ridadena kahes tabelis – ühes tabelis on liikmelisuse andmed ja teises tabelis liikmelisusega seoses toimunud seisundimuudatuste sündmuste andmed. Vanas andmebaasis on seisundite andmeid salvestatud mitmes veerus (iga sündmuse tüübi kohta eraldi veerg). Sihtandmebaasis on iga sündmuse kohta sündmuste tabelis eraldi rida. Selleks, et teisendada vanas andmebaasis olevad seisundite e staatuste andmed veergude kujult ridadeks, viiakse ülekanne läbi mitmes etapis (vt Joonis 27). Esiialgu kantakse üle liikmeavalduse esitamise sündmused, siis beebiliikme seisundi saamise sündmused ning seejärel täisliikme seisundi saamise sündmused. Lisaks salvestatakse vanas andmebaasis veerus *muudetud* kuupäev, millal liige oma viimase seisundi saavutas. Enamikel juhtudel on viimane seisund kas alumni või ebaaktiivne, kuid osadel juhtudel on võimalus saada sealt täiendavaid seisundi kuupäevi.

Vanas liikmelisuse andmebaasis on isikutele määratud id. Seda sama id-d kasutatakse ka uues andmebaasis isikutel isik_id väärtusena, et andmete ülekandmise jaoks oleks seos vana ja uue andmebaasi andmete vahel (vt Joonis 26). Isik id väärtused genereeritakse uues süsteemis

automaatselt (veerul on AUTO_INCREMENT omadus). MariaDBs on see realiseeritud nii, et kui veergu lisada ise väärtus, siis järgmine süsteemi väljastatud väärtus on sellest automaatselt suurem [21]. Seega kasutatud lahendus ei tekita hiljem probleeme uute andmete lisamisel.

```
INSERT INTO d69756_punktidliikmed.liikmelisus
(
    isik_id,
    on_nous_avaldamisega,
    eriala_kood,
    liikmelisuse_seisundi_liik_kood,
    sisseastumise_aasta
)
SELECT id,
       0,
       eriala,
       16,
       astus_kooli
FROM   d69756_ettevalmistus.liikmed;

--
UPDATE d69756_punktidliikmed.liikmelisus
SET    sisseastumise_aasta = NULL
WHERE  sisseastumise_aasta = 0000;
```

Joonis 26. Isikute ülekanne liikmelisuse tabelisse.

```
INSERT INTO d69756_punktidliikmed.liikmelisuse_syndmus
(
    `seisund syndmuse jarel`,
    liikmelisus_id,
    toimumise_aeg,
    reg_aeg
)
SELECT 2,
       a.liikmelisus_id,
       b.t2isliikmeks,
       b.t2isliikmeks
FROM    d69756_ettevalmistus.liikmed AS b
INNER JOIN d69756_punktidliikmed.liikmelisus AS a
ON      b.id = a.isik_id
WHERE   t2isliikmeks !=0000-00-00;
```

Joonis 27. Isikute liikmelisuse kuupäevade ülekanne täisliikme kuupäeva näitel.

5.3.3 Punktide andmete ülekanne

Vanas punktisüsteemis on punktide sisestamiseks kasutatud sama loogikat nagu vanas liikmelisuse tabelis, kus punkti sisestamisel on igale atribuudile (kuupäev, tegevus,

punktisumma) mõeldud oma väli. Kui andmed on sisestatud valesti või andmed hiljem täpsustuvad, siis muudetakse olemasolevat kirjet ning andmete ajalugu ei säili.

Uues andmebaasis on loodud punktide andmete jaoks kaks tabelit – *punkti_saamine* ning *punkti_saamine_syndmus*, kus esimeses tabelis salvestatakse tegevus, valdkond ja punktisumma ning teises tabelis salvestatakse punkti seisundimuutust tinginud sündmustega seotud andmed (ajatemplid, registreerija ning kommentaar). Punktide andmete ülekanne tehti kahes etapis (vt Joonis 28). Ka uues andmebaasis ei säilitata andmemuudatuste ajalugu.

Esimeses etapis kanti üle tegevused, valdkonnad, punktisummad ning ühendades isiku identifikaatori (*isik_id*) punktitableis ning liikmelisuse tabelis, lisati ka liikmelisuse identifikaator. Teise etapina loodi ühendus *punkti_saamine* tabeliga, et kanduks üle õige *punkti_saamine_id* ja toodi üle ettevalmistusala punkti tabelitest teine osa andmeid (ajatemplid, kommentaar).

Vanas punktisüsteemis ei salvestata andmete sisestajaid. Seega registreerija identifikaatoriks on pandud käesoleva töö autori liikmelisuse identifikaator. Põhjuseks on, et kui tulevikus tekib uues tarkvaras küsimusi mõne punkti sisestuse ajaloo kohta, siis on võimalik eeldada, et andmed punktide kohta on uude andmebaasi tulnud andmesiirdest.

```

-- MTG_punktid punktide sisestamine
INSERT INTO d69756_punktidliikmed.punkti_saamine
(
    punkti_id,
    liikmelisus_id,
    tegevus_kood,
    valdkond_kood,
    punkti_saamise_seisundi_liik_kood,
    punktisumma,
    tegevuse_arv
)
SELECT      c.m_id,
            a.liikmelisus_id,
            b.tegevus_kood,
            5,
            1,
            c.punkte_dec,
            c.mitu_dec
FROM        d69756_ettevalmistus.mtg_punktid AS c
INNER JOIN  d69756_punktidliikmed.liikmelisus AS a
ON          a.isik_id=c.isik_id
INNER JOIN  d69756_ettevalmistus.valdkond_tegevus AS b
ON          b.tegevus=c.mida
WHERE      b.valdkond="mtg";
--
-- MTG punkti sündmuse lisamine
INSERT INTO d69756_punktidliikmed.punkti_saamise_syndmus
(
    punkti_saamine_id,
    seisund_syndmuse_jarel,
    registreerija_id,
    toimumise_aeg,
    reg_aeg,
    kommentaar
)
SELECT DISTINCT d.punkti_saamine_id,
                1,
                1094,
                c.aeg,
                c.aeg,
                c.kommentaar
FROM          d69756_ettevalmistus.mtg_punktid AS c
INNER JOIN    d69756_punktidliikmed.liikmelisus AS a
ON           a.isik_id=c.isik_id
INNER JOIN    d69756_ettevalmistus.valdkond_tegevus AS b
ON           b.tegevus=c.mida
INNER JOIN    d69756_punktidliikmed.punkti_saamine AS d
ON           d.liikmelisus_id=a.liikmelisus_id
AND          d.valdkond_kood=b.valdkond_kood
AND          d.tegevus_kood=b.tegevus_kood
AND          d.punktisumma=c.punkte_dec
AND          d.tegevuse_arv=c.mitu_dec
AND          d.punkti_id=c.m_id
WHERE        b.valdkond="mtg";

```

Joonis 28. Punkti andmete ülekanne MTG töögrupi näitel.

5.4 Järgmised tööd

Autor on koostatud laused BEST-Estoniale zip faili kokkupakituna üle andnud. Autor plaanib nende lausete kasutamise ja ajutiste andmekvaliteedi parandamise meetmete kohta läbi viia koolituse. Autor plaanib organisatsioonile tegeliku andmesiirde läbiviimiseks igal juhul appi tulla.

Lõputöö autori hinnangul tuleks järgmisena ette võtta liikmelisuse ning punktide haldamise tarkvara loomine. Ideaalis võiks toimuda mõlema tarkvara loomine samal ajal, et mõlemad rakendused valmiks ühes ajal ja neid saaks hakata üheskoos kasutama. Kui see pole võimalik, siis esmalt on soovitatav luua uus liikmelisuse haldamise rakendus ning siis alles punktide haldamise rakendus, sest punktide andmed sõltuvad liikmelisuse andmetest. Karl Henri Kaljuse lõputöös valminud punktide haldamise kasutajaliides on toimivaks lahenduseks liiga algeline. Üheks võimaluseks on seda täiendada või siis tuleb luua täiesti uus tarkvara vastavalt Karl Henri Kaljuse töös välja toodud nõuetele. Liikmelisuse halduse tarkvara, mis kasutaks uut andmebaasi, ei ole veel üldse loodud. Kui nimetatud tarkvara on valmis, siis tuleks viia läbi käesolevas töös ettevalmistatud andmesiire, loobuda seejärel vanade liikmelisuse ja punktide haldamise rakenduste kasutamisest ning hakata kasutama uusi rakendusi.

5.5 Hinnang tulemusele

Heaks tavaks on tehtud tööle tagasi vaadata ning mõelda sellele, mis läks hästi ning millele oleks võinud rohkem tähelepanu pöörata.

Tulemuse praktiline osa sooritati nii, kuidas see oli algselt planeeritud. Kogu vajalik informatsioon on uude andmebaasi üle kantav ning suure osa tööst saab SQL lausetega automatiseerida. 100% automaatset andmete ülekannet ei ole siiski võimalik koostatud SQL lausetega teha. Veendumaks, et ülekantav andmestik on õige, on vaja autoripoolset järelevalvet ja seega peab ta ka osalema tegelikul andmete ülekandel. Ei saa lõpuni välistada, et siis peab mõnda nendest SQL lausetest muutma, kuna lähteandmed on täienenud.

Ideaalne ennetusmeede mitmeosalistest nimedest tekkivate probleemide lahendamiseks oleks punktisüsteemis punktide sisestuse juures täisnime sisestamise asemel nõuda eraldi eesnime ja

perenime sisestamist. Samuti oleks punktisüsteemis vaja registreerida isiku unikaalne identifikaator, et selle alusel saaks siduda liikmelisuse ja punktide saamise andmeid. Need ennetusmeetmed tähendaks punktisüsteemi olulist täiendamist. Kuna autor teadvustas seda probleemi liiga hilja, siis ta ei jõudnud neid ennetusmeetmeid kasutusele võtta.

Töö koostamisel tuli ilmsiks, et andmete korrastamist oleks võimalik veelgi optimeerida. Optimeerimine mõjutaks vaid üksikuid väärtuseid ning üldpildis tulevast andmete ülekannet see ei mõjutaks. Samas andmekvaliteeti silmas pidades oleks olnud parem protsesse automatiseerida selle asemel, et käsitsi üksikuid ridu muuta (siin all on mõeldud üksikute korduste eemaldamist lähteandmebaasi liikmelisuse tabelist). Paraku halva ajaplaneerimise tõttu jäi see tegemata.

6 Kokkuvõte

Käesoleva töö siht oli toetada üliõpilasorganisatsiooni BEST-Estonia liikmelisuse ja punktisüsteemi uue tarkvara realiseerimist. Töö sisuks oli andmete korrastamine ning nende ülekandmise ettevalmistamine uude MariaDB andmebaasisüsteemi abil loodud andmebaasi. Selleks tuuakse töös välja andmesiirde ja andmekvaliteedi teoreetilised põhimõtted ning selgitatakse valitud lahenduse põhjuseid. Seejärel analüüsitakse olemasolevate andmete kvaliteeti ning luuakse SQL laused andmete üleviimiseks uude loodud andmebaasi. Olemasolevate andmete kvaliteedis leidis mitmeid probleeme, mistõttu ei olnud andmete ülekandmine triviaalne ettevõtmine. Andmete ülekandmiseks koostati kokku 258 lauset ning kogu ettevalmistatud andmesiirde protsessi läbimine võttis umbes 17.5 minutit. Suurima osa sellest ajast (umbes 10 minutit) võttis andmete korrastamiseks mõeldud SQL lausete täitmine.

Töö tulemusena on valminud andmesiirdeks vajalikud SQL laused ning kasutusele võetud ajutised andmekvaliteedi parandamise meetmed kuniks andmete haldamiseks on loodud toimiv tarkvara. Autori hinnangul on töö eesmärgid täidetud.

Tegelik andmete ülekanne uude andmebaasi toimub alles siis, kui on valminud liikmelisuse ja punktide haldamise tarkvara uus versioon, mis kasutab uut andmebaasi, kuhu andmete ülekandmist käesolev töö ette valmistab. Autori hinnangul loob selle töö tulemus eelduse, et ülekanne toimub kiiresti ja probleemivabalt.

Kasutatud kirjandus

- [1] XAMPP [WWW] <https://en.wikipedia.org/wiki/XAMPP> (29.12.2021)
- [2] Kaljus, K.H. Üliõpilasorganisatsiooni BEST-Estonia punktisüsteemi tarkvara uue versiooni arendamine,“ Tallinna Tehnikaülikooli Tarkvarateaduse instituut, 2021 Bakalaureusetöö. [WWW] <https://digikogu.taltech.ee/et/Item/e45719a2-3ab4-43b4-b54c-8407f8abd351> (13.12.2021)
- [3] Leego, S. „Üliõpilasorganisatsiooni BEST-Estonia liikmelisuse arvestuse tarkvara kavandamine,“ Tallinna Tehnikaülikooli Tarkvarateaduse instituut, 2021 Bakalaureusetöö. [WWW] <https://digikogu.taltech.ee/et/Item/f25aa900-7cb2-4c8c-8ddb-f5cf74397155> (13.12.2021)
- [4] „Extract, Transform, Load,“ [WWW] https://www.sas.com/en_us/insights/data-management/what-is-etl.html (13.12.2021)
- [5] Tayi, G.K., Ballaou, D.P. „Examining Data Quality,“ Communications of the ACM [WWW] <https://dl.acm.org/doi/pdf/10.1145/269012.269021> (22.12.2021)
- [6] Azeroual, O., Saake, G., Schallehn, E. „Analyzing data quality issues in research information systems via data profiling,“ International Journal of Information Management, Pages 50-56 [WWW] <https://www.sciencedirect.com/science/article/pii/S0268401218300975> (27.12.2021)
- [7] Kavak, E. „Andmekvaliteedi probleemide lahendamine andmeaidas,“ Tallinna Tehnikaülikooli infotehnoloogia teaduskond, 07.06.2016 Bakalaureusetöö. [WWW] <https://digikogu.taltech.ee/et/Item/57c70e03-5f54-4495-9f3f-892228c60ae6> (30.12.2021)
- [8] Morris, J. „Practical data migration“ [Online] BCS, The Chartered Institute, Pages 14-17 (29.12.2021)
- [9] Eessaar, E. „Andmebaasid II. Teema 4. Andmebaasi füüsiline disain, ehitamine, andmesiire, varundamine“ [WWW] <https://maurus.ttu.ee/download.php?aine=346&document=32766&tyyp=do> (04.01.2021)
- [10] Isikuandmete kaitse seadus [WWW] <https://www.riigiteataja.ee/akt/104012019011>. (30.12.2021)
- [11] Hevo Data [WWW] <https://hevodata.com/> (28.12.2021)
- [12] IRI Nextform [WWW] <https://www.iri.com/products/nextform> (28.12.2021)
- [13] IRI Netform Products [WWW] <https://www.iri.com/products/nextform> (04.01.2022)
- [14] Brackets (text editor) [WWW] [https://en.wikipedia.org/wiki/Brackets_\(text_editor\)](https://en.wikipedia.org/wiki/Brackets_(text_editor)) (29.12.2021)
- [15] Hövel.E. „Mõned mitmekeelsete SQL-andmebaaside disainimustrid“, Tallinna Tehnikaülikooli infotehnoloogia teaduskond, 09.06.2017 Bakalaureusetöö. [WWW] <https://digikogu.taltech.ee/et/Item/3c203fdf-0162-4545-a3d8-fcfa150f5301> (04.01.2022)
- [16] Windows-1252 [WWW] <https://dbpedia.org/page/Windows-1252> (29.12.2021)

- [17] Source lines of code [WWW] https://en.wikipedia.org/wiki/Source_lines_of_code (29.12.2021)
- [18] AIDanial – cloc [WWW] <https://github.com/AIDanial/cloc> (30.12.2021)
- [19] Instant SQL Formatter [WWW] <https://www.dpriver.com/pp/sqlformat.htm> (04.01.2022)
- [20] Unicode Character Sets [WWW] <https://dev.mysql.com/doc/refman/8.0/en/charset-unicode-sets.html> (29.12.2021)
- [21] MariaDB AUTO-INCREMENT [WWW] https://mariadb.com/kb/en/auto_increment/ (04.01.2022)

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Jorgen Ööpik

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

"[Lõputöö pealkiri]" , mille juhendaja on [Juhendaja nimi]

1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh

Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni

autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna

kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni

autoriõiguse kehtivuse tähtaja lõppemiseni.

2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka

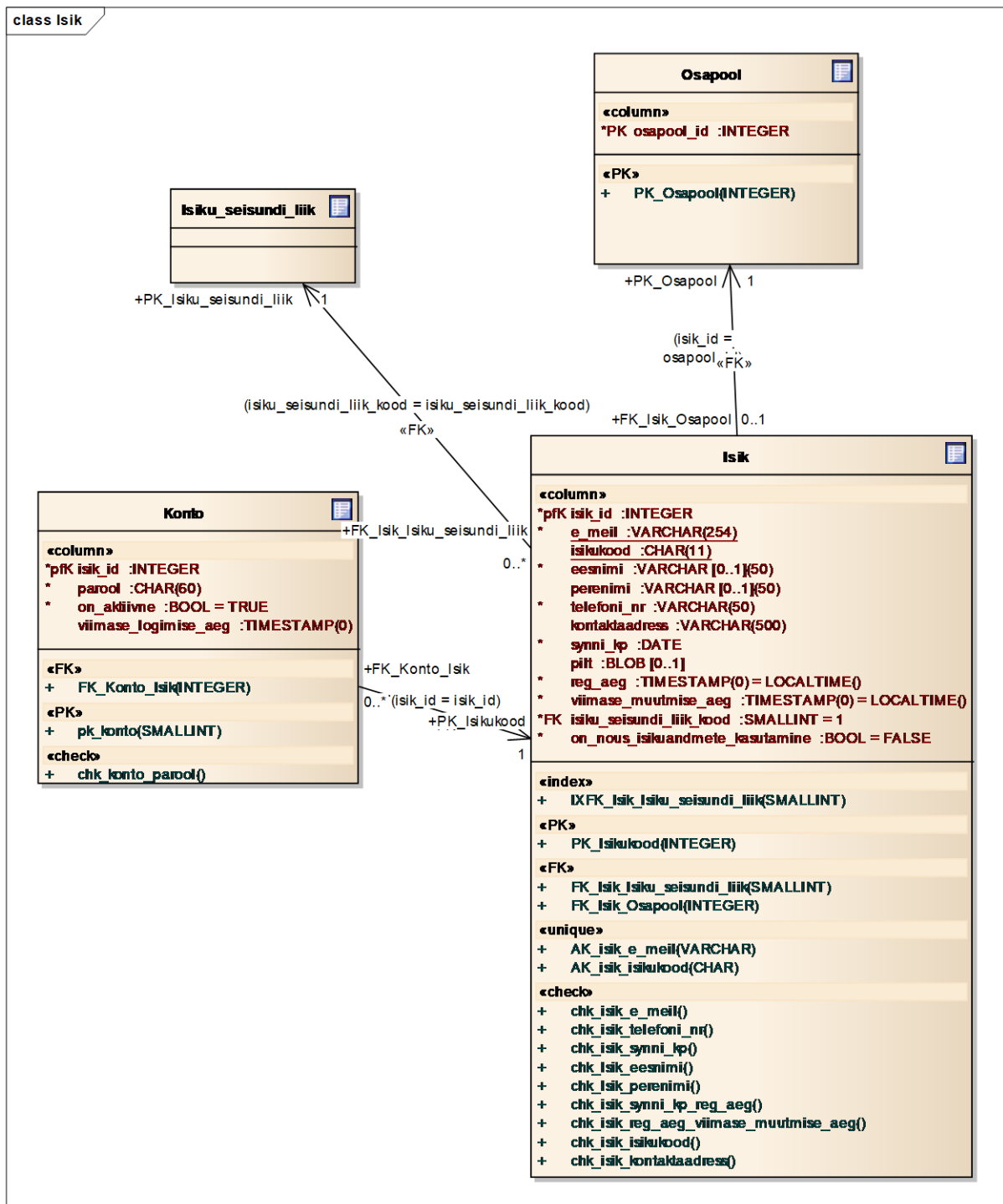
autorile.

3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega

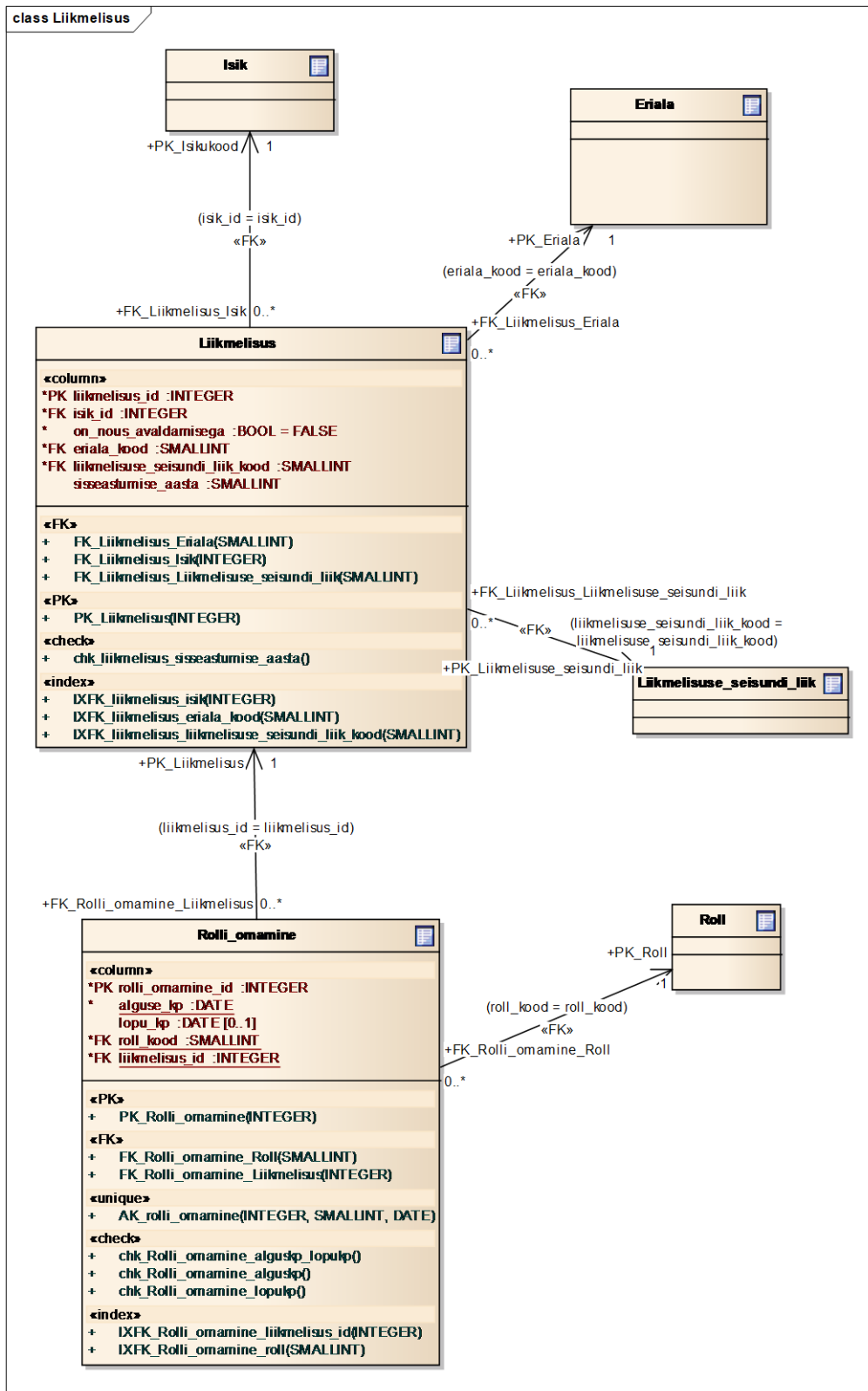
isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi

30.12.2021

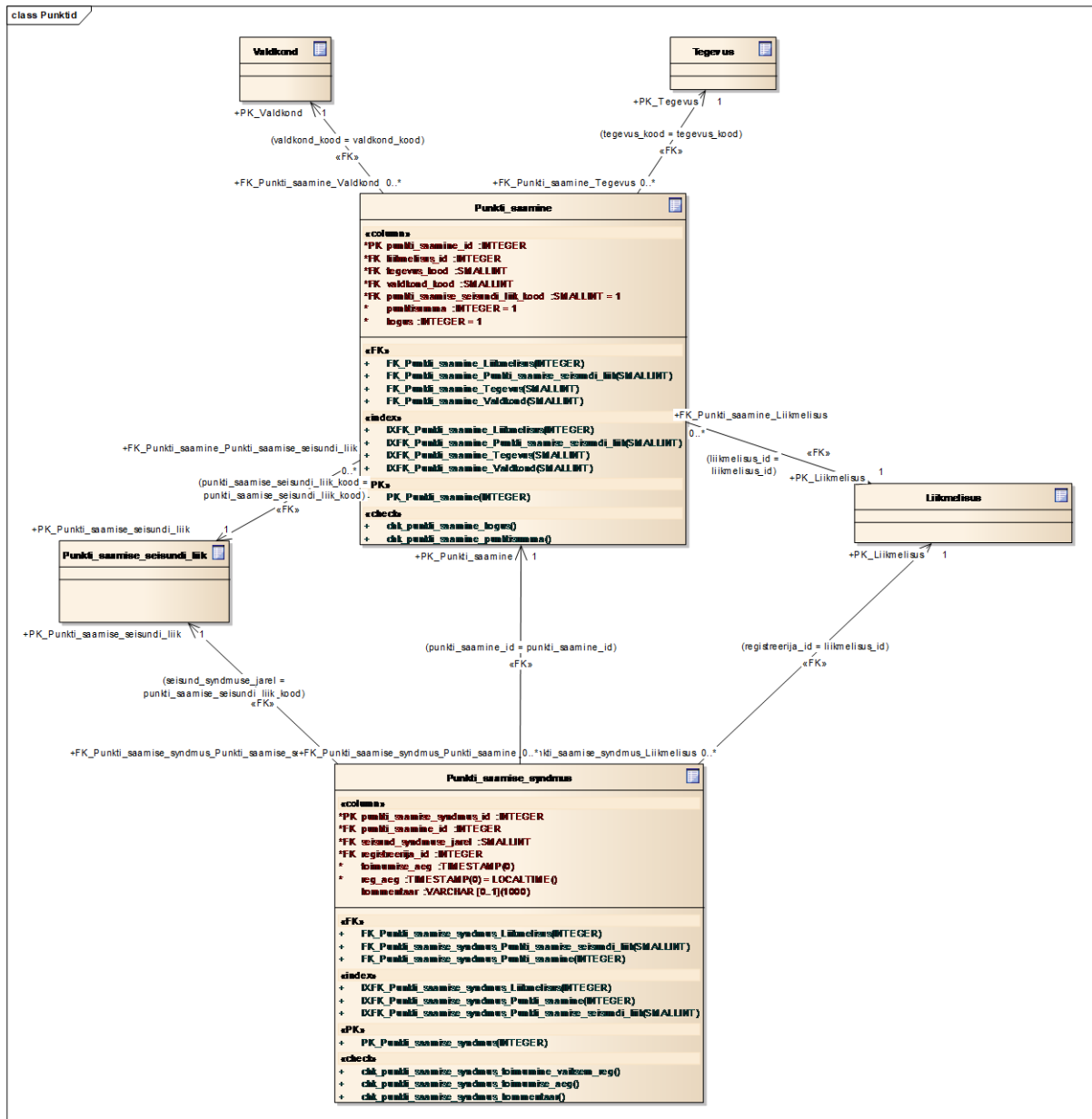
Lisa 2 – Osapoolte ja isikute register BEST-Estonia infosüsteemi uues versioonis



Lisa 3 – Liikmelisuste register BEST-Estonia infosüsteemi uues versioonis



Lisa 4 – Punktide saamise register BEST Estonia infosüsteemi uues versioonis



Lisa 5 – Klassifikaatorite register BEST Estonia infosüsteemi uues versioonis

