

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Raul Velbaum 192823IADB

EKEI pärandüsteemi andmete migreerimine

Bakalaureusetöö

Juhendaja: Priit Rospel
MSc.

Tallinn 2024

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: Raul Velbaum

04.01.2024

Annotatsioon

Lõputöös kirjeldatakse Eesti Kohtuekspertiisi Instituudi (EKEI) andmemigratsiooni, mis on nende uue süsteemi arenduse alamprojekt. Töö annab ülevaate EKEI tegevusest, pärand süsteemist, uuest süsteemist, migratsiooni analüüsist, migratsiooni planeerimisest ja realisatsioonist.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 24 leheküljel, 4 peatükki, 9 joonist, 1 tabelit.

Abstract

EKEI Legacy System Data Migration

The thesis describes the data migration project for the Estonian Forensic Science Institute (EKEI), which aims to replace the legacy system LIMS with a new modular system KEIS2. The text gives an overview of EKEI's organization and activities, the legacy system and the system to which the data is migrated, the project stakeholders and the data migration team. The thesis also presents the analysis, design and implementation of the data migration process, which involves assessing the risks, scope, quality and mapping of the data, choosing the form and tools of the data migration, and performing the data migration in a big bang approach. The text concludes with the results of the data migration project.

The thesis is in Estonian, and contains 24 pages of text, 4 chapters, 9 figures, and 1 table.

Lühendite ja mõistete sõnastik

Andmekvaliteet	Andmete vastavus andmete kasutajate ootustele
EKEI	Eesti Kohtuekspertiisi Instituut
ETL	Andmete väljavõtte teisendamine ja laadimine (<i>Extract Transform Load</i>)
KEIS2	Uus Kohtuekspertiisi Instituudi infosüsteem
LIMS	Kohtuekspertiisi Instituudi pärandüsteem (<i>Laboratory Information Management System</i>)
MVP	Minimaalne töötav toode (<i>Minimum Viable Product</i>)
PL/pgSQL	PostgreSQL andmebaasisüsteemi protseduuriline programmeerimiskeel
RIK	Registrite ja Infosüsteemide Keskus
RIT	Riigi IT Keskus
S3	Amazoni lihtne failiteenus (<i>Amazon Simple Storage Service</i>)

Sisukord

Sissejuhatus	10
1 Olukorra kirjeldus.....	12
1.1 Ülevaade EKEI organisatsioonist ja tegevusvaldkondadest.....	12
1.2 LIMS.....	13
1.3 KEIS2	14
1.4 Projekti osapooled	16
1.5 Andmemigratsiooni meeskond	16
2 Analüüs.....	17
2.1 Riskide hindamine	17
2.2 Migreeritavate andmete skoobi leidmine	18
2.3 Sihtbaaside migreerimise järjekorra leidmine	19
2.4 Mõistete kaardistamine.....	19
2.5 Andmete vastavuse kaardistus.....	21
2.6 Algne andmekvaliteedi hindamine	21
3 Kavand.....	24
3.1 Andmemigratsiooni vormi valik.....	24
3.2 Andmemigratsiooni läbiviimiseks vajalike vahendite valik.....	25
3.3 Andmemigratsiooni protsess	26
3.4 Andmemigratsiooni tulemuste õigsuse kontroll	28
4 Teostus.....	30
4.1 Migratsioonikeskkondade kirjeldus.....	30
4.2 PL/pgSQL skriptid.....	30
4.3 Failide migratsiooni rakendus	31
4.4 Testid	31
4.5 Andmemigratsiooni läbiviimine ja tulemused.....	32
Kokkuvõte	34
Kasutatud kirjandus	35

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks 36

Jooniste loetelu

Joonis 1 EKEI organisatsiooniline ülesehitus [1].....	13
Joonis 2 LIMS arhitektuur.....	14
Joonis 3 KEIS 2 tagarakenduse lihtsustatud arhitektuur	15
Joonis 4 Andmete liikumine andmemigratsioonil andmebaaside vahel.....	19
Joonis 5 EKEI äriinfomudel [4]	20
Joonis 6 Migratsiooniprotsessi ettevalmistamise ja läbiviimise erinevad etapid [5].....	28
Joonis 7 Andmete liikumine arendus-, test- ja migratsioonikeskkondades.....	30
Joonis 8 Failide ja andmete liikumine migratsioonirakenduses	31
Joonis 9 Testpäringu näide	32

Tabelite loetelu

Tabel 1 Enimlevinud migratsioonivormide eelised ja puudused	25
---	----

Sissejuhatus

Eesti Kohtuekspertiisi Instituut on vahetamas kasutatavat pärandüsteemi uue lahenduse vastu. Koos süsteemide vahetusega tuleb tagada, et pärandüsteemi andmed ja dokumendid oleksid kättesaadavad ka uues süsteemis.

Töös käsitletavaks probleemiks on andmete migreerimine pärandüsteemist uude kohtuekspertiisi infosüsteemi. Andmemigratsioon on tegevuste kogum mille käigus kantakse andmed lähtebaasist või baasidest üle sihtbaasi või sihtbaasidesse. Pärandüsteemist migreeritud andmed ja dokumendid peavad olema uues leitavad, vaadeldavad ja muudetavad juhul kui menetluse olek seda lubab.

Töö eesmärk on andmete migratsiooni planeerimise, teostamise ja migratsiooni tulemuste hindamise kirjeldamine ja migratsiooni teostamine. Peale uuele süsteemile üleminekut peab kasutajatel olema võimalik jätkata töid sealt kus need pärandüsteemis pooleli jäid.

Töö autori roll projektis on süsteemide migratsioonieelne analüüs, tegevusplaani koostamine, lahenduste valikud, migratsiooniks vajaliku tarkvara realiseerimine ja migratsiooni tulemuste hindamiseks vajalike testide koostamine. Töö praktiline väljund on migratsiooniks vajalik tarkvara ja migratsiooniprotsessi detailne juhend migratsiooni läbiviijale.

Töö koosneb sissejuhatuses, neljast peatükist ja kokkuvõttest:

Esimeses peatükis antakse ülevaade EKEI organisatsioonist ja tegevusest, kirjeldatakse pärand- ja uue süsteemi ülesehitust ning projektis osalejaid.

Teises peatükis valitakse migratsioonistrateegia, analüüsitakse siht ja lähtebaase, mille tulemusena kaardistatakse olemite ja väljade vastavused, mis on sisendiks andmemigratsiooni läbiviimisel.

Kolmandas peatükis valitakse tehnilised vahendid andmemigratsiooni läbiviimiseks ja planeeritakse migratsiooni edukaks läbiviimiseks vajalikud tegevused.

Neljandas peatükis antakse ülevaade migratsiooniks loodud tarkvaralahendusest ja migratsiooni läbiviimisest.

1 Olukorra kirjeldus

Järgnevalt antakse ülevaade EKEI-st, pärandüsteemist ja süsteemist, kuhu andmed migreeritakse, lisaks tutvustatakse projekti osapooli ja migratsioonimeeskonna koosseisu ning rollide jaotust tiimis.

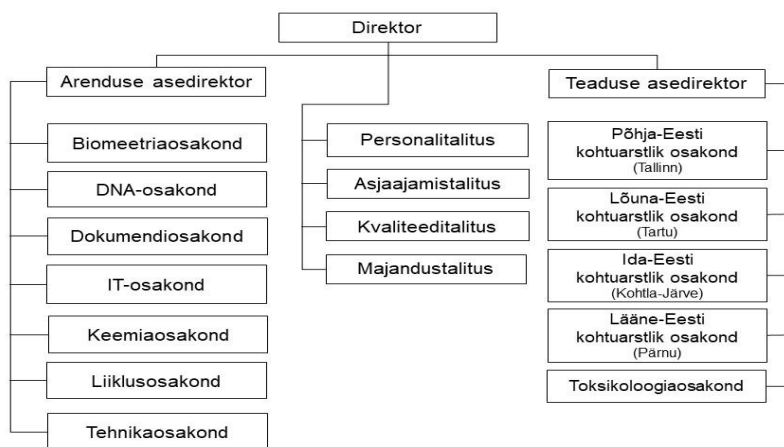
1.1 Ülevaade EKEI organisatsioonist ja tegevusvaldkondadest

„Eesti Kohtuekspertiisi Instituut on riiklik ekspertiisiasutus, mis loodi 01.01.2008. aastal ning on Justiitsministeeriumi hallatav riigiasutus. Instituut on tekkinud kahe asutuse, Eesti Kohtuarstliku Ekspertiisibüroo ning Kohtuekspertiisi ja Kriminallistika Keskuse, ühinemisel.“ [1]

Instituudi põhiline tegevusala on kohtuekspertiiside ja uuringute tegemine kriminaal- ja väärteomenetlustes. Ekspertiisi objektideks on näiteks sündmuspaigalt kogutud asitõendid (sõrmejäljed, DNA proovid, erinevad füüsilised objektid jne). Samuti tegeletakse ekspertiisivaldkondade arendustegevusega, politseiasutuste ekspertiisialase koolitamise ja nõustamisega. „EKEI töö tellijateks on peamiselt kohtu, prokuratuuri, uurimisasutuste ja kohtuvälised menetlejad.“ [1]

„Vastavalt Eesti Kohtuekspertiisi Instituudi põhimäärusele on instituudi üks põhieesmärk kohtuekspertiiside tegemine menetlusasjades. EKEI teostab oma põhieesmärgilist tegevust erapooletult ja asjatundlikult, kõrgel professionaalsel ja tehnilisel tasemel ning majanduslikult efektiivsel viisil, et oleks tagatud töö tulemuste usaldusväärsus ning siseriiklik ja rahvusvaheline tunnustus. EKEI rahvusvahelist tunnustust kinnitab akrediteerimistunnistus. EKEI-s teostatud ekspertiiside tulemused põhinevad teaduslikel alustel.“ [1]

Instituudi struktuuriüksusteks on osakonnad ja talitused. Instituudi organisatsiooniline ülesehitus on kujutatud järgneval joonisel.



Joonis 1 EKEI organisatsiooniline ülesehitus [1]

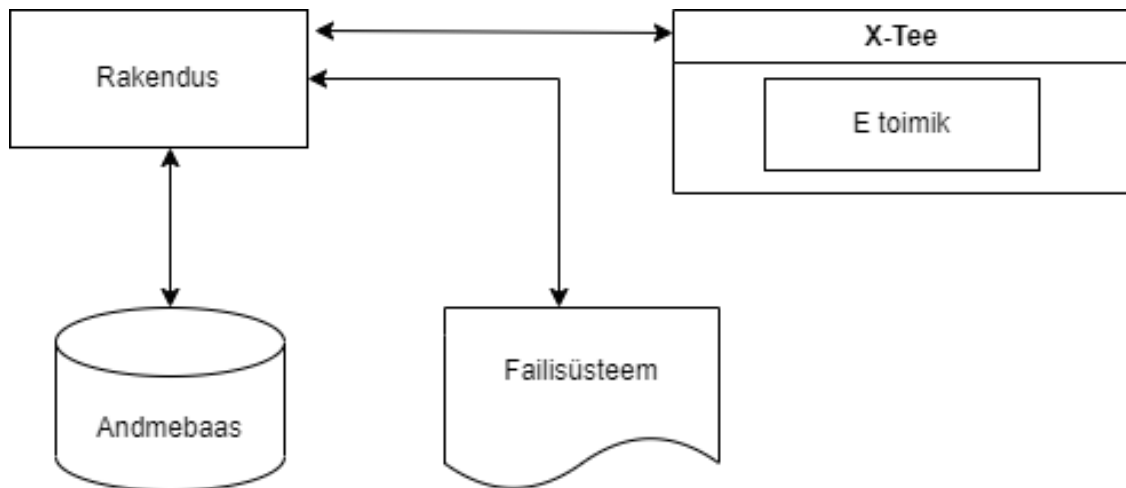
1.2 LIMS

„EKEI võttis olemasoleva kohtuekspertiisi infosüsteemi kasutusele aastal 2011. Tegemist on Soome tarkvaraarendaja loodud tarkvaraga, mille edasiarendamiseks ainulitsents kuulub tarkvara loonud ettevõttele. Sellest tulenevalt sõltub olemasoleva süsteemi edasiarendamise võimekus täielikult ühest arendajast ning viimased aastad on näidanud, et see on väga aeganõudev ning kulukas protsess.“ [2]

Olemasoleva süsteemi funktsionaalsus ei vasta EKEI äriprotsessidele. Välja on arendatud peasjalikult arvepidamist toetav funktsionaalsus, mis võimaldab arhailisel moel jälgida sissetulevaid ja väljaminevaid ekspertiise (dokumendihalduse tasemel). Ekspertiisiaktide koostamine toimub süsteemiväliselt Microsoft Wordi kasutades, mis tähendab et ekspertiisiakti koostamine toimub käsitsi, on veaaltis ja aeganõudev. Puuduvad võimalused akti koostamise käigus taaskasutada varasemates toimingutes kogutud andmeid. [2]

Infovahetus väliste infosüsteemidega on vaevaline ja kasutatava süsteemi tehniline võimekus ei võimalda täisdigitaalse menetluse läbiviimist. Elektroonilise andmevahetuse väljaarendamist takistavad infosüsteemi tehnilised piirangud. [2]

Pärandsüsteem koosneb Java Server Pages rakendusest, mis kasutab andmebaasisüsteemina Oracle 11g'd. Baasis olevate tabelite arv on 743. Dokumendid ja seotud failid on salvestatud nii failisüsteemi kui andmebaasi. Mingil moel on realiseeritud liidestus üle X-Tee E toimikuga. Dokumentatsioonist on olemas süsteemi lõppkasutaja kasutusjuhend. Peale uue süsteemi kasutuselevõttu pärandsüsteemi kasutamine lõpetatakse.



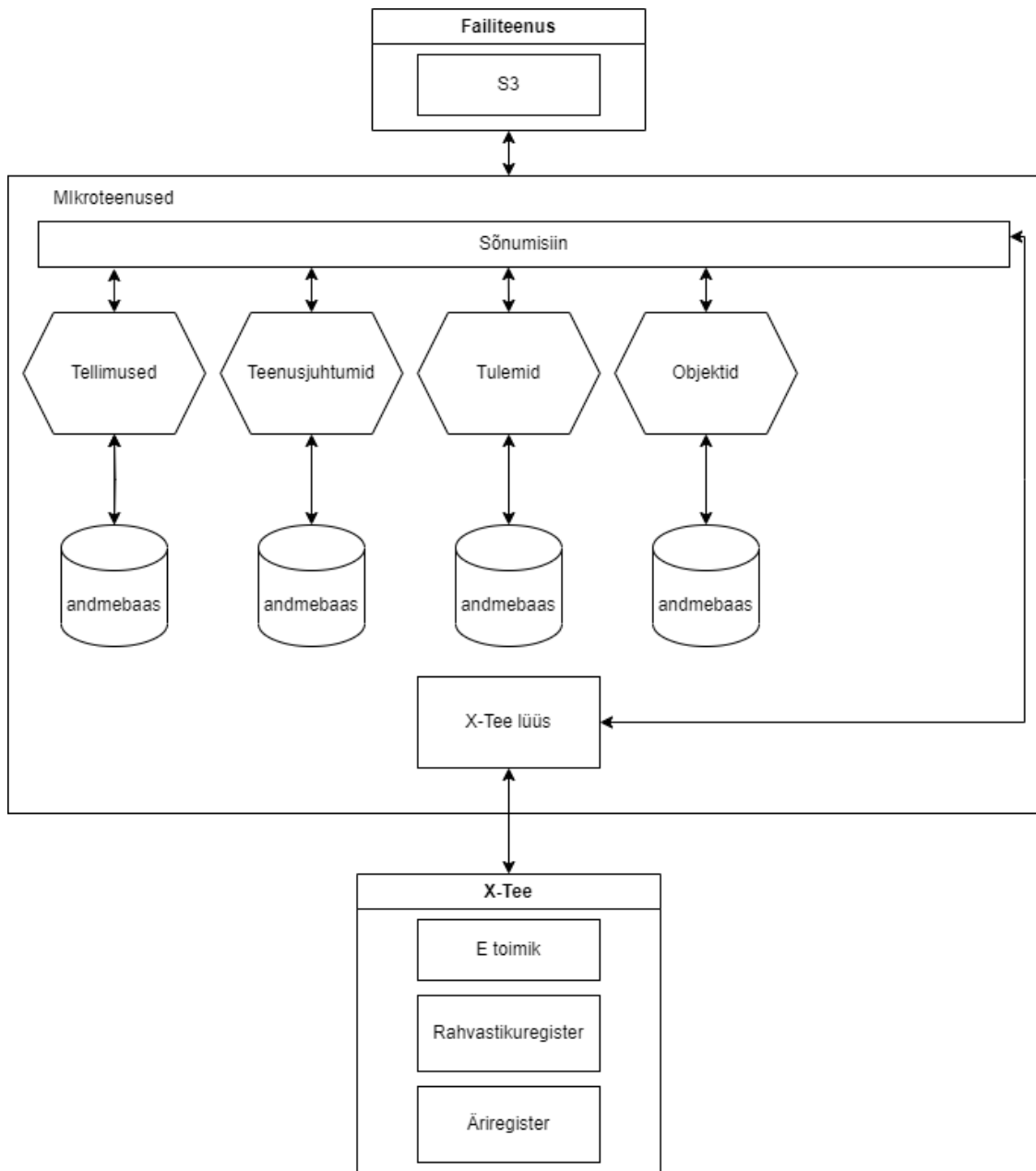
Joonis 2 LIMS arhitektuur

1.3 KEIS2

Arendusprojekti eesmärk on arendada välja EKEI-le uus terviklik infosüsteem, mis aitab digitaliseerida EKEI äriprotsesse, asendada paberipõhised tegevused digitaalsete toimingutega ning vähendada eksperdi ebavajalikke manuaalseid tegevusi ning suurendada kasutaja rahulolu. Projekti raames on plaan arendada välja uus modulaarne süsteem. Projekt on jaotatud etappideks, iga etapi käigus viiakse ellu valitud hulk tegevusi. [2]

Esimeses etapis (MVP) teostati tegevuste kogum, mis katab täielikult pärandsüsteemi poolt pakutava, tellimuste haldamine, ekspertide tööloikude ehk teenusjuhtumite ja asitõendite ehk objektide haldamine, lisaks tulemite (ekspertiisiaktide, uuringutulemuste) koostamine ning haldus. Töös käsitletav andmemigratsioon on samuti arendusprojekti esimese etapi osa.

Uus süsteem on üles ehitatud nullist. Süsteemi tagarakenduse arhitektuuriline lahendus on mikroteenused ja kasutatavaks raamistikuks on .NET core 6. Kuna iga mikroteenus peab olema iseseisvalt tarnitav ja töötav, siis on igal teenusel oma andmebaas. Kasutatavaks andmebaasisüsteemiks on PostgreSQL 15. Teenuste vaheline suhtlus toimub üle sõnumisiini. Dokumendid ja failid säilitatakse S3 hoidlas. Üle X-Tee toimub suhtlus väliste infosüsteemidega: e toimiku, rahvastiku- ja äriregistriga.



Joonis 3 KEIS 2 tagarakenduse lihtsustatud arhitektuur

1.4 Projekti osapooled

Projekti osapooled on EKEI, kes on projekti- ja andmete omanik, RIK, kes vastutab arenduse, andmemigratsiooni ja infosüsteemi halduse eest ja RIT kes haldab tehnilist infrastruktuuri mille peal süsteem jookseb. RIT on ka andmemigratsiooni lõplik läbiviija, kellele tarnitakse migratsiooni läbiviimiseks vajalikud vahendid (tehnilised juhendid ja tarkvara).

1.5 Andmemigratsiooni meeskond

Meeskonda kuuluvad uue süsteemi ärianalüütik, kelle vastutusala on administratiivsed ülesanded ja suhtlemine EKEI-ga, pärand süsteemi haldur kes vastutab andmeparanduste eest pärand süsteemi andmebaasis ja töö autor, kelle roll on analüüs, andmemigratsiooni arendus ja testimine.

2 Analüüs

Järgnevalt kirjeldatakse migratsiooni kavandamisele eelnenud riskide hindamist, migreeritavate andmete skoobi hindamist, baaside migreerimise järjekorra leidmist, mõistete ja andmeväljade kaardistust ning esialgset andmekvaliteedi hindamist.

2.1 Riskide hindamine

Sõltuvalt uuringust, ebaõnnestuvad täielikult või ületavad mitmekordselt algset eelarvet ja ajahinnangut 70 — 90% kõigist andmemigratsiooni projektidest. Pärandsüsteemi andmekvaliteet on väga problemaatiline, dokumentatsioon puudub ja andmemudelite erinevusest tulenevalt võib andmete üle kandmine olla keeruline, lisaks pole võimalik projekti läbiviimisesse kaasata inimesi, kel puudub riigipoolne turvakinnitus.

Risk on definitsiooni järgi ohu realiseerumise tõenäosus ja järgnevalt on autoril välja toodud töö põhilised ohud ning võimalikud lahendused nende vältimiseks.

Esimene probleem, mida tuleb vältida on andmekadu või mõistete muutumine andmemigratsiooni käigus. Vigade leidmiseks kasutatakse testpäringuid.

Teiseks võivad andmed küll üle kanduda, aga vaatamata sellele, et nad võivad kasutajaliideses õiged välja näha, kuid rakenduse funktsionaalsed nõuded ei pruugi sellegipoolest täidetud olla (näiteks mingit kasutuslugu pole nende andmetega võimalik lõpuni viia). Sedalaadi vigade vältimiseks testitakse migreeritud andmeid uue rakendusega ja tehakse läbi võimalikud kasutuslood.

Suhtlus väliste süsteemidega ei pruugi migreeritud andmetega töötada, mistõttu on vajalik võimalusel ka seda testida. Probleemiks on siin andmete konfidentsiaalsus, mistõttu pole võimalik tegelikke andmeid näiteks E Toimiku testkeskkonda üles laadida.

Migratsiooniprotsess peab mahtuma ettenähtud ajavahemikku. Seetõttu tuleb testida täismahus kogu migratsiooni ja veenduda, et ajakulu selleks ei ole suurem kui ette nähtud.

Kuna piiratud arvul inimestel on ligipääs andmetele, siis iga inimese haigus või muu ettearvamatu põhjus võib panna migratsiooniprojekti venima ja sellega nihutada ka KEIS2-e kasutuselevõtu tähtaega. Selle vältimiseks tuleb arendus- ja migratsioonivahendid valida nii, et keegi teine saab vajadusel töö üle võtta.

Vead võivad ilmned peale andmemigratsiooni. Võimalike migratsioonijärgselt väljatulevate probleemide lahendamiseks tuleb säilitada seosed siht- ja lähtetabelite vahel.

Migratsioon võib läbi kukkuda ükskõik milliste asjaolude kokkusattumiste tõttu, selle riski realiseerumise tõenäosus on väga suur (eelpool mainitud 70 - 90 protsenti). Seetõttu tuleb töötada välja varuplaan, mida sel juhul teha.

2.2 Migreeritavate andmete skoobi leidmine

Kõigi andmete migreerimine pärandüsteemist sihtsüsteemi pole enamasti vajalik, sest pärandüsteemi andmed sisaldavad erinevaid logikirjeid, rakenduse parameetreid ja muud, mis ei oma sihtsüsteemi seisukohalt olulist tähendust. Kui neid andmeid peaks peale migratsiooni mingil põhjusel vaja minema, siis on nad vajadusel kättesaadavad arhiveeritud pärandüsteemi baasist.

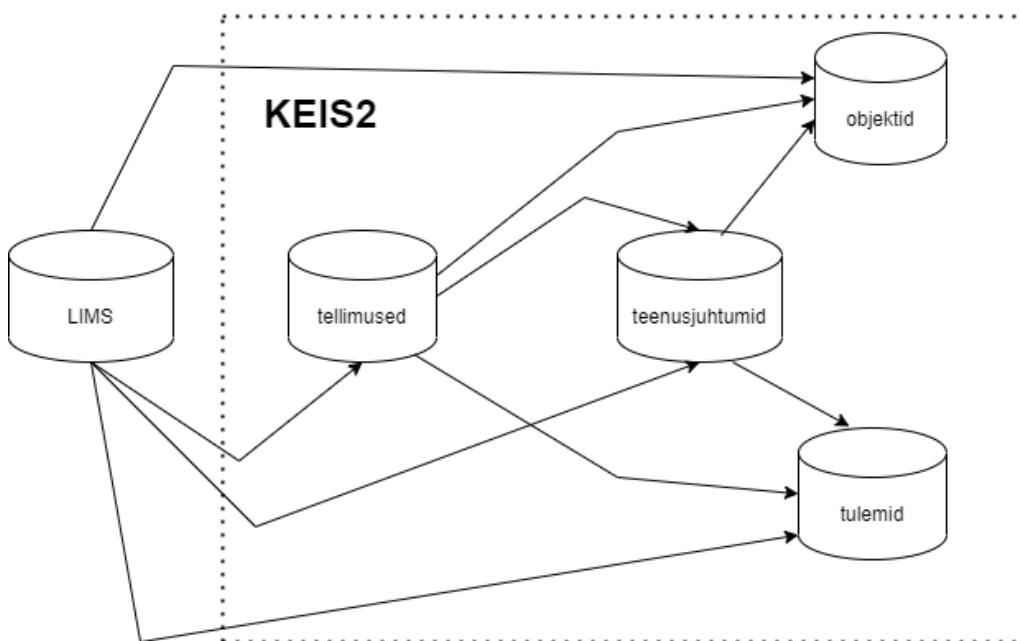
Migreeritavate andmete skoobi leidmiseks viidi läbi analüüs pärandüsteemi baasil, et leida sihtsüsteemi seisukohalt olulised tabelid.

Lähtebaasi tabelite arv on 743, millest tuli välja valida tabelid, mis kannavad sihtsüsteemi kontekstis vajalikku teavet. Kõrvale jäeti kõik tabelid milles on null kirjet, millega vähenes tabelite arv 117 tabelini. Lisaks jäeti välja kõik tabelid mis sisaldasid logisid, ajutisi andmeid ja pärandüsteemi toimimiseks vajalikke abiandmeid. Allesjäänud tabelites olevate andmete sisu järgi otsustati koos pärandüsteemi halduriga kas neis olev info omab uues süsteemis kasulikku teavet. Selle protsessi käigus vähenes

eeldatavalt vajalike tabelite kogus 36 tabelini. Esialgne otsus oli viia kõik dokumendid uude süsteemi üle.

2.3 Sihtbaaside migreerimise järjekorra leidmine

Esimese sammuna vaadatakse andmeid sihtsüsteemi baasides ja leitakse iga teenuse kohta, milliste teiste baaside andmeid ta sisaldab. See määrab, mis järjekorras andmeid migreeritakse. Kui samade olemite andmed esinevad mitmes baasis, siis vaadatakse nende detailsust. Baas, kus olemit kõige detailsemalt kirjeldatakse, saab selle olemit jaoks esmaseks sihtbaasiks. Selle analüüsi tulemusena pannakse kirja baaside migreerimise järjekord, mis on kujutatud järgneval joonisel. Siit on näha, et esmajärjekorras on tellimused, teenusjuhtumid, objektid ja tulemid tuleb migreerida peale neid.



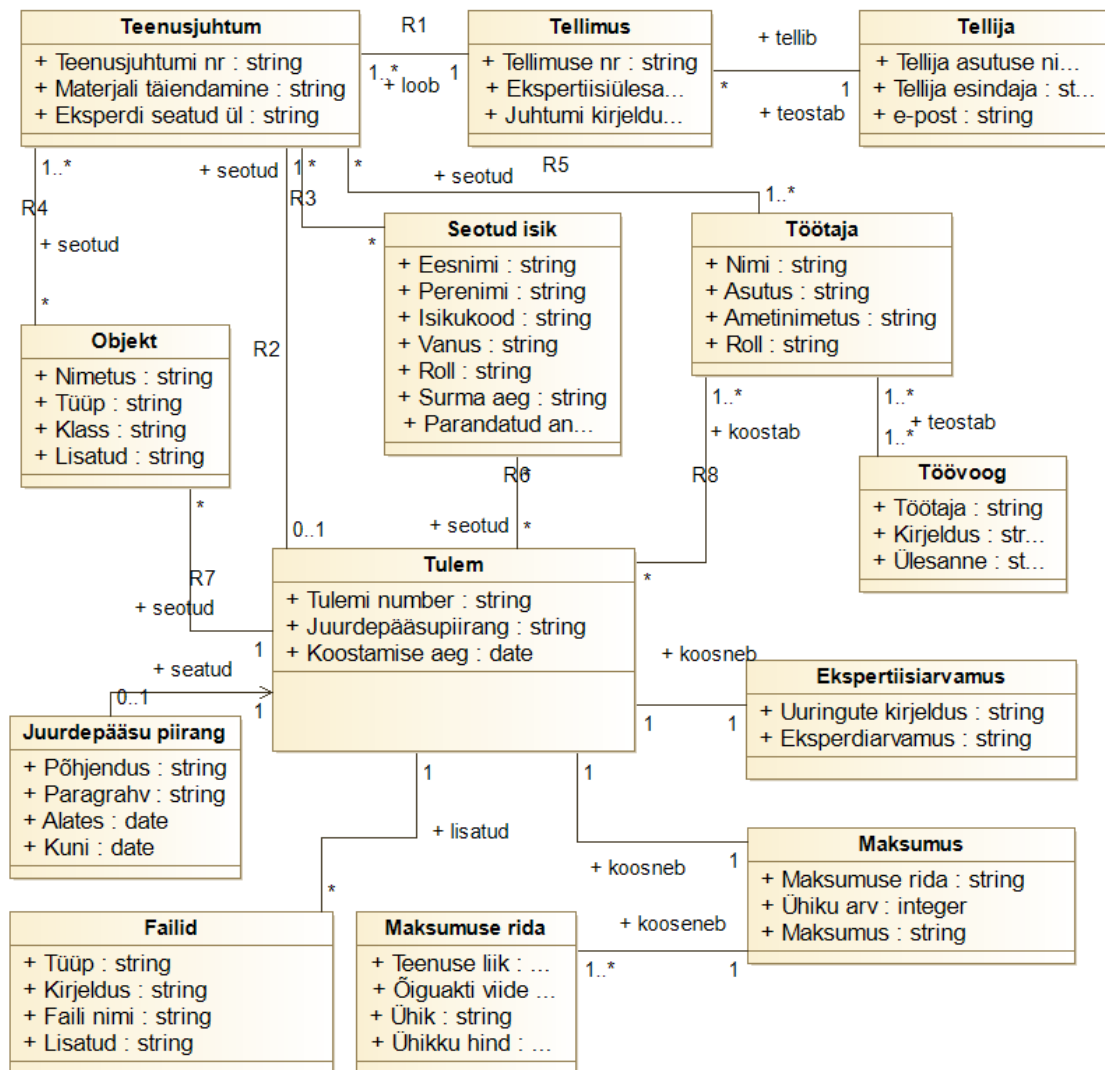
Joonis 4 Andmete liikumine andmemigratsioonis andmebaaside vahel

2.4 Mõistete kaardistamine

Migratsiooni jaoks on oluline omada arusaama kuidas organisatsioon toimib ja mõistete süsteemist, mida organisatsioonis kasutatakse. Esmases lähenduses oleks vaja kirja panna mõisted ja seosed nende vahel. Siinkohal pole mõeldud mitte mõistete süsteemi pärandisüsteemis vaid seda mis mõisted on organisatsioonis tegelikult kasutusel. [3]

Üks moodus selle tegemiseks on mõistete süsteemi kirja panek olemidiagrammina. See olemidiagramm peaks esitama ettevõtet kasutatavate mõistetenä ja seostena nende mõistete vahel. Kuna pärandisüsteemi ja uue süsteemi andmemudelite vahel on detailsuse erinevused, siis on saadud mõistete mudel hea vaheformaat üle mille teha andmete vastavuse kaardistust. [3]

Töö autoril on selleks otstarbeks kasutada kolleegi, Kristo Kleemani loodud EKEI äriinfomudel tema magistritööst.



Joonis 5 EKEI äriinfomudel [4]

2.5 Andmete vastavuse kaardistus

Andmete vastavuse kaardistamisel leitakse sihtbaasi tabelite väljade andmetele vastavad väljad lähtebaasis, andmete teisendusreeglid, navigatsioon ja lisatakse kommentaarid kui leitakse võimalikke probleeme andmeteisendustes. Kõik vastavused ja teisendused peavad saama andmete omaniku heakskiidu. Töö tulemusena luuakse dokument, mis sisaldab andmete vastavusi ja teisenduste kirjeldusi. Sihtbaasi tabelite kohta luuakse vastavuste tabelid, mille veergudeks on:

- Sihtbaasi tabeli välja nimi
- Välja kirjeldus
- Välja andmetüüp
- Välja pikkus
- Kas väli on kohustuslik
- Välja piirangud (unikaalne, peab kuuluma vahemikku jne)
- Navigatsiooni kirjeldus (kuidas jõuda vajalike andmeteni mööda seoseid)
- Teisenduse kirjeldus (viide liikide või tüüpide tabelile, NULL väärtuse asendamine mingi väärtusega jne)
- Lähtebaasi nimi
- Lähtebaasi tabeli nimi
- Lähtebaasi tabeli välja nimi
- Kommentaar (selgitus kus ja miks probleem tekkis ja miks ta on konkreetse migratsiooni puhul oluline, kas see on kriitiline)

Liikide, tüüpide ja staatuste vastavuse kirjeldamiseks leitakse nende vastavused ja pannakse kirja eraldi tabelitesse. See dokument on aluseks andmemigratsiooni päringute kirjutamisele ja seda parandatakse töö käigus pidevalt vastavalt leitud vigadele.

2.6 Algne andmekvaliteedi hindamine

Andmekvaliteedi dimensioonid on õigsus, ajakohasus, täielikkus, reeglipärasus ja ühekordsus. [5]

- „Õigsus (*Accuracy*) näitab, mil määral vastavad andmed tegelikkusele

- Täielikkus (*Completeness*) näitab, mil määral on olemas kõik nõutud andmed“
 - Ajakohasus (*Timeliness*) näitab, mil määral andmete värskus ja kättesaadavus vastab vajadustele ja nõuetele
 - Reeglipärasus (*Orderliness*) näitab, mil määral andmete formaat ja struktuur vastab nõuetele
 - Ühekordsus (*Uniqueness*) näitab, mil määral esineb andmetes duplikaatkirjeid“
- [5]

Ajakohasuse hindamine pole andmemigratsiooni seisukohalt võimalik, eeldus on et andmekogu on ajakohane. Õigsuse hindamiseks peaks olema kasutatav mingi etalonandmekogu, mida töö autoril kasutada ei olnud või reeglistik, mida pärandisüsteemil süsteemiülelel tasemel pole.

Seega dimensioonid mida hinnata on: täielikkus, reeglipärasus ja ühekordsus. Turvakaalutlustel on andmebaasiserver ligipääsetav ainult kaugtöölauda kaudu ja ilma välise internetiühendusega, seetõttu on hinnangute andmiseks võimalik kasutada vaid SQL päringuid.

Järgnevalt on välja toodud kuidas ja mida hinnati. Leitud probleemid dokumenteeritakse ja moodused kuidas need lahenda ja nende prioriteetsus otsustatakse koostöös andmete omanikuga. Andmekvaliteedi esmane hinnang anti dokumentide puudumise protsendi järgi, mis oli üle kolme. Ainuüksi selle parameetri järgi otsustades on olukord andmekvaliteedi osas väga halb. See ei tähenda see, et dokumendid puuduksid, vaid seda, et andmebaasis on suur hulk probleemseid andmeid. Ühe koha väljavõtte järgi saab anda hinnangu ka muude andmete kvaliteedile, sest inimesed, kes need sisestanud on samad.

Sihtsüsteemis unikaalsete väljadele kaardistustulemuste järgi vastavaid välju pärandisüsteemis kontrollitakse unikaalsuse koha pealt. Selleks kasutatakse päringut, mis leiab kirjed, kus selle välja väärtus on korduv.

Tüüpide kaardistuse tabeli järgi leitakse, kas pärandisüsteemi baasis esineb tüüpe, liike, staatuseid jm, mida seal ei peaks olema, samuti uuritakse kas on välju, mille väärtused ei vasta etteantud vormingule.

Kohustuslike sihtbaasi tabeli väljadele vastavatel pärandüsteemi tabeli mittekohustuslike väljade väärtuseid kontrollitakse väärtuse olemasolu järgi.

3 Kavand

Järgnevas tuuakse välja migratsiooniprotsessi planeerimise olulised osad, milleks on andmemigratsiooni ja vahendite valik, migratsiooni arendusprotsessi ja tulemuste hindamise kavandid.

3.1 Andmemigratsiooni vormi valik

Andmemigratsiooniks kasutatakse enamasti kahte meetodit, suure paugu andmemigratsioon (*Big Bang Data Migration*) ja etapiline andmemigratsioon (*Trickle Data Migration, Phased Data Migration*). [6]

Suure paugu lähenemises liigutatakse kõik andmed lähtekeskonnast sihtkeskkonda korraga suhteliselt lühikese aja jooksul. Süsteemid on kasutajatele kättesaadamatud nii kaua, kui andmed liiguvad ja läbivad teisendusi, et vastata sihtsüsteemi nõuetele. Migratsioon viiakse tavaliselt läbi nädalavahetusel või planeeritud katkestusperioodil, kui kasutajad eeldatavasti ei kasuta rakendust. [6]

Suure paugu lähenemisviis võimaldab lõpetada migratsiooni lühima võimaliku ajaga ja säästab vana ja uue süsteemi samaaegsest käiguhoidmisest. [6]

Etapilise andmemigratsiooni puhul on kogu protsess jagatud alammigratsioonideks, millel on oma eesmärgid, ajakavad, ulatus ja kvaliteedikontrollid. Etapilise migratsiooni korral töötavad vana ja uus süsteem paralleelset ja andmete ülekandmist väikeste portsjonitena. Selle tulemusena on rakendus on kättesaadav 24/7. [6]

Selline lähenemine võtab palju rohkem aega ja lisab keerukust. Migratsioonil peab jälgima, millised andmed on juba üle kantud ja tuleb tagada andmete sünkroniseerimine süsteemide vahel. [6]

Järgnevas tabelis on välja toodud mõlema lähenemise eelised ja puudused.

Tabel 1 Enimlevinud migratsioonivormide eelised ja puudused

Migratsiooni vorm	Eelised	Puudused
suure paugu andmemigratsioon	vähem kulukas, vähem keeruline, võtab vähem aega, kõik muudatused toimuvad korraga	suur oht ebaõnnestumiseks, nõuab töökatkestust
etapiline andmemigratsioon	vähem võimalusi ebaõnnestumiseks, ei vaja töökatkestust	kallim, võtab rohkem aega, vajab lisaressursse, et hoida kahte süsteemi samaaegselt töös, tuleb leida viis andmete sünkroniseerimiseks süsteemide vahel

Projekti ressursid (aeg, raha, tööjõud) on piiratud, võimalus andmekatkestuseks on olemas nädalavahetustel kui EKEI-s tööd üldiselt ei toimu ning andmemahud mida üle kanda on suhteliselt väikesed, siis valiti vaatamata tõsiste riskide olemasolule suure paugu andmemigratsioon.

3.2 Andmemigratsiooni läbiviimiseks vajalike vahendite valik

Andmemigratsiooniks vajaliku tarkvara valikul lähtutakse järgmistest tingimustest:

1. Kasutatav tarkvara peab olema vabavaraline
2. Andmed, mida migreeritakse on konfidentsiaalsed, mistõttu andmeid ei saa viia RIT-st välja.
3. Andmemigratsiooni läbiviijaks on RIT-i süsteemiadministraator(id), kes peavad olema võimelised neile antavate tarkvaraliste vahenditega migratsiooni läbi viima juhendi põhjal ilma põhjaliku eeltreeninguta
4. Andmemigratsioon viiakse läbi süsteemi füüsilises asukohas

5. Andmemigratsioon tuleb läbi viia piiratud ettemääratud ajavahemiku jooksul

Esimesed kaks punkti välistavad tasuliste ETL vahendite kasutamise, teine punkt määrab, et veebipõhised tööriistad samuti ei sobi. Vabavaraliste ETL vahendite kasutamise probleemiks on, et migratsiooni läbiviivad administraatorid ei valda neid.

Seetõttu jäävad valikusse migratsioonirakenduse kirjutamine või SQL skriptide kasutamine. Kuna oluline on nii arenduse paindlikkus kui lahenduse töökiirus, siis langes otsus PL/pgSQL kasutamise kasuks.

Failide migreerimise vahendiks on võimalik kasutada vahendeid mille jaoks on olemas ametlikud Amazoni teegid. Need on Python, Java, .NET ja javascript. Valik langes .NET C# konsoolirakendusele, sest .NET on RIK-is soovitatav arendusplatvorm, sel on paralleeltöötuse võimekus ja ka vajalikud S3-e teegid.

Selleks, et kanda andmed üle Oraclest PostgreSQL-i on variantideks andmebaasi skeemi käsitsi üleviimine ja andmete eksport-import läbi mingis formaadis vahefailide (näiteks csv) abil, mis on aeglane, töömahukas ja veaaltis või kasutada mõnda vahendit, millega kogu protsess oleks lihtsam.

Otsustati kasutada vabavaralist rakendust ora2pg, mis loob andmebaasi skeemi skripti ja ekspordib andmed skriptina, mis sisaldab INSERT lauseid ja mille saab importida PostgreSQL-i tekstipõhise kasutajaliidese psql abil. Ora2pg on saadaval Dockerhubist konteinerina, mistõttu on selle kasutamine administraatoritele väga lihtne.

3.3 Andmemigratsiooni protsess

Kuna andmemigratsiooni protsess sõltub samaaegselt toimuvast süsteemi arendusest, siis peab migratsiooni arendus peegeldama süsteemi arendusprotsessi ja metoodikat. KEIS2 arendus toimub agiilselt, SCRUM metoodikat järgides. Sprint kestab kaks nädalat, mille järel valmib uus versioon, mis tähendab, et iga kahe nädala järel sihtsüsteemi baaside skeemid võivad muutuda. Seetõttu viiakse ka andmemigratsiooni arendus läbi iteratiivselt.

Iteratiivne protsess vähendab migratsiooni ebaõnnestumise võimalust. Esiolulised tulemused saadakse migratsiooniarenduse varajases faasis, mis võimaldab alustada nende lahendamise võimalikult vara. Iga järgneva iteratsiooni käigus leitavate küsimuste lahendamise tekitab teadmiste kogum, mida saab kasutada tulevikus arenduse käigus avastatavatele probleemidele lahenduste leidmiseks. [7]

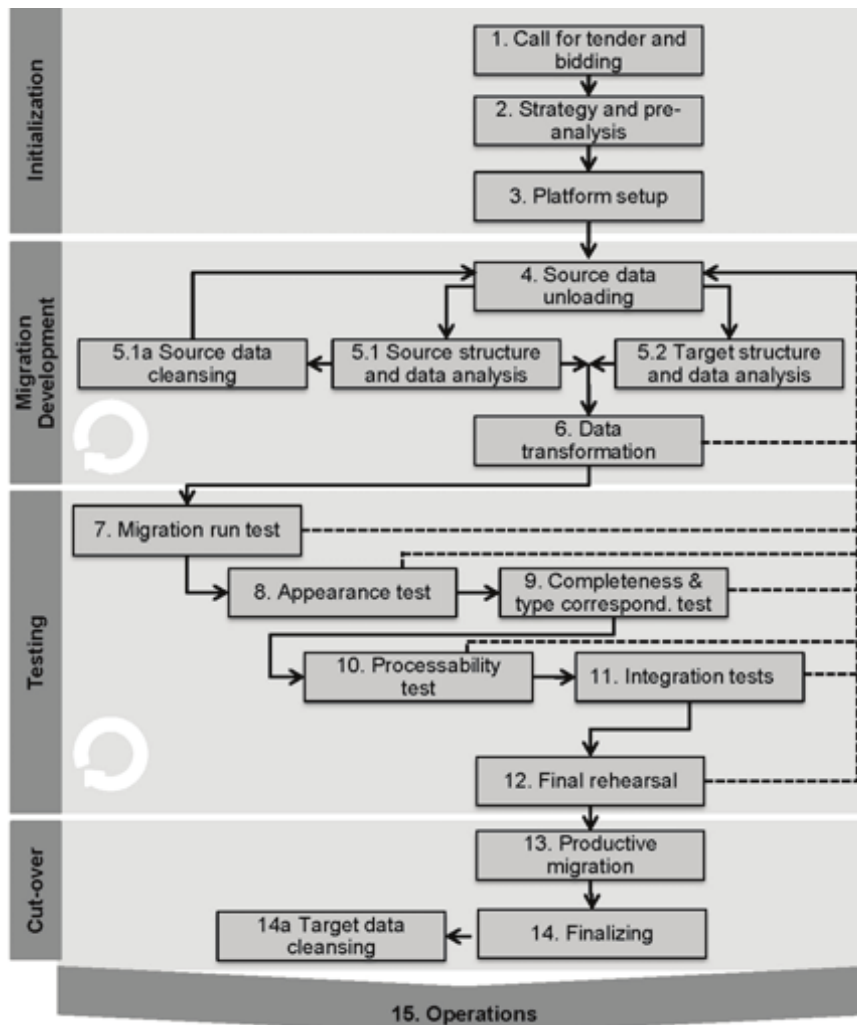
Migratsiooniprotsessi aluseks võetakse töös „Towards the Industrialization of Data Migration“ [8] kirjeldatud protsess, mis koosneb neljast etapist. Need on migratsiooni eeltööd, arendus, testimine ja täideviimine. Joonisel neli on näidatud igas etapis tehtavad sammud ja liikumised nende vahel. Selle töö kontekstis pole oluline eeltööde esimene samm, hanke läbiviimine. Arendus ja testisamme korratakse iteratiivselt senikaua kuni vigade tase jääb alla kokkulepitud piiri.

Eeltööde käigus pannakse paika üldine plaan ja ajakava ning seatakse üles keskkonnad arenduse ning testimise läbi viimiseks.

Arenduses laetakse lähtebaasi andmed arendus ja testkeskkonda, viiakse läbi lähtebaasi ja sihtbaaside analüüs, vajalikud lähteandmete muudatused ja migratsiooniprogrammide kirjutamine.

Testis viiakse läbi testmigratsiooni läbijooks, kontrollitakse andmete esitlust rakenduses, testitakse andmete õigsust ja täielikkust, lisaks testitakse andmete vastavust uue süsteemi nõuetele ja viiakse läbi testid kontrollimaks ühendusi väliste süsteemidega. Viimane samm enne migratsiooni on migratsiooni peaproov, kus simuleeritakse kõiki migratsiooni samme ja vajadusel pannakse kirja, mis läks valesti.

Läbiviimisel seatakse üles tootmiskeskond, viiakse läbi migratsioon (nii andmebaasid kui failid) ja vajadusel tehakse andmetega järeltegevused.



Joonis 6 Migratsiooniprotsessi ettevalmistamise ja läbiviimise erinevad etapid [8]

3.4 Andmemigratsiooni tulemuste õigsuse kontroll

Iga iteratsiooni lõpus sooritatakse valminud skriptidega migratsioonitest. Selle käigus ilmnevad vead, mis enamasti on andmekvaliteedi probleemid, registreeritakse ja otsustatakse koos andmete omanikuga, mis moel need lahendada. Migratsioonitest võimaldab ka hinnata migratsioonile kuluvat aega.

Testkeskkonnas kontrollitakse uuel süsteemil, kas kasutajaliideses kuvatakse andmed õieti, võrreldakse pärandisüsteemi andmeid uue süsteemi andmetega. Vaadeldakse kas andmed on samad, kas vormingud kattuvad ja kas andmete asukoht on õige. Nii kontrollitakse andmete kaardistuse õigsust. Väljatulnud vead parandatakse ja dokumentatsioon uuendatakse.

Uue süsteemi peal kontrollitakse ka migreeritud andmete vastavust infosüsteemi nõuetele. Selleks vaadatakse kas rakenduses esineb migreeritud andmetega vigu ja tehakse andmetega läbi olulised kasutuslood. Tõrgete ilmnemisel vead parandatakse.

Testkeskkonnas kontrollitakse ka kas migreeritud andmetega liidestused väliste süsteemidega toimivad.

Migratsioonitulemuste õigsuse kontrolliks kasutatakse teste. Kontrollitakse kas kõik soovitud kirjed on sihtbaasist lähtebaasi üle kantud ja seda, kas objekti olulised atribuudid on muutmata kujul üle kantud. Testid realiseeritakse SQL päringutega. Testide tulemused salvestatakse edasiseks analüüsiks andmebaasi.

Kirjete arvu võrdlemisega siht- ja lähtetabelis saadakse teada, kas migratsiooniprotsessi käigus on kõik kirjed üle kandunud.

Atribuutide võrdluseks kasutatakse päringuid mis võrdlevad väärtuse olemasolu ning väärtusi siht- ja lähtetabelites. Atribuudid, mida võrrelda valitakse välja koostöös andmete omanikuga.

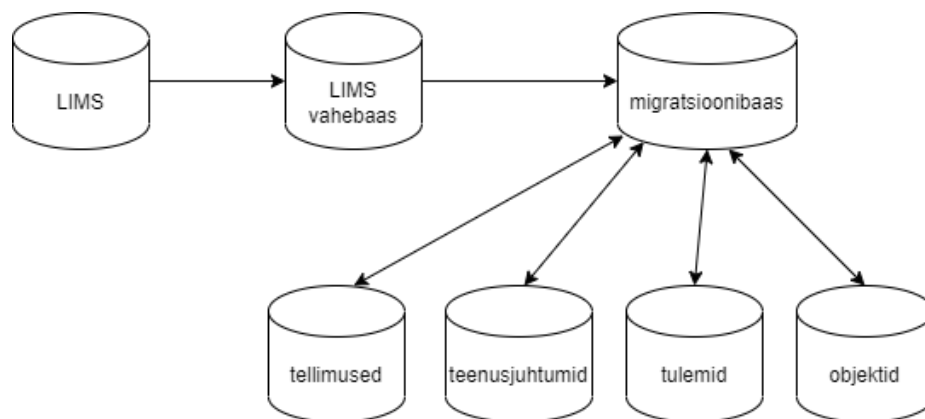
Failide migreerimise testimiseks võrreldakse failide migreerimise rakendusega S3-e üles laetud failide arvu pärandisüsteemi kirjete arvuga. Samuti kontrollitakse valikuliselt kas failid on rakenduses kättesaadavad.

4 Teostus

Selles peatükis kirjeldatakse migratsiooni läbiviimiseks loodud tarkvara ja antakse ülevaade andmemigratsiooni läbiviimisest ning tulemustest.

4.1 Migratsioonikeskkondade kirjeldus

Andmebaasid ja andmete liikumine nende vahel arendus-, test- ja migratsioonikeskkonnas on näidatud joonisel 7. LIMS baasist tehti tõmmis (*dump*), mis taastati LIMS vahebaasis. Ora2pg-ga kanti andmed LIMS vahebaasist migratsioonibaasi. Migratsioonibaasi salvestati lisaks lähteandmetele ka migratsiooni seosekirjed, abiinfo ja testide tulemused. KEIS2 baasid ühendati migratsioonibaasiga PostgreSQL-i välisserveritena (*foreign server*). Migratsiooniskriptid jookсутati migratsiooniserveri vastu, kasutades psql-i.



Joonis 7 Andmete liikumine arendus-, test- ja migratsioonikeskkondades

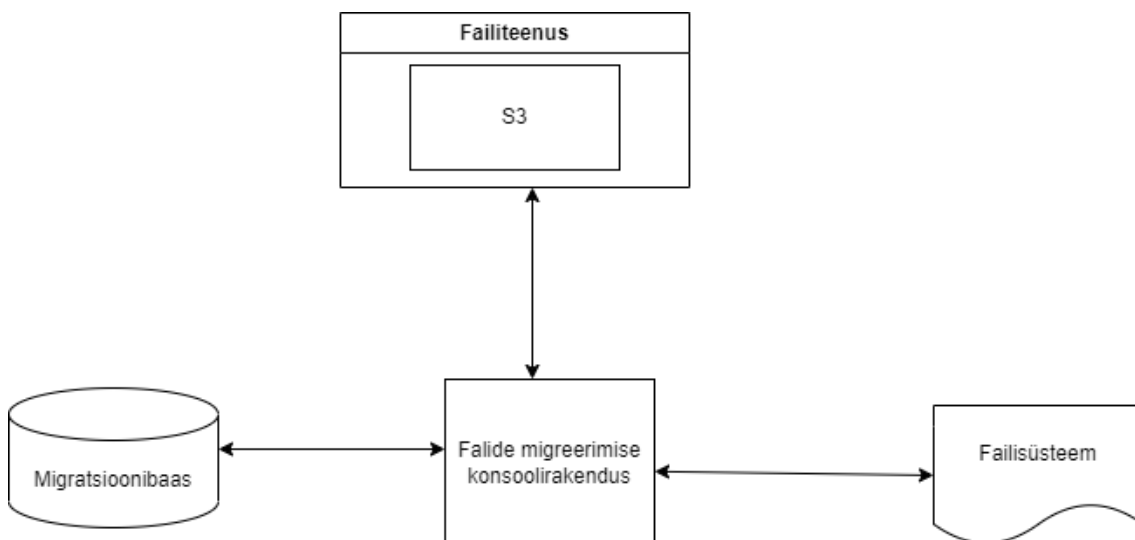
4.2 PL/pgSQL skriptid

Andmebaaside migratsiooniks kirjutati 29 PL/pgSQL skripti. Skriptid kirjutati nii lühikesed kui võimalik, et veasituatsiooni korral oleks lihtne viga üles leida. Skriptid on transaktsioonilased, sest see võimaldab jätkata peale vea parandamist sealt kus protsess vea tõttu seiskus. Kõik andmeteisendused, mis olid kokku lepitud sooritada andmesiirde

ajal teostati PL/pgSQL-i funktsioonide abil. See lähenemine võimaldas andmeteisendused ühiktestidega enne migratsiooni ära testida.

4.3 Failide migratsiooni rakendus

Failide erineva asukoha probleem lahendati andmebaasis asunud failide eksportimisega failisüsteemi. Failide migreerimiseks kirjutati .NET core 6/c# konsoolirakendus. Rakendus loeb migratsioonibaasist välja üleslaetavate failide asukohad, laeb failid S3-e ülesse ja logib tulemused migratsioonibaasi. Töö kiirendamiseks töötab rakendus paralleelselt. Failide üleslaadimine on jaotatud protsessori tuumade vahel nii, et üks tuum tegeleb korraga ühe faili üleslaadimisega. Joonisel 8 on näidatud andmete ja failide liikumine migratsioonirakenduse, migratsioonibaasi, failisüsteemi ja S3-e vahel.



Joonis 8 Failide ja andmete liikumine migratsioonirakenduses

4.4 Testid

Testid loodi samuti PL/pgSQL skriptide abil. Migratsioonibaasi loodi tabelid, et säilitada testitulemused migratsioonitulemuste hindamiseks. Testid viidi läbi analüüsisiosas määratud andmeväljade alusel.

Testid loodi järgmise mustri järgi. Leiti olemist seda iseloomustav unikaalne väärtus. Vana ja uus olem seostati selle võtme järgi FULL OUTER JOINIGA mis annab võimaluse leida erinevad või puuduvad olemid mõlemal pool. Ülejäänud väärtused, mida võrrelda, lisati päringule [8]. Järgnevalt on näidatud kuidas võrreldi juhtumite

numbrite klappivust uues ja pärandüsteemis. Võtmeks mille järgi kahte süsteemi andmeid võrreldi on tellimuse number KEIS2-s ja tellimuse number LIMS-s.

```
SELECT
    uus.tellimuse_nr,
    vana.s_requestid,
    uus.juhtumi_nr,
    vana.u_requestno,
    CASE WHEN
        uus.tellimuse_nr IS NOT NULL AND vana.s_requestid IS NOT NULL
    THEN 'OK'
    ELSE 'NOT OK'
    END olemas,
    CASE WHEN
        uus.juhtumi_nr = vana.u_requestno
    THEN 'OK'
    ELSE 'NOT OK'
    END numbrid
FROM tellimus.tellimus uus
FULL OUTER JOIN sapphire.s_request vana ON vana.s_requestid =
uus.tellimuse_nr
```

Joonis 9 Testpäringu näide

4.5 Andmemigratsiooni läbiviimine ja tulemused

Enne andmemigratsiooni viidi läbi lähteandmete seisuga analüüs LIMS-is . Selleks kasutati arendusel kasutatud andmekvaliteedi teste, kus testiti eelnevalt migratsioonitestide käigus esinenud veasituatsioone. Testide tulemusel, selgus, et baasis on kolm tellimuse kirjet, mille juhtuminumber ja staatus on sama, lisaks tuli välja, et üks kirje tulemite tabelis evib seni esinemata tüüpi. Vead parandati LIMS-is pärandüsteemi halduri poolt.

Migratsioonile eelnenud õhtul suleti kasutajate LIMS-le ligipääs ja tehti baasi hetkeseisust tõmmis migratsiooni läbiviimiseks. Tõmmis taastati vahebaasi ja andmed liigutati sealt migratsioonibaasi.

Migratsioon toimus tõrgeteta, skriptid jooksid läbi seitsme tunni jooksul ja failid laeti RIK-i S3-e kolme tunniga. Ajaliselt mahtus migratsiooniprotsess etteantud ajaakna (12 tundi) sisse ja jäi piisavalt aega üle ka ettenähtud järeltegevuste sooritamiseks.

Peale migratsiooni viidi läbi kontroll andmete ülekande terviklusele, andmete omanikuga kokku lepitud reeglite alusel. Testid kõrvalekaldeid kokkulepetest ei leidnud.

Lõppkasutaja testid näitasid mõningaid kõrvalekaldeid kasutajaliideses. Enamasti olid need situatsioonid, kus pärandüsteemi valideerimata tekstiväljadele oli sisestatud erinevat teavet, mis omas mõtet ainult konkreetse kasutaja jaoks pärandüsteemi kasutajaliidese kontekstis. Need tulemused edastati EKEI-le koos selgitustega.

Kasutuslugude valikuline test kõrvalekaldeid ei tuvastanud nagu ka integratsioonitest väliste süsteemidega.

Alati on võimalik, et mingid probleemid ilmnevad alles mõne aja pärast. Seetõttu otsustati jätta migratsioonikeskkond mõneks ajaks üles, et oleks vajadusel võimalik väljatulevaid vigu parandada. Migratsioonibaasist ja pärandüsteemi baasist tehti arhiivikoopia. Migratsioonibaasi säilitamine on vajalik tulevikus ilmnevate vigade parandamiseks kui see vajalikuks peaks osutuma.

LIMS asendati kasutajate jaoks KEIS2-ga ja EKEI töötajad jätkasid oma tööd uue süsteemiga.

Kokkuvõte

Seoses Eesti Kohtuekspertiisi Instituudi uue süsteemi kasutuselevõtuga tekkis vajadus andmete ja failide ülekandmiseks pärand süsteemist uude süsteemi. Töö eesmärk oli andmemigratsiooni teostamine uue süsteemi juurutamise võimaldamiseks.

Esiteks loodi ülevaade organisatsiooni tegevusaladest, pärand süsteemi olukorrast ja uuest süsteemist. Panti paika projekti osapooled ja migratsioonitiim koos vastutusaladega.

Seejärel viidi läbi analüüs, kus hinnati seonduvaid riske, määrati migreerimisele minevate andmete skoop, leiti andmebaaside migreerimise järjekord, kaardistati ettevõttes kasutusel olevad mõisted ja loodi tabelite ja andmeväljade vastavusdokument. Lisaks viidi läbi esialgne andmekvaliteedi hindamine pärand süsteemis.

Kavandifaasis valiti andmemigratsiooni läbiviimise vorm ja tööriistad migratsiooni läbiviimiseks. Samuti panti paika migratsiooni arenduse ja läbiviimise protsess ning viis kuidas hinnata migratsiooni tulemusi.

Viimasena vaadeldi teostust, andmete ja failide üleviimist lähtesüsteemist sihtsüsteemi, tulemuste hindamist ja migratsiooni tulemusi.

Lõplik migratsioon õnnestus, andmed ja failid migreeriti. EKEI töötajad said jätkata ekspertiiside ja uuringute tegemist uue süsteemiga, mis on pärand süsteemist parema kasutajakogemuse ja funktsionaalsusega.

Kasutatud kirjandus

- [1] „Eesti Kohtuekspertiisi Instituudi koduleht,“ [Võrgumaterjal]. Available: <https://www.ekei.ee/et>.
- [2] Registrate ja Infosüsteemide Keskus, „KEIS2 Lähteülesanne,“ 2019. [Võrgumaterjal]. Available: <https://riigihanked.riik.ee/rhr-web/#/procurement/1573297/documents/source-document?group=B&documentOldId=13221941>.
- [3] J. Morris, Practical Data Migration, 3rd Edition, BCS, The Chartered Institute for IT, 2020.
- [4] K. Kleemann, „Kohtuekspertiisi infosüsteemi tulemite koostamise mooduli loomine,“ 06 03 2022. [Võrgumaterjal]. Available: <https://digikogu.taltech.ee/et/item/ffc8ceaf-4924-4d99-9f86-092d0158c9f6>.
- [5] Eesti andmehalduse metoodikaprojekt, „Andmekvaliteedi juhised,“ 2020. [Võrgumaterjal]. Available: https://www.stat.ee/sites/default/files/2022-03/Andmekvaliteedi%20juhised_2020.pdf.
- [6] Oracle, „Successful Data Migration,“ [Võrgumaterjal]. Available: <https://www.oracle.com/docs/tech/middleware/successful-data-migration-technical-brief.pdf>.
- [7] C. S. K. H. Florian Matthes, „Testing & Quality Assurance in Data Migration,“ 2011. [Võrgumaterjal]. Available: http://www.klaushaller.net/media/icsm11_testing_data_migration.pdf.
- [8] K. Haller, „Towards the Industrialization of Data Migration,“ 2009. [Võrgumaterjal]. Available: http://www.klaushaller.net/media/haller_datamigration_caise2009.pdf.
- [9] D. M. D. D. M. G. Kevin Kline, Pro Database Migration to Azure: Data Modernization for the Enterprise, Apress, 2022.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Raul Velbaum

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „EKEI pärandsüsteemi andmete migreerimine“, mille juhendaja on Priit Rospel
 - 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.

04.012024

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.