

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Allan Bernard 192181IAAM

GIT põhise avaliku sektori sisuhaldussüsteemi arhitektuuri innovatsioon

Magistritöö

Juhendaja: Alari Krist
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: Allan Bernard

16.05.2024

Annotatsioon

Käesolev magistritöö tegeleb avaliku teabe halduse infosüsteemi kirjeldamise, analüüsi ja parendamisega.

Töö eesmärk on kirjeldada avaliku teabe halduse infosüsteemi hetkeseisu, analüüsida selle tehnoloogilist sobivust ning pakkuda välja lahenduse eskiis mis tagaks parema vastavuse kehtivate normide ning strateegiliste eesmärkidega.

Magistritöö koostamisel lähtus autor riiklikest digitaalvaldkonna strateegiatest, mittefunktsionaalsetest nõuetest ning digiriigi tulevikusuundadest. Seatud eesmärgi saavutamiseks, kasutab autor põhiliselt disainmõtlemise ja domeenjuhitud disaini põhimõtteid.

Magistritöö aitab teadvustada avaliku teabe ajakohasena hoidmise vajalikkust ning selle tulemeid saab kasutada sisendina detailanalüüsi hankeprotsessis, mis oleks järgmine samm, magistritöös kirjeldatud lahenduste tellimisel. Täiendavalt annavad analüüsi tulemusena tehtud järeldused sisendit uuritud infosüsteemi omanikule, kasutusel olevate tehnoloogiate jätkusuutlikkuse, võimalike alternatiivide ja innovatsioonivõimaluste kohta.

Lõputöö on kirjutatud Eesti keeles ning sisaldab teksti 49 leheküljel, 5 peatükki, 14 joonist, 3 tabelit.

Abstract

Innovation of GIT-based Public Sector Content Management System Architecture

This master's thesis deals with the description, analysis, and improvement of a public information management system.

The aim of the work is to describe the current state of the public information management system, analyze its technological suitability, and propose a solution that ensures better compliance with existing norms and strategic goals.

The problem addressed by the thesis is due to the fragmentation of state public information and the lack of full utilization of automation capabilities in information updating. This makes public information management more complex, causes non-compliance with existing norms, and negatively impacts the achievement of the state's strategic objectives.

In compiling the thesis, the author adhered to national digital strategies, non-functional requirements, and future directions for the digital state. To achieve the set goal, the author primarily uses the principles of design thinking and domain-driven design.

The master's thesis helps to raise awareness of the need to keep public information up-to-date, and its results can be used as input for the detailed analysis procurement process, which would be the next step in ordering the solutions described in the thesis. Additionally, the conclusions drawn from the analysis provide input to the owner of the examined information system regarding the sustainability of the technologies in use, possible alternatives, and innovation opportunities.

The thesis is in Estonian and contains 49 pages of text, 5 chapters, 14 figures, 3 tables.

Lühendite ja mõistete sõnastik

RIA	Riigi Infosüsteemi Amet
SWOT	<i>Strenghts, Weakensses, Opportunities, Threats (i.k.)</i> analüüsimetod mis seisneb lahenduse tugevuste, nõrkuste, ohtude ja võimaluste sõnastamises.
AvTS	Avaliku Teabe Seadus
EESTI 2035	Riigi pikaajaline arengustrateegia „Eesti 2035“
DÜAK 2030	Eesti digiühiskonna arengukava 2030
MKM	Majandus- ja Kommunikatsiooniministeerium
RIA 21-25	Riigi Infosüsteemi Ameti Strateegia 2021 – 2025
ARVA	Artiklivaramu
TOGAF	<i>The Open Group Architecture Framework (i.k.)</i> levinud arhitektuurikarkass nii äri- kui tehnilise arhitektuuri kirjeldamiseks.
ArchiMate	Avatud lähtekoodiga arhitektuuri modelleerimiskeel mis toetab arhitektuuri kirjeldamist, visualiseerimist ja analüüsimist.
BPMN	<i>Business Process Model and Notation (i.k.)</i> levinud notatsioonikeel äriprotsesside kirjeldamiseks.
C4 mudel	Tõhus graafiline notatsioonitehnika tarkvarasüsteemide arhitektuuri kirjeldamiseks.
UML	<i>Unified Modeling Language (i.k.)</i> universaalne modelleerimise keel süsteemide disaini kirjeldamiseks.
SDG	<i>Single Digital Gateway (i.k.)</i> ehk Ühtse Digivärava määrus [1]. EL regulatsioon nr. 2018/1724 mis etableerub ühtse digivärava ligipääsuks teabele, protseduuridele ja abiteenustele.
TARA	RIA poolt pakutav riiklik autentimisteenus millega asutus saab oma e-teenuses kasutajaid autentida ID-kaardi, mobiil-ID, smart-ID ja Euroopa Liidu eID abil.
GovSSO	Riigi SSO (<i>Single Sign-On (i.k.)</i>) on RIA poolt pakutav teenus, millega asutused saavad oma e-teenustesse lisada nii siseriiklike kui Euroopa Liidu piiriüleste autentimismeetodite toe koos seansihaldusega.
CMS	<i>Content Management System (i.k.)</i> ehk Sisuhaldussüsteem
AS-IS	Süsteemi hetkeseis või hetkel evitatud lahendus

TO-BE	Süsteemi soovitatav seis tulevikus või tulevikus soovitatav lahendus
GraphQL	Avatud lähtekoodiga andmepäringu ja andmete manipuleerimise keel mis võimaldab deklaratiivsete API otspunktide genereerimist.
SHS	vt. CMS.
DMS	<i>Document Management System (i.k.)</i> ehk Dokumendihaldussüsteem
ACL	<i>Access-Control List (i.k.)</i> privileegide loend mille abil on defineeritud vastava ressursi kasutamine.
IAM	<i>Identity and Access Management (i.k.)</i> ehk kasutajate autentimise ja autoriseerimise süsteem.
AD	<i>Active Directory (i.k.)</i> on Microsofti poolt arendatud teenus mis lubab süsteemiadministraatoril hallata kasutajaõigusi ning ligipääse võrguressurssidele.
IP ACL	<i>IP Access Control List (i.k.)</i> reeglite valim mille abil kontrollitakse võrguliiklust ja jõustatakse turvapoliitika IP pakettide lubamisel või keelamisel kahe IP aadressi vahel.
OAuth2	<i>Open Authorization (i.k.)</i> karkassi versioon 2.0 on enamlevinud standard mille abil saab veebiressurs kasutaja autentida, et taotleda ligipääsu teisele veebiressursile või rakendusele, kasutaja nimel.
OIDC	<i>OpenID Connect (i.k.)</i> on OAuth2 karkassi edasiarendus kus lisaks kasutaja autentimisele saab kasutaja ka autoriseerida.
DXP	<i>Digital Experience Platform (i.k.)</i> integreeritud tööriistade ja tehnoloogiate kogum mis lubab organisatsioonidel luua, hallata ja optimeerida sisu, funktsionaalsust ning kasutajakogemust üle erinevate platvormide.

Sisukord

Jooniste loetelu	9
Tabelite loetelu	10
Sissejuhatus	11
1 Magistritöö eesmärk.....	13
1.1 Valdkonna ülevaade	13
1.2 Probleemi püstitus.....	13
1.3 Magistritöö aktuaalsus.....	15
1.4 Magistritöö eesmärk, kasutatud metoodika ja eeldatav tulem.....	19
1.5 Magistritöö käsitusala.....	20
1.6 Autori roll	21
2 Organisatsiooni ärianalüüs	22
2.1 Huvitatud osapooled.....	22
2.2 Strateegia- ja motivatsioonimudel.....	24
3 Taust.....	26
3.1 ARVA tutvustus.....	26
3.2 ARVA tehniline ülevaade (AS-IS).....	27
3.3 Artiklite uuendamise protsess (AS-IS).....	30
4 Süsteemianalüüs.....	32
4.1 Wiki.js	32
4.1.1 Levik.....	32
4.1.2 Jätkuarendused ja tootetugi.....	32
4.1.3 Järeldus	33
4.2 Wikide/CMS'ide kasutamise sobivus	35
4.2.1 Fookus on ebavajalikel funktsionaalsustel.....	35
4.2.2 Jäigad andmestruktuurid.....	36
4.2.3 Piiratud rollide ja õiguste haldus.....	36
4.2.4 Semantilised piirangud	37
4.2.5 Järeldus	37

4.3 GIT	37
4.3.1 GIT puudused ARVA kontekstis	38
4.3.2 GIT funktsioonid üle x-tee.....	39
4.3.3 Järeldus	39
4.4 Duplikaatide tuvastamine kasutades plagiaadituvastustarkvara	40
4.4.1 Funktsionaalne sobivus.....	40
4.4.2 Mittefunktsionaalne sobivus	40
4.4.3 Järeldus	41
4.5 Hinnang ARVA hetkeseisule.....	41
4.5.1 Järeldus	44
5 Lahendus	45
5.1 ARVA tehniline ülevaade (TO-BE).....	46
5.2 ARVA CORE.....	48
5.2.1 Teavituste edastamine ja tagasiside küsimine X-Tee või IP ACL lüüsi vahendusel	50
5.3 ARVA CMS.....	51
5.4 Artiklite uuendamise protsess (TO-BE)	54
5.5 Duplikaatide automaatkontroll.....	57
Kokkuvõte	59
Kasutatud kirjandus	61
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	64
Lisa 2 – Artiklivaramu olemasoleva lahenduse SWOT analüüs	65
Lisa 3 – Töö tulemuste kokkuvõte.....	67

Jooniste loetelu

Joonis 1. Avaliku teabe haldamise protsessi parendamisest saadav üldine väärtus (autori koostatud).....	18
Joonis 2. Avaliku teabe haldamise protsessi muudatustega mitte tegelemisest tekkiv üldine kahju (autori koostatud).....	18
Joonis 3. Organisatsiooni hierarhia ja Artiklivaramu infosüsteemi seos (autori koostatud).....	22
Joonis 4. Huvigruppide skeem (autori koostatud).....	23
Joonis 5. Organisatsiooni motivatsiooni- ja eesmärkmudel ning seos ressursidega (autori koostatud).....	25
Joonis 6. ARVA liidestuste komponentjoonis, AS-IS, C1 tasemel (autori koostatud)...	27
Joonis 7. Artiklite uuendamise AS-IS protsess ülevaatlikult (autori koostatud).....	31
Joonis 8. ARVA rakenduse TO-BE kontseptvaade (autori koostatud).....	46
Joonis 9. ARVA CORE liidestuste komponentjoonis, TO-BE, C1 tasemel (autori koostatud).....	48
Joonis 10. Pakutava lahenduse sisuhalduskomponendi joonis C1 tasemel (autori koostatud).....	52
Joonis 11. Artikli uuendamise protsess ülevaatlikult, TO-BE (autori koostatud).....	55
Joonis 12. Artikli uuendamise protsess teabevaldaja vaatest, TO-BE (autori koostatud).....	56
Joonis 13. Info muutmise TO-BE protsess kolmanda osapoole vaatest (autori koostatud).....	57
Joonis 14. Duplikaatinfo automaatkontrolli ja masinõppe protsess minimalistlikult (autori koostatud).....	58

Tabelite loetelu

Tabel 1. Wiki.js tarkvara kasutused avaliku sektori veebilehena (autori koostatud)	32
Tabel 2. Artiklivaramu AS-IS lahenduse SWOT analüüs (autori koostatud).	65
Tabel 3. Töö tulemuste kokkuvõte (autori koostatud).	67

Sissejuhatus

Käesoleva magistritöö eesmärgiks on teostada eelanalüüs mille põhjal saab Riigi Infosüsteemi Amet kaasajastada riigi ülest infoartiklite haldust.

Töös antakse vastused järgmistele küsimustele:

- Milliseid võimekusi on vaja lisada või milliseid parendusi teha hetkel implementeeritud artiklite halduse protsessis või tarkvaras?
- Kas praegune karbitoodetel põhineva lahendus on optimaalne?
- Mil viisil oleks praegust artiklite haldamise süsteemi vaja täiendada, et paremini joonduda rahvusvaheliste normide ning riikliku ja asutusesisese strateegiaga?
- Millistele funktsionaalsetele ja mittefunktsionaalsetele nõuetele peavad võimalikud muudatused artiklite halduse infosüsteemis vastama?

Magistritöö koosneb sissejuhatusest, viiest peatükist, kokkuvõttest, kasutatud kirjanduse loetelust ja lisadest:

- Esimene peatükk kirjeldab teemavaldkonda, määratleb probleemi, selgitab selle aktuaalsust, täpsustab magistritöö käsitusala, seab oodatavad tulemused ja sõnastab autori rolli tulemuste saavutamisel.
- Teises peatükis annab autor ülevaate uuritavast organisatsioonist ning analüüsib lahendust mõjutavaid strateegiaid ja osapoolte motivatsioon ja organisatsiooni ressursse.
- Kolmas peatükk kirjeldab praegu evitatud lahenduse tausta ja olemust ning annab ülevaate selle protsessidest ja tehnoloogiatest.
- Neljandas peatükis viib autor läbi olemasoleva lahenduse süsteemianalüüsi, käsitledes esmalt olulisemaid tehnilisi aspekte ning hüpoteese ning tehes iga

aspekti kohta järelduse. Peatüki lõpuks annab autor hinnangu süsteemi hetkeseisule ning teeb kokkuvõtva järelduse.

- Viies peatükk kirjeldab magistritöö tulemusi, sõnastades vajalikud muudatused süsteemides ja protsessides ning andes suuniseid järgnevateks tegevusteks.

Autor soovib tänada Riigi Infosüsteemi Ametit ning ARVA meeskonna liikmeid: Õie-Mari Aasmäe, Anneli Asser, Mari Ader ja Janar Nagel, kes võimaldasid autorile ligipääsu artiklivaramut puudutavale dokumentatsioonile ning panustasid enda aega autori küsimustele vastamisel.

Autor soovib avaldada sügavat tänu oma abikaasale, Nele Bernardile, kelle ennastsalgava toetuseta ei oleks käesolev magistritöö valminud.

1 Magistritöö eesmärk

Käesolevas peatükis annab autor ülevaate magistritöö valdkonnast, kirjeldab probleemi olemust, töö eesmärki ning käsitusala. Autor annab ülevaate töö vajalikkusest ning kirjeldab eesmärkide saavutamiseks läbiviidavaid tegevusi ja autori rolli töö koostamisel.

1.1 Valdonna ülevaade

Artiklivaramu ehk ARVA on Riigi Infosüsteemi Ameti arendatav lahendus mille eesmärk on koondada ühele platvormile riiklikud infoartiklid ning pakkuda viisi kuidas erinevad riigiasutused aga ka teised infosüsteemid saavad riiklikku avalikku teavet enda digikeskkondades kuvada või kodanikele selle info abil täiendavaid teenuseid pakkuda [2].

ARVA on artiklite hoidmise ja haldamise keskkond mis on liideste kaudu ühendatav teiste keskkondadega ning milles olev teave on ristkasutatav [2]. Hetkel kasutavad ARVA't teabe allikana hetkel Riigiportaal (eesti.ee) [3], Sündmusteenuste platvorm [4] ning Bürokratt juturobot [5].

Teabe uuendamine ja ajakohastamine käib teabeomanike poolt, kelleks on tavaliselt teised riigiasutused. Teabeomanikud ajakohastavad teavet ARVA's käsitsi, paralleelselt ja sõltumata enda keskkondadest ning ARVA pakub neile selleks kasutajaliidest Riigi Infosüsteemi Ameti hallatavas keskkonnas.

Tegemist on evitatud rakendusega.

1.2 Probleemi püstitus

Riigi avalike digikanalite teave on killustunud ja korduv, puudub ülevaade duplikaatinfo hulgast, aegunud info hulgast ning info uuendamine ei kasuta ära automatiseerimise kõiki võimalusi.

See muudab avaliku teabe haldamist keerulisemaks ning mõjutab negatiivselt riigi strateegiliste eesmärkide saavutamist.

Olukorra on autori hinnangul põhjustanud pinnapealse regulatiivse vastavuse ja kiire tarne eelistamine, funktsionaalselt sobivama ja adaptiivse süsteemi välja töötamise asemel. Tulemuseks on mitte terviklik lahendus kus info tekkekohas hõivamise ja hoidmise ning hajusa sünkroniseerimise asemel luuakse üha keerulisemaid mehhanisme info tekkimise koha ja info hoidmise koha kahepoolseks sünkroniseemiseks. Olemasoleval süsteemil puuduvad võimekused duplikaatinfo tuvastamiseks ja teabevaldaja automaatseks teavitamiseks.

Tuginedes autori koostatud olemasoleva lahenduse SWOT analüüsile (Lisa 2) võib probleemiga mitte tegelemine kaasa tuua järgnevad negatiivsed ilmingud:

- Sisulise vastavuse puudumine seatud strateegiatega mõjutab negatiivselt strateegiates sõnastatud tulemuseesmärkide saavutamist.
- Seatud eesmärkideks ebasobiva lahendusega jätkamine võib tekitada olukorra kus regulatsioonide või nõuete täienemisel ei ole ka vormilise vastavuse saavutamine, seatud aja- ja eelarveraamides, enam võimalik mis omakorda põhjustab mainekahju ja võib päduda ettekirjutusega.

Probleem mõjutab otseselt:

- Umbkaudu 60 RIA partnerit¹ riigiasutuste või avalik- õiguslike ettevõtete näol kellel on Avaliku Teabe Seadusest tulenevalt kohustus omada ja hallata avalikke veebilehti.
- Umbkaudu 1500 avaliku sektori kodulehe [6] haldureid kelle kohustuseks on tagada info asja- ja ajakohasus.

Probleem mõjutab kaudselt:

¹ Partnerasutuste suurusjärk täpsustatud autori poolt ARVA meeskonnaga läbi viidud intervjuu käigus 11.03.2024

- Eesti Vabariigi kodanikke ja nende võimet saada avaliku sektori digikanalitest aja- ja asjakohast teavet.
- Eesti Vabariigi Valitsust, läbi digiriigi kuvandi nii riigi siseselt kui rahvusvaheliselt.

1.3 Magistritöö aktuaalsus

Autorile teadaolevalt pole riigiülese avaliku sektori veebilehtede teabe haldamise ja sünkroniseerimise lahenduste osas lõputöid koostatud.

Ka levinud teaduskirjandus käsitleb riigi käsutuses olevast infost rääkides, peaasjalikult asutustes, riiklikes andmekogudes, registrites ja andmebaasides salvestatud teavet, mitte spetsiifiliselt riigi poolt, riiklikes digikanalites, avalikustatud ja üldsusele suunatud informatsiooni.

Magistritöö on vajalik, et hinnata tegevusi mis on vaja sooritada ning tuvastada süsteemid mis on vaja luua, et joondada riigi avaliku teabe haldus kehtivate normide ning riigi strateegiliste eesmärkidega.

Töö tulemused panustavad paremini integreeritud, automatiseeritud ning funktsionaalselt sobivama avaliku sektori teabehalduse lahenduse juurutamisesse.

Nõuded avaliku teabe haldusele lähtuvad ennekõike Avaliku Teabe Seadusest (AvTS) [7].

AvTS 4. peatükk, 2. jagu sõnastab teabe avalikustamise kohustuse üldkasutatavas andmesidevõrgus. AvTS §31 sõnastab, et: „*Riigikogu Kantselei, Vabariigi Presidendi Kantselei, Õiguskantsleri Kantselei, Riigikontroll, kohtud, valitsusasutused ja avalikõiguslikud juriidilised isikud on kohustatud pidama teabe avalikustamiseks veebilehte.*“. Sama seaduse §32 seab nõuded veebilehe ja mobiilirakenduse pidamisele ning lõikes 1 loetletud kohustuste hulgas on: „*avaldab veebilehel või mobiilirakenduses päevakohast teavet*“ (AvTS §32, lõige 1, punkt 2) ning: „*ei või avaldada veebilehel või mobiilirakenduses teavet, mis ei vasta tegelikkusele või on eksitav*“ (AvTS §32, lõige 1, punkt 3).

Oluline on siinjuures see, et AvTS ei tee vahet sellel, kas teabe avalikustab teabevaldaja või mõni teine osapool kellel on AvTS §31 alusel ning AvTS §28 loetletud ulatuses kohustus teavet avaldada. Kohustus tagada teabe ajakohasus ja korrektsus ning vastutus normi rikkumise eest, lasub teabe avalikustajal.

Käesolevalt puudub mehhanism avaliku teabe duplikaatide tuvastamiseks. Puudub mehhanism teabe avaldaja automaatseks teavitamiseks teabe muutumisest ning puudub ka mehhanism teabevaldajale teabe kasutamisest ülevaate andmiseks. Defineeritud protsessi puudumise tõttu ei saa autori hinnangul eeldada, et praegune olukord on vastavuses seadusandja tahtega AvTS mõttes.

Tegevused erinevates allikates teabe ajakohasena hoidmiseks viiakse tihti läbi käsitsi, ning defineerimata protsessidega. Illustreeriva näitena saab tuua Postimees Gurpi poolt 16.11.2022 avaldatud artikli juhtumist, kus perekond jäi riigiasutuse veebilehel info uuendamata jätmise tõttu ilma lapse sünnitoetusest. Ametnikule omistatud tsitaat artiklist: „*Pean vabandama selle pärast, et info ei olnud kõikides kohtades õigel ajal uuendatud ja tekitas segadust. See oli inimlik eksitus ja see info on nüüd korrigeeritud,*“ [8].

Strateegia „Eesti 2035“ (EESTI 2035)

Eesti pikajaline strateegia [9] ja selle tegevuskava [10] võeti Riigikogu poolt vastu 2021. aasta kevadel. EESTI 2035 tegevuskava uuendatakse igaastaselt, viimane uuendus kinnitati Vabariigi Valitsuse poolt 25.04.2024.

EESTI 2035 sõnastab ühe pikaajalise sihina valdkonna nimega Riigivalitsemine. Nimetatud valdkonnale on EESTI 2035 tegevuskavas muuhulgas seatud järgmised eesmärgid:

1. „*Eesti on edasipürgiv ja demokraatlik riik*“ mille üldiseks mõõdikuks on seatud „*Avalike teenustega rahulolu*“ [10];
2. „*Avalikud teenused on kvaliteetsed, etteaimatavad ja kättesaadavad igas piirkonnas, tagades inimeste põhiõigused*“, üldise mõõdikuga „*KOV-i teenustega rahulolu*“ [10];

Soovitud muutuse sisuna on Vabariigi Valitsus muuhulgas rõhutanud vajadust kasutajakesksete otseste avalike teenuste kättesaadavuse parandamiseks ning teenuste

seiresüsteemide arendamiseks. Samuti platvormipõhise lähenemise arendamist andmete ja digilahenduste vallas ning digiriigi arendamisel pideva katsetamise kultuuri ja korralduse arendamist koostöös erasektori ja vabakonnaga [10].

Avaliku teabe ajakohasus ning võimekused seda tagada mõjutavad otseselt EESTI 2035 tegevuskava eesmärke ning EESTI 2035 strateegia pikaajalist sihti. Nende võimekuste suurendamine joondub EESTI 2035 tegevuskavas loetletud tegevustega.

Digiühiskonna Arengukava 2030 (DÜAK 2030)

DÜAK 2030 [11] on Majandus- ja Kommunikatsiooniministeeriumi (MKM) visioonidokument, mis lähtub EESTI 2035 strateegiast ning võeti vastu 2021. aasta detsembris.

Visiooni üldeesmärgist tulenevalt seab MKM sihi „*tagada 2030. aastaks parim digiriigi kogemus.*“. Selle saavutamise põhiliseks mõõdikuks seatakse (sarnaselt ja joondudes EESTI 2030'ga) „*avalike digiteenustega rahulolu*“ [11].

DÜAK 2030 sõnastab põhimõtted mida tuleks arengukava raames tegevuste kavandamisel ja elluviimisel järgida ja rõhutab usaldusväarsuse ning probleemide ennetamise vajadust ning vajadust olla uuendusmeelne ja proovida uusi lähenemisviise [11].

Riigi Infosüsteemi Ameti Strateegia 2021 – 2025 (RIA 21-25)

RIA visioon on sõnastatud kui: „*Loome ja kaitseme koos maailma parimat digitaalset ühiskonda.*“ [12].

RIA missioon on sõnastatud kui: „*Kompetentsikeskusena tagame Eesti (e-)riigi usaldusväarsuse ja mugavad teenused.*“ [12].

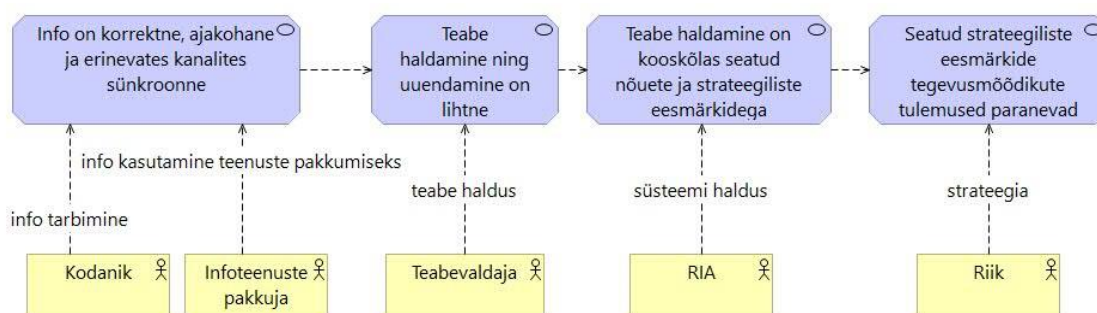
RIA 21-25 on organisatsiooni sisene strateegilisi fookusi ja põhieesmärke seadev dokument mille eesmärgiks on DÜAK 2035 edukas elluviimine RIA ülesannete piires [13].

Ühe strateegilise eesmärgina sõnastatakse: „*Tagame kesksed IKT platvormid järgmise põlvkonna digiteenuste pakkumiseks.*“. See on osa fookussuunast milleks on väärtuse pakkumine kliendile ning selle põhiliseks mõõdikuks on klientide rahulolu RIA

pakutavate teenustega. Eesmärki on täiendavalt selgitatud järgmiselt: „Meie pakutavad lahendused toetavad nutikat riigivalitsemist ja mugavat kodaniku vajadustest lähtuvat avalike teenuste pakkumist.“ [13].

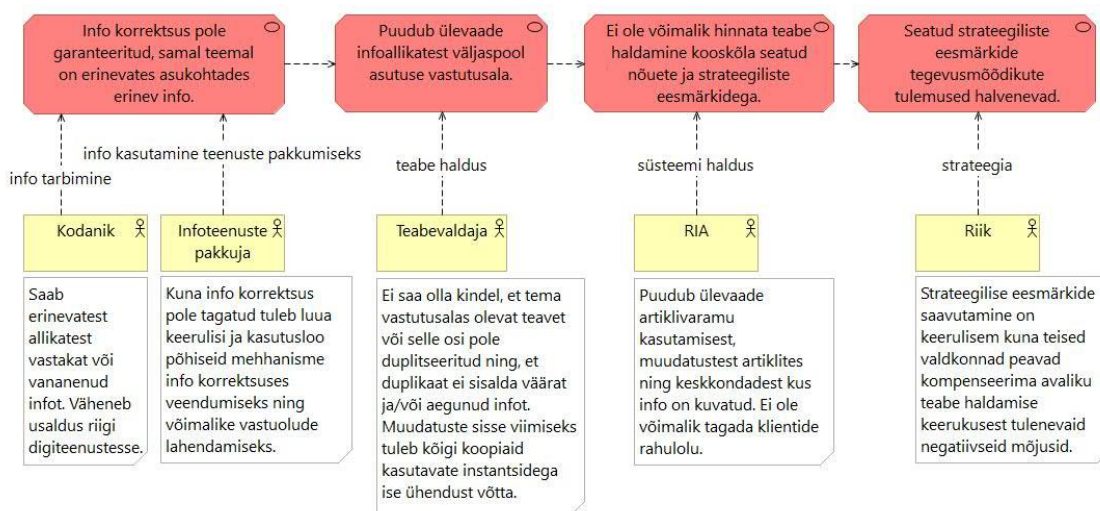
Selle eesmärgi täitmiseks tuuakse ühe tegevusena välja kesksete alusplatvormide ja tehnoloogiate arendamise mis lihtsustaksid avalike teenuste pakkumist.

Alloleval joonisel toob autor välja üldise väärtuse mida ARVA protsesside parendamine endaga kaasa toob:



Joonis 1. Avaliku teabe haldamise protsessi parendamisest saadav üldine väärtus (autori koostatud).

Järgnevalt kirjeldab autor negatiivseid ilminguid mis kaasnevad ARVA protsessidega mitte tegelemisega:



Joonis 2. Avaliku teabe haldamise protsessi muudatustega mitte tegelemisest tekkiv üldine kahju (autori koostatud).

Kokkuvõtteks

Autori hinnangul viitavad magistritöö aktuaalsusele tõsiasi, et praegune lahendus ei suuda tagada vastavust kehtivate normidega ning lahendusel on mõju enamike põhiliste huvigruppide strateegilistele eesmärkide saavutamisele. Käesolev magistritöö on sisendiks avaliku teabe haldamise protsesside parendamisele.

1.4 Magistritöö eesmärk, kasutatud metoodika ja eeldatav tulem

Käesoleva magistritöö eesmärk on kirjeldada Artiklivaramu (ARVA) infosüsteemi hetkeseisu, analüüsida selle tehnoloogilist sobivust ning pakkuda välja lahenduse kontsept mis tagab parema vastavuse kehtivate normide ning strateegiliste eesmärkidega.

Seatud eesmärgi saavutamiseks, kasutab autor meetoditena põhiliselt:

- *Lean* lähenemist
- Disainmõtlemise põhimõtteid;
- Domeenjuhitud disaini põhimõtteid;
- SWOT analüüsi

Modelleerimiseks ja vaadete koostamiseks kasutab autor järgmisi raamistikke või keeli:

- TOGAF ja ArchiMate;
- BPMN;
- C4 arhitektuurimudel;
- UML;

Analüüsietappidena viib autor läbi:

- Olemasoleva olukorra kirjeldamise;
- Olemasolevate tehnoloogiate ja protsesside analüüsi ning järelduste tegemise;
- Sobivama tulevikuolukorra ning selle nõuete kirjeldamise;

Töö tulemusena autor:

- Kirjeldab ning analüüsib ARVA hetkeseisu.
- Kaardistab ARVA protsessid ja seosed huvitatud osapoolte, nõuete ning strateegiliste eesmärkidega;
- Annab hinnangu ARVA praeguse lahenduse tehnoloogilisele sobivusele, strateegiliste eesmärkide saavutamiseks;
- Annab hinnangu olemasoleva lahendusega jätkamise mõistlikkusele ning kui vaja pakub välja uue visiooni ja lahenduse eskiisi;

1.5 Magistritöö käsitlusala

Magistritöö käsitlusalasse kuulub:

- ARVA't mõjutavad regulatsioonid ja strateegiad;
- Huvitatud osapoolte kaardistus ning nende strateegia ja motivatsiooni analüüs;
- Olemasoleva ARVA infosüsteemi kirjeldamine;
- ARVA's kasutusel olevate või planeeritavate valmistoodete sobivusuuring;
- ARVA võimekuste ja võimaluste analüüs;
- Loodavate võimekuste ja uute protsesside kirjeldamine;
- Loodavate infosüsteemide eskiisjoonised ja komponentide kirjeldused;

Magistritöö käsitlusalasse ei kuulu:

- RIA eksisteeriva taristu, sisemiste süsteemide arhitektuuri ja spetsiifiliste tehnoloogiate kirjeldamine, osas mis puudutab evitust, käitlemist ja turvalisust;
- Detailne süsteemianalüüs (sh. olemi-suhte diagrammid, liideste kirjeldused, evituse plaan jne.);

- Eksisteerivate RIA asutusesiseste või riiklike mittefunktsionaalsete nõuete analüüs ja hindamine ARVA kontekstis.
- Ekraanivaadete prototüüpimine;
- Lahenduse arendusplaan, mahuhinnangud, teekaart või rahastusmudelid;
- Õiguslik analüüs lahenduse teostamiseks;
- Tarkvaraarenduse planeerimine, teostamine või hangete korraldamine;

1.6 Autori roll

Autor töötas aastatel 2020-2021 artiklivaramu arendusmeeskonna juhina (*team lead i.k.*) ettevõttes Nortal AS mis teostas ARVA rakenduse täiendarendusi Riigi Infosüsteemi Ametile. Autor töötab praegu samuti ettevõttes Nortal AS, riigitehnoloogiate arhitektina, mis seisneb digiriigi lahenduste analüüsis ja visioneerimises, tehnilisest aspektist lähtuvalt, ning erinevate osapoolte nõustamises.

Autor tugineb käesolevas magistritöös oma töö käigus saadud teadmistele ja kogemustele mida autor on ARVA osas täiendanud 2024 aastal Riigi Infosüsteemi Ameti ARVA meeskonna erinevate huvitatud osapooltega suheldes ning analüüsides evitatud lahenduse dokumentatsiooni.

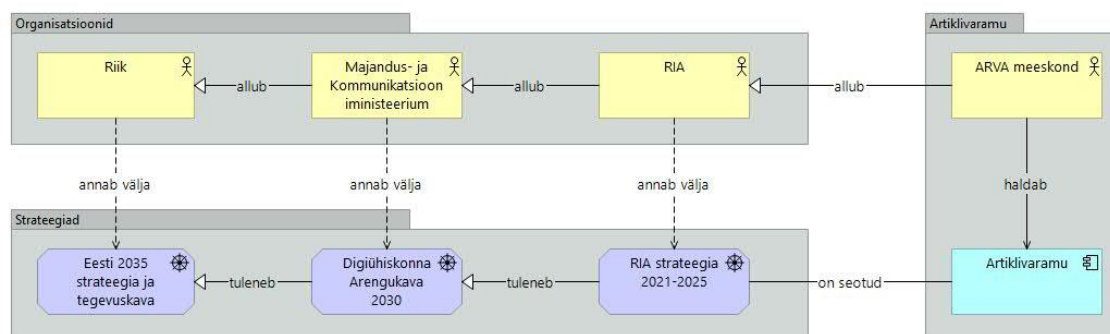
Autor täidab magistritöö koostamisel arhitekti ja RIA välise konsultandi rolli ning kombineerides varasemad teadmised ja magistritöö käigus läbiviidava analüüsi, realiseerib autor magistritööle seatud eesmärgid.

Magistritöö sisuks olevad tegevused on autor teostanud ainuisikuliselt. Kõik kasutatud kirjandusallikad ja teiste osapoolte sisendid toob autor välja.

2 Organisatsiooni ärianalüüs

Riigi Infosüsteemi Amet on Majandus- ja Kommunikatsiooniministeeriumi haldusalas tegutsev riigiamet, mis koordineerib riigi infosüsteemi arendamist ja haldamist, korraldab infoturbega seotud tegevusi ja käsitleb Eesti arvutivõrkudes toimuvaid turvaintsidente. RIA konsulteerib avalike teenuste osutajaid, kuidas oma infosüsteeme nõuetekohaselt hallata ja teostab nende üle järelevalvet. Lisaks on RIA Euroopa Liidu struktuuritoetuste rakendusüksus [14].

Alljärgneval joonisel (Joonis 3) kirjeldab autor seoseid organisatsioonide ja strateegiate vahel mis on relevantset Artiklivaramu analüüsil.



Joonis 3. Organisatsiooni hierarhia ja Artiklivaramu infosüsteemi seos (autori koostatud).

2.1 Huvitatud osapooled

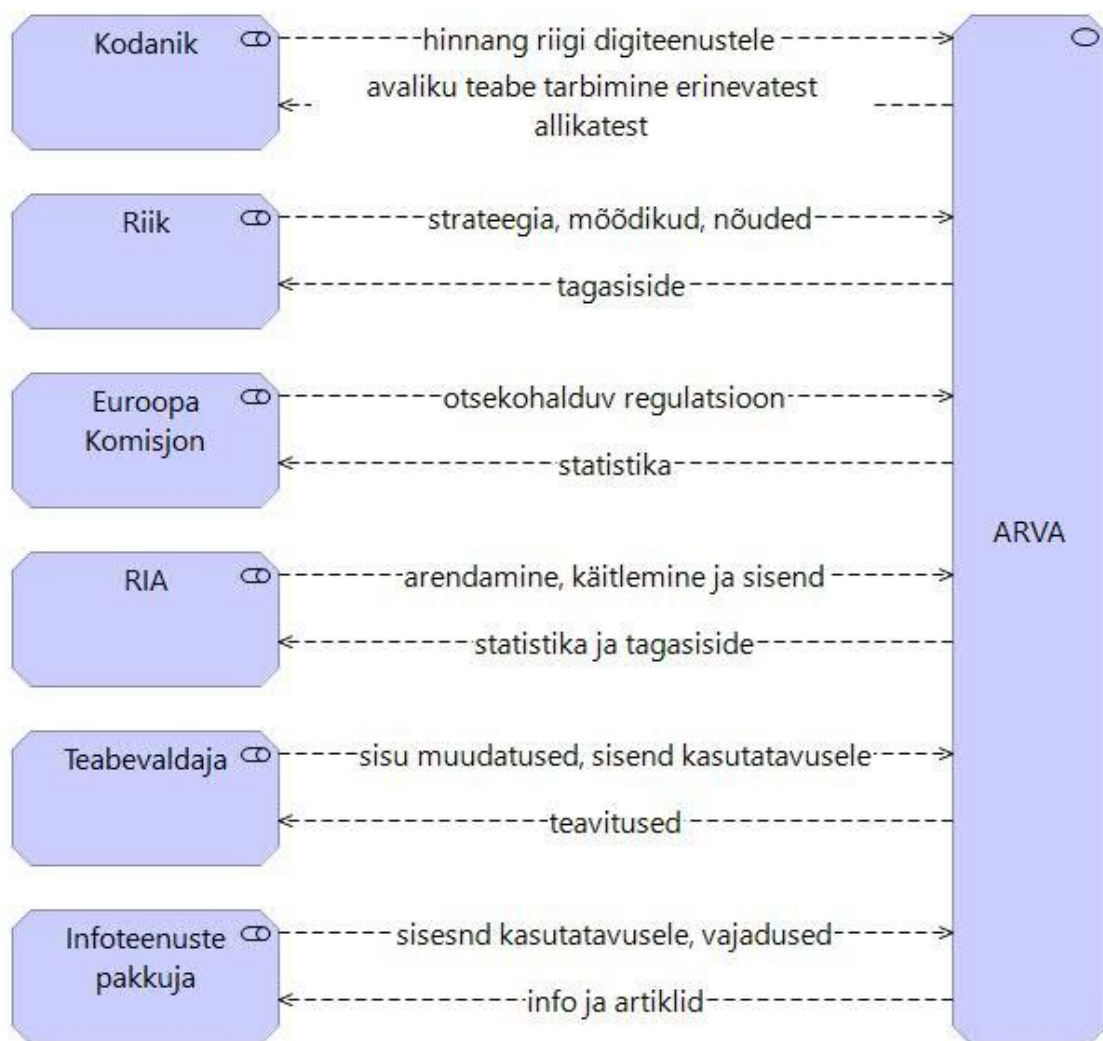
Alljärgnevalt (Joonis 4) toob autor välja põhilised huvigrupid ning nende põhilised motivaatorid Artiklivaramu kontekstis.

Kodaniku all mõistab autor antud kontekstis nii füüsilisi kui juriidilisi isikuid kes avalikku teavet riiklikest infoallikatest tarbivad. Kuna juriidiline isik tarbib avalikku teavet ainult läbi konkreetse füüsilisest isikust esindaja või töötaja siis on selline üldistus autori hinnangul õigustatud.

Kodaniku all on mõeldud lisaks Eestile ka Euroopa Liidu riikide kodanikke ning alalise elamisloaga isikuid. Muud riikliku teavet otsivad/tarbivad isikud (nn. turistid, väljaspool Euroopa Liitu asuvad ettevõtted või riigiasutused) ei ole käesoleva töö käsitusallas.

Infoteenuste pakkuja on keegi (tavaliselt projekt või initsiatiiv) kes pakub olemasoleva avaliku teabe pinnalt tuletatud digitaalseid teenuseid. Näiteks sündmusteenuste [4] meeskond kes agregeerib ja kuvab avaliku teabe vastavalt kodaniku personaalsele olukorrale. Või Bürokratti [5] initsiatiiv mis liidestab juturoboti avaliku teabe koguga. Infoteenuse pakkuja ei halda teavet vaid ainult kasutab seda.

Infoomanik on avaldatud teabe valdaja. Reeglina on infoomanik ise ka info põhiliseks avaldajaks. Põhimõtted mille alusel infoomanikud määratakse, teabe valduse üle andmise ning ilma valdajata teabega tegelemise protsessid ei ole käesoleva töö käsitlusalas.



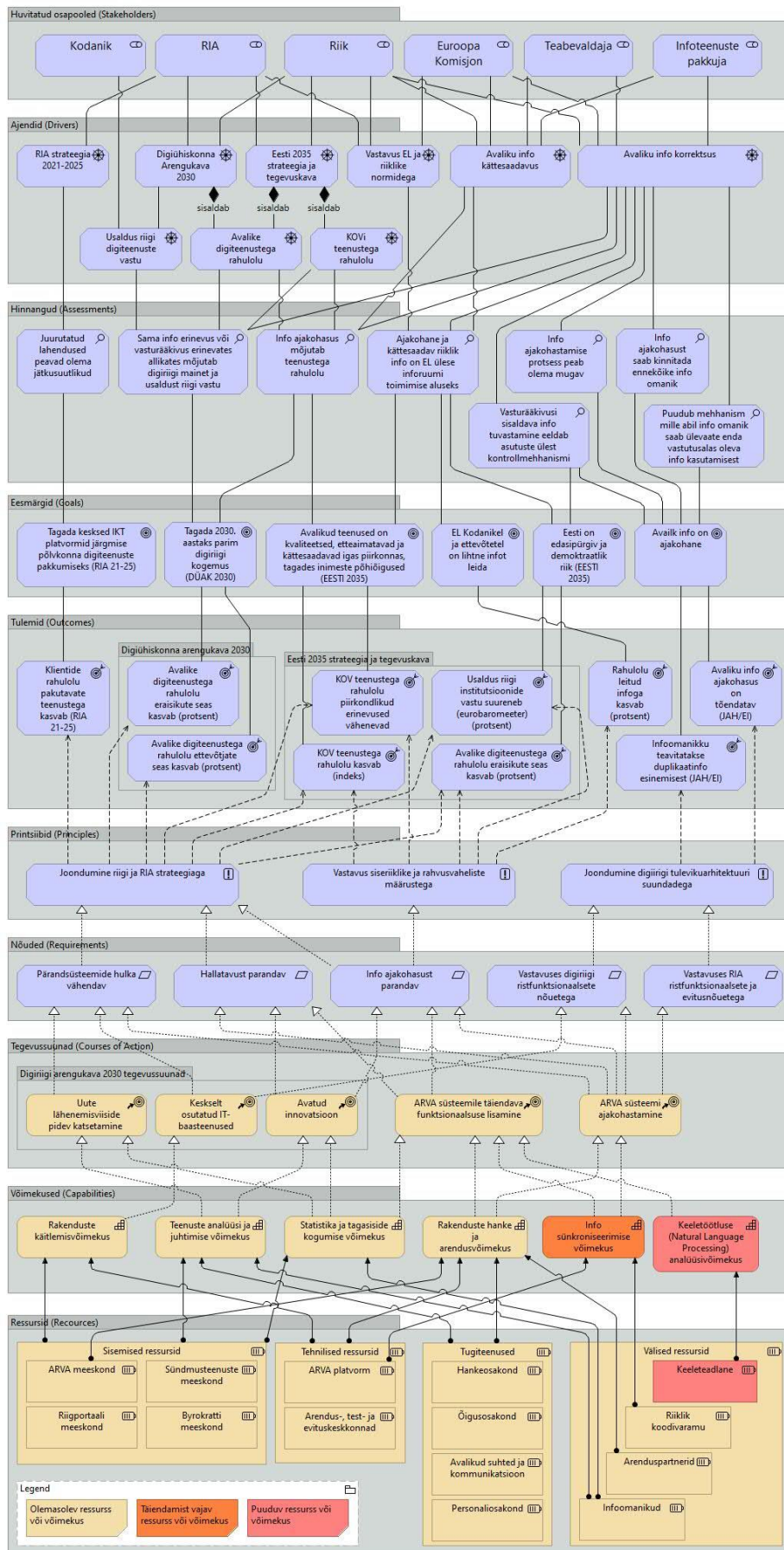
Joonis 4. Huvigruppide skeem (autori koostatud).

2.2 Strateegia- ja motivatsioonimudel

Selleks, et pakkuda välja arhitektuurivisioon viis autor läbi üldise ARVA äristrateegia, ajendite ning eesmärkide analüüsi mille tulemusel koostas ArchiMate notatsioonis strateegia- ja motivatsioonimudeli (Joonis 5).

Mudeli loomiseks ja analüüsi käigus, sooritas autor järgmised tegevused:

- kaardistas huvitatud osapooled ja hindas nende põhilisi ajendeid;
- sõnastas ajendite põhjal hinnangud;
- kaardistas nii ajenditest tulenevad kui ka autori enda poolt tuletatud strateegilised eesmärgid (kui eesmärk tuleneb otseselt riiklikust strateegiast on vastav strateegiadokument lühendina eesmärgi juures eraldi välja toodud);
- kaardistas või sõnastas tegevuseesmärgid (sarnaselt strateegilistele eesmärkidele on eraldi välja toodud tegevuseesmärgid mis tulenevad riigi strateegiadokumentidest);
- sõnastas üldised printsiibid ja nõuded millest tuleks tegevuseesmärkide saavutamisel lähtuda;
- tõi välja tegevussuunad mida tuleks rakendada (eraldi on välja toodud käesoleva töö raames analüüsitava probleemi kontekstis relevantseid tegevussuunad mis on sõnastatud riigi strateegiadokumentides);
- kaardistas tegevussuundade rakendamiseks vajalikud võimekused ning nende võimekuste tagamiseks vajalikud ressursid.



Joonis 5. Organisatsiooni motivatsiooni- ja eesmärgmudel ning seos ressurssidega (autori koostatud)

3 Taust

Käesolevas peatükis kirjeldab autor Artiklivaramu olemasolevat lahendust ja selle põhimõtteid nii nagu nad on teostatud ja sõnastatud praegu. Kirjelduste aluseks on Artiklivaramu dokumentatsioon, mille on koostanud Riigi Infosüsteemi Amet, autori isiklik kogemus Artiklivaramu eelmise iteratsiooni arendusel, 2021. aastal ning struktureerimata intervjuud Artiklivaramu praeguse meeskonnaga.

3.1 ARVA tutvustus

ARVA on lahendusena nii Riigi Infosüsteemi Ameti andmetel kui ka autori hinnangul maailmas unikaalne. Autor ei suutnud tuvastada ühtegi teist lahendust mille eesmärk ja funktsioon oleks, riigi üleselt, kõigi avaliku sektori veebilehtede informatsiooni koondamine ja riskasutatavalt kättesaadavaks tegemine.

ARVAs kui „ühtse tõe allikas“ sisalduva info eest vastutavad infoomanikud, kel on kohustus infot hoida ajakohasena ning seda vajadusest lähtuvalt muuta [2].

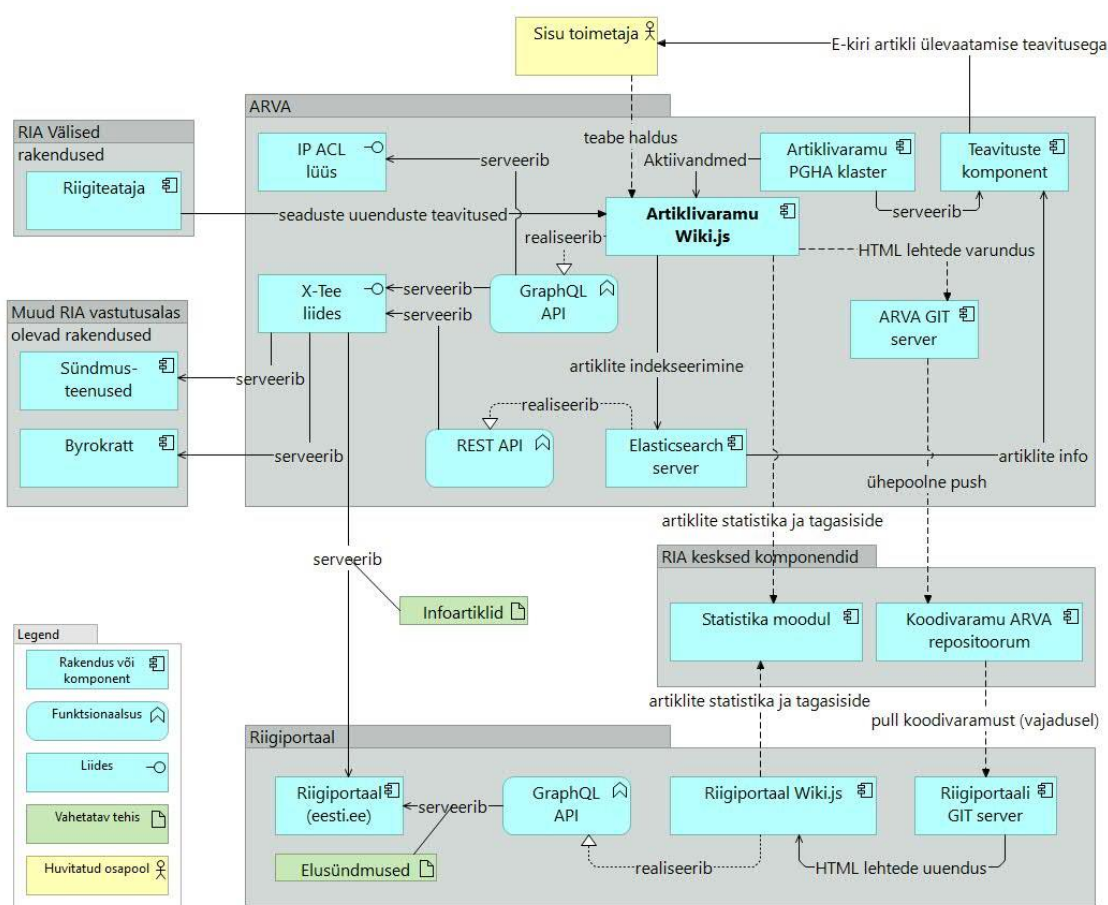
ARVA baasfunktsionaalsused [2]:

- Artiklite lisamine, muutmine, kustutamine;
- Teksti visuaalidega täiendamise võimalus;
- SDG märgendite lisamise võimalus;
- Süvalingi lisamine ehk teenus artikli sees;
- Toimib otsing kaustapuus (artiklitel on küljes silt/märksõna);
- Kõiki artikleid saab esitada masinloetavalt;
- Asutuste kontaktid ja KKK on toodud eraldi artiklitena;

- Liidestus Riigiteatajaga kus artikli metainfos on sellega seotud seadused ning mõne seaduse muutumisel ilmub artikli juurde vastavasisuline märge;
- Teavitus süsteem mis annab artikli omanikule märku vajadusest artikkel üle vaadata tulenevalt seaduse muutumisest või tagasisidest;
- Asutusekesksete kasutajagruppide moodustamine;
- Liidestus TARA ja GovSSO'ga kasutajate autentimiseks;

3.2 ARVA tehniline ülevaade (AS-IS)

Alljärgnevalt (Joonis 6) kirjeldab autor ARVA süsteemi komponente ja liidestusi üldistatult ning selles ulatuses mis autori hinnangul on relevantne. ARVA komponendid on (sarnaselt teistele RIA vastutusalas olevatele rakendustele) täiendavalt liidestatud erinevate käidukäitlus, monitooringu ja ligipääsusüsteemidega, mille kirjeldamine ei ole käesoleva magistritöö käsitlusalas.



Joonis 6. ARVA liidestuste komponentjoonis, AS-IS, C1 tasemel (autori koostatud)

Artiklivaramu keskse halduskomponendina on kasutusel Wiki.js nimeline sisuhaldusrakendus [14]. Tegemist on avaliku vabavaralise sisuhalduslahendusega mis on kirjutatud Node.js raamistikku kasutades ning toetab liidestumist mitmesuguste andmeallikatega. Wiki.js on paigaldatav kõrgkäideldavalt.

ARVA meeskond on viinud Wiki.js võimekuste täiendamiseks tootesse sisse muudatusi. Valdavalt on muudatused realiseeritud moodulitena, mis on Wiki.js autori poolt soovitatud viis oma toote täiendamiseks. Samas on ARVA arendustööde käigus esinenud ka olukordi kus muudatused moodulite abil pole võimalikud ning muudatusi on tehtud Wiki.js algkoodis.

Algkoodi muudatused on seotud järgmiste teemadega:

- GovSSO autentimisvõimaluse lisamine. Muudatuse tingis vajadus siduda iga Wiki.js kasutaja tema isikukoodiga, mille tulemusel tuli muuta Wiki.js administraatori paneeli kasutajate vaadet.
- Välja saadetud teavituste monitooring. Muudatuse tingis vajadus omada ülevaadet välja saadetud teavitustest, mille tulemusel lisati administraatori paneeli vaadete valikusse menüüpunkt „Teavitussüsteemi monitooring“ alampunktiga „Teavitused“, millele vajutades avaneb uus vaade nimega „Teavitused“.
- Artikliomaniku töölaual kaustapuu muutmise nupu eemaldamine. Muudatuse tingis vajadus piirata väliste kasutajate ja/või sisutoimetajate võimalust muuta töölaua üldisi navigatsioonielemente. Selleks tuli vastav funktsionaalsus artikliomaniku vaatest eemaldada terviklikult (mitte pelgalt visuaalse komponendi peitmise läbi).

Peamine mehhanism Wiki.js liidestumiseks teiste süsteemidega on GraphQL abil realiseeritud API otspunktid. Aktiivandmete salvestamiseks ja lugemiseks kasutab ARVA PostgreSQL kõrgkäideldavat klastrit millega liidestumist Wiki.js toetab standardselt. Wiki.js toetab standardselt ka automaatset GIT serveriga sünkroniseerimist (HTML lehtede varundamiseks) ning GIT repositooriumi kasutamist tõe allikana (*single source of truth i.k.*) [15].

ARVA GraphQL API võimekusi kasutavad nii artiklivaramu X-Tee liides kui ka IP ACL lüüs, viimane on võimaldamaks ARVA API'le ligipääsu nendel välistel süsteemidel millel X-Tee võimekus puudub. Hetkel on ARVA API liidestujaid kolm:

- Riigiportaal (eesti.ee);
- Sündmusteenuste rakendus;
- Bürokratt rakendus;

Kõik nimetatud rakendused kasutavad X-Tee liidest. Hetkel ei kasuta ükski väline osapool IP ACL lüüsi.

ARVA X-Tee alamsüsteem

RIA turvaserveris registreeritud „artiklivaramu“ nimelise X-Tee alamsüsteemi näol on tegemist kahe REST tüüpi teenusega:

- Artiklite päringu teenus nimetusega „arva“, mis võtab sisendiks GraphQL vormingus päringu ja tagastab ühe artikli andmed vastavalt päringus defineeritud väljadele. Artikkel tuvastatakse numbrilise identifikaatori järgi [16].
- Artikli otsingu teenus nimetusega „otsing“, mis võtab samuti sisendiks GraphQL vormingus päringu ning tagastab ühe või mitme artikli andmed [16]. Teenus on küll registreeritud aga pole hetkel evitatud.

Riigiteataja liidestus

ARVA on liidestatud Riigiteatajaga saamaks sealt sisendit seaduste muutumise kohta. Seda võimekust kasutab ARVA teavituste komponent ja info selle kohta, et artikliga seotud õigusaktid on muutunud, lisatakse ka artiklile.

Teavituste komponent

Teavituste komponendi põhiline funktsionaalsus on edastada teabevaldajale e-kirja teel meeldetuletus, talle kuuluva teabe üle vaatamise kohustusest. Seda teeb teavituste komponent ühel kolmest juhust:

1. Ajatatud ja regulaarse protsessina üks kord aastas.

2. Juhul kui artikli kasutajatagasiside halveneb.
3. Artikliga seotud õigusaktides toimub muudatusi.

Teavituste komponendis salvestatakse iga artikli kohta mitu asutuse kontakti, erinevatel tasemetel. Kui teavitusele ei reageerita (tagasisuunav viide e-kirjas) siis saadab Teavituste komponent kirja järgmise tasandi e-posti aadressile, päädudes vastutava asutuse üldinfo aadressiga.

Elasticsearch

Võimaldamaks efektiivsemat otsingut ja artiklite ning nende osade masinloetaval kujul esitamist, indekseeritakse kõik artiklid Elasticsearch serveris. Elasticsearch on avatud lähtekoodiga otsingu ja analüütikatööriist ning dokumentandmebaas mis pakub põhjalikku API liidest [18].

ARVA seos RIA kesksete komponentide ja riigiportaaliga

Riigiportaal on hetkel ainus (ARVA mõistes) väline komponent mis kasutab andmeallikana koodivaramus asuvat ARVA repositooriumi. Kuigi Riigiportaali enda arhitektuur jääb käesoleva magistritöö käsitlusalast välja, pidas autor siiski vajalikuks joonisel 6 kirjeldada Riigiportaali liidestumist väga üldisel tasemel. Seda põhjusel, et Riigiportaal kasutab koodivaramust artiklite taaskasutamiseks enda koopiati Wiki.js rakendusest ning joonisel 6 kirjeldatud liidestus on lähenemisviis mida hetkel soovitatakse implementeerida ka teistel välistel osapooltel.

Nii Riigiportaal kui ARVA edastavad artiklite kasutuse statistikat ning kasutajate tagasisidet (vastavalt SDG määruse [1] nõuetele) kesksele statistika moodulile. Kuna tegemist on RIA keskse komponendiga millele, autorile teadaolevalt, ei plaanita asutuse välist ligipääsu tekitada siis hetkel evitatud lahenduses ei ole realiseeritud väliste osapoolte võimalust statistikat edastada.

3.3 Artiklite uuendamise protsess (AS-IS)

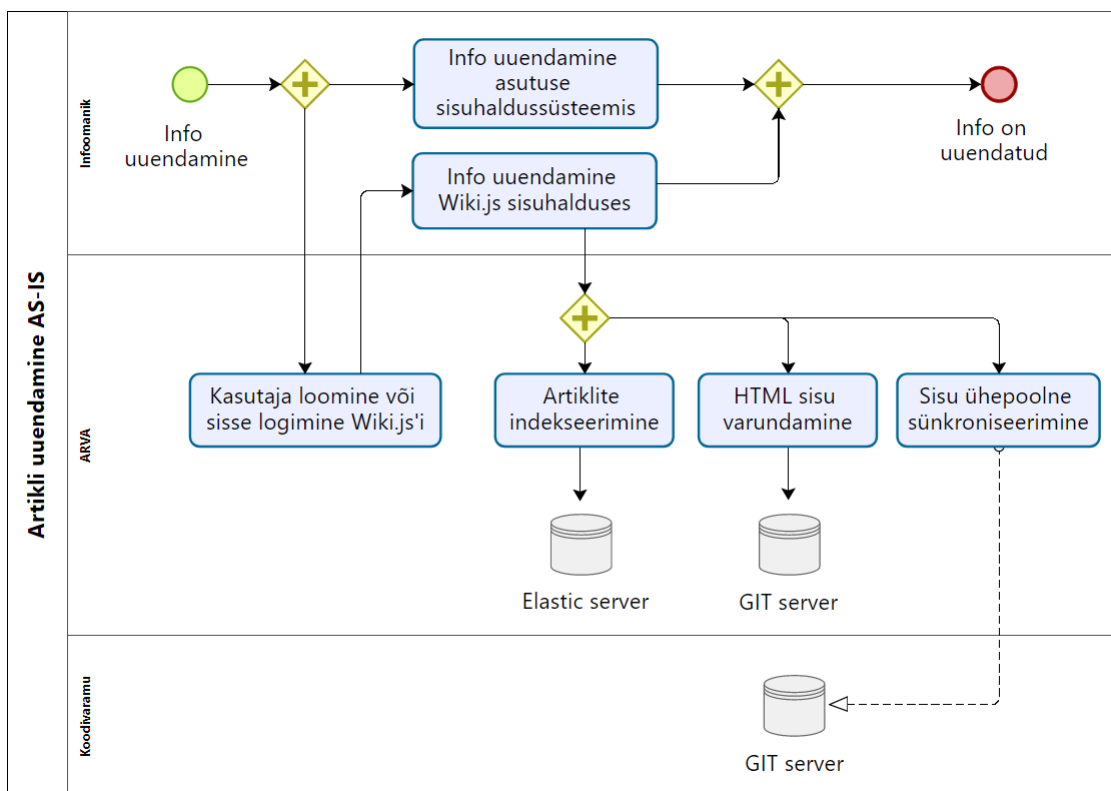
Alljärgnevalt (Joonis 7) kirjeldab autor artiklite uuendamise protsessi (protsess on sama ka uue artikli lisamisega) ülevahtlikult. Viimase evitatud versiooni kohaselt algatab info uuendamise teabevaldaja. Seda kuidas ja mis hetkel info uuendatakse teabevaldaja enda

keskkonnas ARVA praegune mehhanism ei jälgi, info uuendamine ARVA's on paralleelprotsess.

Teabevaldaja peab sisse logima ARVA sisuhaldusrakenduses. Välise kasutaja puhul kasutades GovSSO/TARA lahendust, sisekasutaja puhul RIA AD (*Active Directory i.k.*) lahendust ning tegema muudatused Wiki.js kasutajaliidese abil. Pärast muudatuse salvestamist teabevaldaja kasutajatekond ARVA's lõpeb ning edasiste ja taustaportsesside osas teda ei informeerita ning temalt sisendit ei võeta.

Ei ARVA ega teabevaldajate infosüsteemid ei ole võimelised tuvastama erisusi kummaski keskkonnas evitatud teabe osas. Protsess selleks on struktureerimata, manuaalne ning kasutajaküsitluste¹ põhjal varieerub asutuste vahel, sõltudes konkreetsetest ametnikest ning töökorraldusest.

Hetkel on teabevaldaja põhiline motivaator ARVA's salvestatud info ajakohastamiseks, kohustus tagada info korrektsus Riigiportaalis, tulenevalt SDG määrusest.



Joonis 7. Artiklite uuendamise AS-IS protsess ülevaatlilikult (autori koostatud)

¹ Läbi viidud RIA poolt, 2022 aastal.

4 Süsteemianalüüs

4.1 Wiki.js

Olemasolev lahendus Wiki.js näol võimaldas ARVA meeskonnal, elementaarse sisuhalduse kiirelt ja madalate kuludega evitada ning pakkuda teenuseid riigiportaalile ja sündmusteenustele.

4.1.1 Levik

Tegemist on levinud CMS'iga millel on arvestatav kasutajabaas 9349 [18] evitatud instantsi näol. Autor suutis tuvastada kolm juhtu kus Wiki.js'i kasutatakse avaliku sektori poolt:

Tabel 1. Wiki.js tarkvara kasutused avaliku sektori veebilehena (autori koostatud)

Riik	Wiki.js instants	Kasutusala
Nigeeria	https://wiki.nimc.gov.ng/	Digitaalse identiteedi teabekogu.
Ukraina	https://wiki.naqa.gov.ua/	Kõrghariduse kvaliteedikontrolli agentuuri teabekogu.
Kambodža	https://kms.ncdd.gov.kh/	Tervishoiu süsteemide haldamise teabekogu.

Igal tuvastatud juhul kasutati Wiki.js'i sihtotstarbeliselt, ehk kindla domeeni teabekoguna (nn. „wiki“ i.k.). Autor ei suutnud tuvastada ühtegi juhtu kus Wiki.js oleks kasutusel asutuste üleselt avaliku teabe haldamiseks või keskse avaliku teabe repositooriumina.

4.1.2 Jätkuarendused ja tootetugi

Kuigi Wiki.js koodibaasi panustab 30 arendajat [19], ning sellel on kümneid toetajaid on valdavalt siiski tegemist ühe vabatahtliku¹ poolt hallatava rakendusega. Rakenduse

¹ Nicolas Giard, (<https://github.com/NGPixel>)

omanik on koodi panustanud 1824 korral, kõik ülejäänud panustajad kokku 111 korral, nendest individuaalselt ainult üks panustaja 14 korral ning ülejäänud vähem kui 10 korral. [19] [20].

Säärane sõltuvus ühest omanikust teeb lahendusest pikas perspektiivis riskantse valiku, juhul kui Wiki.js omanik ei ole enam võimeline projekti sama intensiivsusega panustama. Tegemist on väikese ja fokuseeritud kogukonna poolt toetatud rakendusega mis laialdasest kasutamisest hoolimata, ei ole võrreldav suuremate organisatsioonide poolt juhitud vabavaralise tarkvaraga [22].

Autor ei suutnud tuvastada ühtegi spetsiifilist tehnoloogiapartnerit või protsessi mille abil oleks võimalik hankida probleemide, küsimuste, või täiendarenduse vajaduse korral kommertstuge Wiki.js'ile. Kõik säärased vajadused peab RIA katma sisemiste ressursside või arenduspartnerite abiga.

Märgiks sellest, et Wiki.js omaniku huvi toodet edasi arendada on raugemas on 3nda versiooni evituse pikk viibimine ning tööde aeglane areng. Viimane versiooniuuendus evitati autori poolt 20. Veebruaril, 2022 aastal ehk praeguseks enam kui 2 aastat tagasi. Ning viimane sisuline uuendus (*major release i.k.*) 6. Septembril, 2020 aastal [22].

Wiki.js omanik ei tee sellest ka ise saladust. Tõlgitud tsitaadid projekti infolehel [23]:

Omanik vastab korduma kippuvatele küsimustele uuenduste ajakava, prioriteetide ning täiendusettepanekute osas: „*Wiki.js on ehitatud pea ainuisikuliselt 1 arendaja poolt, ilma kompensatsioonita, tema vabast ajast (tööpäevade õhtud ja nädalavahetused).*“

Omanik kirjeldab kuidas saaksid soovijad teda aidata: „*Te võite panustada majanduslikult selleks, et seda projekti toetada. Toetused aitavad katta serverite ja võrguteenuste kulusid koos mõningate arendustööriistade kuludega.*“

Kui piisavalt suur hulk püsiannetusi on saavutatud, võib juhtivarendajal olla võimalik töötada projektiga täiskohaga.“

4.1.3 Järeldus

Autori hinnangul on Wiki.js omanik, hr. Giard, teinud suurepärasest tööd ning sisuliselt üksi, tootnud äärmiselt professionaalse ja laialt kasutatud tarkvara. Samuti on autori

hinnangul asjakohane, ARVA meeskonna poolt Wiki.js kasutusel võtmine esmase lahendusena tänu sisseehitatud toele GIT serveriga integreerumiseks.

Autori reservatsioonid Wiki.js jätkuva kasutamise osas püsiva ja keskse ARVA komponendina ning selle integreerimisel välistesse asutustesse, tulenevad ennekõike tootestatud Wiki'de ja CMS'ide üldisest sobimatuses selleks otstarbeks (autor analüüsib seda aspekti täiendavalt, käesoleva töö peatükis 5.2) ning riskidest mida kommertstoe puudumiseta ja ühest isikust sõltuva vabavaralise toote kasutamine endas kätkeb.

Wiki.js koodibaasis muudatuste tegemine on ka ARVA meeskonna hinnangul keeruline. Tuleb loota, et koodibaasi kas omanik võtab muudatused vastu või on RIA sunnitud tegema hargnemise (*i.k. fork*) autori koodist.

Autorile teadaolevalt ei ole RIA otstarbeks tehtud muudatusi seni suudetud Wiki.js koodibaasi sisse viidud, mistõttu RIA on juba praegu olnud sunnitud Wiki.js'ist enda klooni looma. Autori hinnangul on RIA poolt kasutatav kloon sellega juba praegu muutunud RIA hallatavaks pärandvaraks (*legacy i.k.*). Autori hinnangul ei ole sellises olukorras täiendarenduste osas vahet selle ja rätsepatööna tellitud tarkvara haldamise vahel. Pigem vastupidi, ilmselt oleks efektiivsem hallata väiksema funktsionaalsuste hulga, spetsiaalset rakendust mille koodistiili ja põhimõtetega meeskond algusest peale tuttav on ning mis on spetsiifiliselt RIA ja riiklike mittefunktsionaalsete nõuetega joondatud [24].

Wiki, CMS või DMS (*Document Management System i.k.*) pakub häid võimalusi sisu kuvamiseks ja kollaboratiivseks muutmiseks. Samas arvestades riiklikult hallatava avaliku teabe hulka, seotud osapoolte hulka, ning semantilisi iseärasusi, vajavad sellised nn. „karbitooted“ lõpuks ikkagi arvestatavat täiendarendust ning modifikatsioone, et vastata infoühiskonna üha kasvavatele nõuetele.

Olemasolevas lahenduses on juba tehtud täiendarendusi ja kohandusi, mugandamaks Wiki.js toodet RIA ja ARVA vajadustele. Samuti on autorile teadaolevalt käimas täiendav analüüs Wiki.js õiguste ja privileegide süsteemi kohandamiseks kuna õigustega seoses on rakenduses tekkinud ettearvamatuid tulemusi, GraphQL API päringute vastustes.

4.2 Wikide/CMS'ide kasutamise sobivus

Järgnevalt analüüsib autor Wiki ja CMS kategooria rakenduste üldist sobivust. Isegi kui autor jätab kõrvale Wiki.js'i või muud väikesed ja/või vabavaralised sisuhaldussüsteemid ning hindab ka kommertslike või kommertstoega tooteid või lausa DXP (*Digital Experience Platform i.k.*) rakendusi jõuab autor seisukohale, et need on põhimõtteliselt ebasobivad ARVA otstarbeks.

4.2.1 Fookus on ebavajalikel funktsionaalsustel

Wiki-põhised ja levinud CMS lahendused on loodud ennekõike silmas pidades sisu kuvamist, kollaboratiivset muutmist ja lõppkasutajale info kättesaadavaks tegemist [25]. See sisaldab funktsionaalsusi nagu visuaalsed stiilid ja teemad, kommenteerimine, meedia ja sotsiaalmeedia haldus. ARVA kontekstis on tegemist üleliigsete ning potentsiaalselt lihtsalt keerukust lisavate võimekustega [26].

Ka ebavajalikud funktsionaalsused vajavad pidevaid uuendusi ja hoolt sarnaselt kasulikule koodile. Nende komponentide mitte hooldamise tagajärjeks on laienenud ründevektorid ja kõrgem turvarisk [27]. Väheste ressursside olukorras, kui RIA'l tekib vajadus seda tarkvara ise hallata, hõivab see ressursse mida võiks rakendada põhifunktsioonide täiendamiseks [24].

Väikesest hulgast või üksikust arendajast sõltuvad vabavaralised projektid kätkevad endas toimepidevuse riski. Nagu ka Wiki.js näitel näha, on tegemist aeglase evolutsiooniga toodetega kus uuendatud versiooni tarneajad aina pikenevad. Sõltumine sellistest projektidest riigi kesksete süsteemide loomisel lisab turvariski mida on keeruline hallata [22] [27].

Erinevalt kommertstoega vabavaralistest toodetest kus sääraseid riske on võimalik lepinguliselt hallata (tootja võtab uuendamise vastutuse enda peale) või rätseplahendusena tellitud tooted kus toimepidevuse plaani saab toote elutsükli osana ette arvestada.

Suurte ja suletud lähtekoodiga või kommertslikult pakutavate toodete puhul on endiselt probleemiks suur hulk ebavajalikku funktsionaalsust. DXP'de ja dokumendihaldustarkvarade näol on tegemist väga võimekate ning laia kasutusvaldkonnaga tarkvaraga. Selle rakendamine spetsiifilise süsteemi nagu ARVA

kontekstis ei ole see millele need süsteemid optimeeritud on. Ka oleks suletud lähtekoodiga tarkvara kasutamine vastuolus digiriigi tulevikuarhitektuuri suundadega [28].

4.2.2 Jäigad andmestruktuurid

Wiki-põhised ning levinud CMS lahendused lähtuvad üldjuhul spetsiifilistest andmemudelitest või formaatidest oma sisu salvestamisel. Sealjuures on tihti probleemiks justnimelt metaandmete ja semantilise andmestiku salvestamine.

Nende struktuuride mugandamine viisil, et see imiteeriks keerulisemaid hierarhilisi mudeleid võib põhjustada ulatuslikke ja laialivalguvaid kohandusi või semantiliselt keerulist dokumentatsiooni, kus rakenduse sisemise loogika mõisted ning nende tegelik kasutus ARVA kontekstis lahknevad.

See läheb vastuollu domeenijuhitud disaini põhimõtetega kus hallatavuse tagamiseks on oluline piiratud kontekst (*bounded context i.k.*) ja ühtne arusaam terminitest (*ubiquitous language i.k.*) [29]. Näiteks võib juba ainuüksi terminil „artikkel“ olla väga erinev äriiline tähendus Rahvusringhäälingu konteksti ning rahvusvahelise õiguse konteksti vahel.

Lähtudes eeldusest, et andmed on kasutusel ühe rakenduse siseselt on arendamise ja haldamise hõlbustamiseks mõistlik andmemudel koostada vastavalt rakenduse otsesest vajadusest, mitte andmete migreeritavuse või riskasutuse aspekti silmas pidades.

Sama oht on ka rätseplahenduste puhul kuid seal on see risk võimalik maandada seades vastavad mittefunktsionaalsed nõuded andmete, metaandmete ja semantilise andmestiku salvestamisele mille läbi andmed salvestamise viis ja struktuur seotakse lahti rakendusest millega neid hallatakse.

4.2.3 Piiratud rollide ja õiguste haldus

Kuigi enamik Wiki ja CMS lahendusi pakuvad erineval tasemel ligipääsukontrolli võimalusi ei ole see piisavalt võimekas tagamaks üleriigiliselt ja erinevate organisatsioonide poolt hallatud info jaoks vajalikku õiguste ja ligipääsude granulaarsust. Eriti olukorras kus tulevikus on mõistlik siduda väliste osapoolte rollid ja privileegid riigi keskse volituste haldamise infosüsteemiga [30].

Säärase paindlikkuse saavutamine nõuab taaskord arvestatavaid mugandusi mis on valmistoodete puhul kohmakad ning vea-altid.

4.2.4 Semantilised piirangud

Wikid ja CMS'id on üles ehitatud ümber kontseptide nagu üksteisega seotud „lehed“ või „postitused“ mis seavad ebaloomulikud piirangud artiklivaramule kui avaliku teabe talletajale ning sünkroniseerijale üle erinevate asutuste ning kontekstide.

4.2.5 Järeldus

Wiki-põhised ja CMS valmislahendused ei ole praktilised ARVA spetsiifiliste vajaduste katmiseks.

4.3 GIT

GIT on vabavaraline hajus versioonikontrolli süsteem (DVCS ehk *Distributed Version Control System i.k.*). Tavaliselt kasutusel tarkvaraarenduses, võimaldab see mitmel arendajal töötada sama koodibaasiga jälgides ja hallates muudatusi ajas. GIT kasutab mitmeid arhitektuurilisi ja algoritmilisi kontsepte mis teevad selle vägagi sobivaks, valdavalt teksti kujul esitatud, info talletamisel [31].

ARVA kontekstis on olulisemad nendest:

- **Hajus arhitektuur** – erinevalt tsentraliseeritud versioonikontrollisüsteemidest on GIT hajutatud, mis tähendab, et iga panustaja (*contributor i.k.*) omab terviklikku koopiat kogu repositooriumist (versioonikontrolli termin andmekogule mida versioneeritakse) sh. muudatuste ajalugu. Lisaks kollaboratsioonile tagab see toimepidevuse andmekao vastu. Iga repositooriumi kloon on täiemahuline varukoopia viimase sünkroniseerimise ajahetkega. GIT võimaldab asünkroonset või täielikult võrguvälist (*i.k. offline*) andmete haldamist [31].
- **Hetketõmmised ja olekud** – GIT olemuslik funktsionaalsus on hoida andmete ajalugu kui ajajoont hetketõmmistest (*snapshot i.k.*) ja olekutest (*state i.k.*). Iga GIT muudatus mida Inglise keeles nimetatakse *commit*, representeerib terviklikku hetketõmmist repositooriumi kõigist failidest selle ajahetkel. Erinevalt

varasematest versioonikontrolli süsteemidest ei tööta GIT inkrementaalsete muudatustega [32] [31].

- **Räsid** – GIT kasutab krüptograafilisi algoritme (varasemalt SHA-1 nüüd SHA-256), et luua unikaalne tuletatud räsi (*hash i.k.*) iga muudatuse ja iga objekti kohta (nn. failid, kaustad, sildid). Kuna need räsid arvutatakse eelnevate räside põhjal tagab GIT repositooriumi ajaloo muutmatuse (*immutability i.k.*), välistades ajaloo muutmise eksituse või pahatahtlikkuse tõttu [31].
- **Suunatud atsükliline graaf** – DAG (*Directed Acyclic Graph i.k.*) on matemaatiline kontsept graafi teoorias mida GIT kasutab salvestamiseks eellane-tuleneja (*parent-child i.k.*) suhteid muudatuste vahel. See struktuur võimaldab funktsionaalsusi nagu hargnemine (*branching i.k.*) ja mestimine (*merging i.k.*), mis omakorda võimaldab erinevatel osapooltel töötada samaaegselt repositooriumi samade objektidega ilma üksteist mõjutamata [32].

GIT põhjalik versioonikontrolli võimekus tagab muudatuste ja nende ajaloo suurepärase jälgimise. See võimaldab ARVA muutusi mitte ainult auditeerida vaid ka jälgida ning analüüsida artiklite ja sisu evolutsiooni.

GIT pakub väga selget viisi kuidas andmeid erinevate infoomanike andmekogudes uuendada ja sünkroniseerida. Allikas (*i.k. origin*) tehtud muudatused on võimalik evitada kloonidele automaatselt, ilma, et oleks vajadus sõltuda manuaalsest uuendamisest või teavitustest [32].

4.3.1 GIT puudused ARVA kontekstis

Hoolimata ARVA poolt võimaldatud abstraktsioonist võib GIT'i operatsioonide enda keerukus siiski põhjustada täiendavat koormust süsteemiadministraatoritele olukorras kus automaatse mestimisega esineb probleeme.

GIT repositooriumitel võivad tekkida jõudlusprobleemid suure hulga failide ning mahuka muudatuste ajaloo puhul [32]. Kuigi ARVA sisu teksti ja metaandmete salvestamisel see probleemiks ei osutuks siis kuna GIT ei ole optimeeritud haldamiseks binaarfaile (pildid, PDF dokumendid jne.) mis on mõnikord artiklite osaks, nende salvestamine repositooriumis võib jõudluse probleemi sellegipoolest tekitada.

4.3.2 GIT funktsioonid üle x-tee

Analüüsi käigus uuris autor ka GIT funktsionaalsuste kättesaadavaks tegemist X-Tee vahendusel. Kuigi see oleks kontseptuaalselt võimalik siis kaasneksid sellega mitmed mõlema tehnoloogia jaoks ebakonventsionaalsed seadistused.

Andmevahetusmudel – X-Tee on üles ehitatud teenusorienteeritud arhitektuurile (*Service Oriented Architecture i.k.*) [34] kus andmeid vahendatakse järgalt defineeritud päring-vastus mustris. GIT operatsioonide edastamiseks tuleks need operatsioonid kapseldada konkreetse teenuse päringuteks ning vastusteks mis eeldaks adapterrakenduste olemasolu nii teenuse tarbija kui pakkuja poolel ja tähendaks sisuliselt GIT funktsionaalsuse duplitseerimist X-Tee teenustena.

- **Jõudlus ja efektiivsus** – GIT operatsioonid võivad vahetada arvestaval hulgal andmeid. Kuna ARVA sisus¹ pole midagi mis vajaks ilmtingimata X-Tee pakutavat täiendavat turvalisust ja auditeerimist ei ole otstarbekas X-tee platvormi koormamine artiklite sünkroniseerimiseks Koodivaramuga.

Autori hinnangul ei looks see täiendav keerukus piisavat väärtust ning lihtsam ja praktilisem on siduda ARVA lokaalsed repositooriumid, koodivaramus asuva keske ja avalikult kättesaadava repositooriumiga.

4.3.3 Järeldus

GIT kasutamine riigi artiklite tehnilise salvestus- ja sünkroniseerimismehhanismina on jätkuvalt asjakohane. GIT'i detsentraliseeritus, auditeeritavus ning vabavaraline olemus joonduvad Digiriigi Tulevikuarhitektuuri suundadega.

GIT on laialt levinud ja hästi mõistetud tehnoloogia ning sellele ehitatud hoiustusmehhanism jätab võimaluse ARVA alusandmetega liidestada ka süsteeme mida ARVA praegune meeskond või käesoleva töö autor hetkel ette ei näe.

ARVA repositooriumi GIT funktsionaalsuste kättesaadavaks tegemine X-Tee vahendusel pole mõistlik.

¹ Tegemist on avaliku teabega

4.4 Duplikaatide tuvastamine kasutades plagiaadituvastustarkvara

ARVA meeskond on püstitanud hüpoteesi, et artiklite info duplikaatide tuvastamiseks võiks kasutada mõnd levinud plagiaadituvastustarkvara. Järgnevalt analüüsib autor seda hüpoteesi funktsionaalse ja mittefunktsionaalse sobivuse aspektist.

4.4.1 Funktsionaalne sobivus

Plagiaadi tuvastamiseks mõeldud tarkvara on reeglina loodud akadeemiliste tekstide kontrollimiseks, eesmärgiga tuvastada mitte-originaalset loomingut. Avaliku sektori dokumendid kasutavad tihti standardiseeritud keelekasutust, õiguslikke termineid ning korduvalt taaskasutatud formaate. Sealjuures pole avaliku sektori dokumentides kombeks viiteid lisada.

Avaliku sektori tekstide formaalse ja tihti korduva tekstikasutuse tõttu vajaks traditsiooniline plagiaadituvastustarkvara hoolikat seadistamist või isegi modifikatsioone, et vältida vale-positiivseid tuvastusi.

Kommertsiaalselt toodetud plagiaadituvastusmehhanismide üheks väärtuspakkumiseks on liidestus suurte teadustöö ja loomingu andmebaasidega, mis riigi artiklite sarnasuse tuvastamisel lisandväärtust ei loo.

Autor ei suutnud leida ühtegi kommertslikku valmislahendust mis pakuks peaaesjalikult tekstides sisalduva duplikaatinfo tuvastamist.

4.4.2 Mittefunktsionaalne sobivus

Riigi peaarhitektide töögrupp sõnastas kolm aastat tagasi põhimõtte vältida sõltuvusi kindlate seotud tehnoloogiatega. Tsitaat dokumendist Digiriigi Tulevikuarhitektuuri Suunad: „*On tähtis vältida sõltuvust ühte kindlasse partneritesse ja seotud tehnoloogiatesse. Vastasel juhul tekib olukord, kus partner dikteerib üksinda hinda ning muutub raskemaks tehnoloogiamigratsioon. Kontrollküsimus: Kas meie kasutatavad rakendused ja nende koodi on meil võimalik edasi arendada ka teiste partneritega? Kas meie rakendused on meie jaoks avatud lähtekoodi litsentsidega (MIT License, etc).*“ [28].

Autori hinnangul eksisteerib vähe avatud lähtekoodiga plagiaadikontrolli tööriistu, leitud näidete funktsionaalsused on äärmiselt piiratud ning nende rakendamine ilma täiendava tarkvaraarenduseta ei ole võimalik isegi esmaseks kontseptsiooni toetuseks.

Sherlock [35] – Räsipõhine algoritm mis on sobilik kahe lihtteksti võrdlemiseks ning vajab eraldi implementatsiooni laiema hulga dokumentide töötlemiseks ja ei toeta muid algoritme.

Plagiarism-Checker [36] – Sarnaselt Sherlockile on tegemist elementaarse tekstivõrdlus tööriistaga mis võrdleb kaht dokumenti omavahel. Mõõndustega on võimalik tööriista kasutada ka piiratud kujul parafraseeritud sisu tuvastamiseks.

WCopyfind [36] – On üks vanemaid tekstivõrdlustööriistu mis on loodud ainult Windows'i keskkonnas kasutamiseks ja baseerub suuresti tekstide sõna-sõnalt võrdlemisele.

JPlag [38] – fokuseeritud programmikoodi analüüsile ja programmikoodis plagiaatide leidmisele. JPlag põhiliseks väärtuspakkumiseks on hea toimekindlus obfuskaatsiooni¹ vastu, mis saavutatakse läbi toetatud programmeerimiskeelte spetsiifilise süntaksi analüüsi. Paraku ei ole see meetod sobilik loomuliku keele töötlemiseks.

4.4.3 Järeldus

Autori hinnangul muudavad funktsionaalsed ja mittefunktsionaalsed ebasobivused plagiaadikontrollitarkvara valmistoodete kasutamise ebamõistlikuks võrreldes rätseplahendusena valminud tarkvaraga.

4.5 Hinnang ARVA hetkeseisule

Autori hinnangul, mis baseerub isiklikul kogemusel ARVA arendamisega aastatel 2020 ning 2021, aastal 2024 evitatud lahenduse dokumentanalüüsil ning intervjuudel ARVA meeskonnaga, tulenevad praeguse lahenduse põhilised probleemid kahest juurpõhjusest:

1. ARVA kui kontsept kujunes välja riigiportaali vajadustest lähtuvalt.

¹ Koodi modifitseerimine viisil, et see ei oleks loetav või tunduks erinev inimkasutajale kuid käsud kompilaatorile või rakendusserverile oleksid identsed obfuskeerimata koodiga.

Süsteemi ja kontspetsiooni loomise motivaatoriks oli vajadus riigiportaalis kuvada, ühtse visuaalse stiiliga ning ühtse kategoriseerimise põhimõttega, teiste asutuste vastutusalas olevat infot.

See tingis näiteks mõiste nagu „kaust“ või „kataloog“ (kasutatakse tihti sünonüümina) tekkimise ARVA ärikeeles. Kausta mõiste praeguses domeenis on veidi üle-laetud ning võib tähendada nii infoomaniku põhise jaotust mis grupeerib ühe asutuse või valdkonna info ning riigiportaali menüüpunktide põhise jaotust mis grupeerib infot riigiportaalis vastavalt selle asukohale. Mõiste kahe kasutuse vahel on enamjaolt seos kuna riigiportaal valdavalt lähtub teemapõhisest grupeerimisest.

Conway seaduspära [38] ütleb, et organisatsiooni süsteemid ja protsessid peegeldavad tihti organisatsiooni kommunikatsiooni ja koostöö vormi. Selle edasiarendus infosüsteemidele [39] oleks, et infosüsteemi osad ja funktsioonid kipuvad peegeldama organisatsiooni osi ja funktsioone.

Riigiportaali olid sisutoimetajad kelle eesmärgiks oli koondada, läbi vaadata ja kinnitada (viimane autoriteet enne avaldamist) kogu info mida riigiportaalis kuvatakse. Samas ei olnud riigiportaali toimetajad (üksikute eranditega) info tekkeallikaks. Artikleid koostasid (ja koostavad ka praegu) vastutavate asutuste töötajad, tuginedes asutuste sisestele protsessidele, kasutades asutuste siseseid tööriistu.

ARVA senist arengut on juhtinud RIA organisatsioonisisene kommunikatsioonimudel.

2. Kontsepti keskmes on sisuhaldus mitte info haldus.

Sisuhaldus on laiem mõiste kui info haldus. Info on andmed koos metaandmetega (andmetega andmete kohta) mis on organiseeritud ja töödeldud viisil, et neil oleks tähendus ning väärtus. Sisu (*i.k. content*) on sarnaselt infole, organiseeritud andmete ja metaandmete kogum ent sellel on täiendav funktsioon, kanda teatud kindlat sõnumit. Kui info lisab andmetele konteksti siis sisu lisab infole sõnumi.

Nii RIA kui teabevaldajate poolt on süsteemi põhilisteks kasutajateks sisuhaldusega tegelevad inimesed, nõunikud, ametnikud, toimetajad. Süsteemi esmane disain lähtub nende kasutajate vajadustest ning nii oli algselt põhjendatud tavapäraste sisuhaldustööriistade kasutusele võtmine.

Kuna tegemist on valdavalt mitte tehniliste kasutajatega pidi mistahes lahendus järgima WYSIWYG (akronüüm inglisekeelsest terminist „*what you see is what you get*“) põhimõtet kus sisu saab hallata nii, et see on kuvatud sarnaselt nagu see paistab lõppkasutajale.

See aga tähendab väga jäika eraldatust info allika ja ARVA vahel. Sisuliselt on ARVA infoomaniku vaates lihtsalt riigiportaali sisuhaldusrakendus. Enamikel teabevaldajatel polnud sellele isegi juurdepääsu ning alles 2024 aasta esimeses kvartalis, RIA poolt evitatud lahendus tekitas teabevaldajale juurdepääsu ARVA sisuhaldusele. Teabevaldajate vajadus ARVA's infot uuendada tuli peasjalikult vajadusest hoida ajakohasena enda vastutusallas olev info riigiportaalis.

Ühtse digivärava määrus [1] tugevdas seda vajadust veelgi, sätestades (muuhulgas) üle-Euroopalise infoportaali *Your Europe*¹ loomise mis ühendab riiklike portaalide võrgustiku. Sama määrus kehtestas Euroopa Liidu liikmesriikidele kohustuse riiklike portaalide loomiseks.

2020 aasta 12. detsembriks oli liikmesriikidel kohustus käivitada ühtse digivärava portaalid kust Euroopa Liidu kodanikud saavad kergelt leida kogu informatsiooni Euroopa Liidu ja siseriiklike regulatsioonide ning protsesside kohta mis puudutavad mitmeid isiklike või tööalaseid sündmusi nagu reisimine, pension, haridus, tööhõive, tervishoid, jne [1].

Määrus seadis ka konkreetseid tulemuseesmärke ja mõõdikud. Tulemuseesmärke olid põhiliselt seotud sellega kui kiiresti ja mugavalt kasutajad info leiavad ning kui täpne see on. Mõõdikuteks oli põhiliselt kasutajate tagasiside ning infoteenuste kasutusaktiivsus. Kokkuvõtvalt võib öelda, et mõõdikud olid otsisid kasutajalt vastust küsimusele: „*Kas leidsid mida otsisid?*“ [1].

Eesti oli tänu juba eksisteerivale ning aktiivselt arendatavale riigiportaalile SDG määrusega kooskõlla jõudmiseks üsna heas seisus, samas paremat lahendust vajab justnimelt info sünkroniseerimine riigiportaali ning infot omavate asutuste vahel.

¹ <https://europa.eu/youreurope/index.htm>

Lisaks riiklikule strateegiale oli nüüd ka jõustunud määrus mis seadis infoomanikele kohustuse infot lisaks enda sisuhaldusele ka riigiportaali sisuhalduses asjakohasena hoida.

Muutunud olukorrast arusaamist riiklikult ja RIA tasandil peegeldas initsiatiiv luua 2020 aastal kontseptsiooni tõestus (*Proof of Concept i.k.*) ARVA'st kui kesksest avaliku info hoidlast¹. Erinevalt varasemast nähti ARVAs võimalust integreerida informatsioon üle mitme siseriikliku portaali ning kasutades riigiportaali liita see Euroopa raamistikuga. Initsiatiiv seadis senisest suuremat rõhku protsesside automatiseerimisele ning info täiendavale kategoriseerimisele ja metaandmetega rikastamisele.

4.5.1 Järeldus

Praegu evitatud lahendus, baseerub endiselt fundamentaalselt puudlikel ning aegunud lähtepunktidel mis takistavad seatud strateegiliste eesmärkide saavutamist infoallikate koostoime ja ajakohasuse osas.

Ka ei joondu valitud arhitektuur Eesti digiriigi aluspõhimõtetega teabe haldamise osas kus info hõivatakse ja hoitakse selle tekkekohas ning seda ei duplitseerita.

Kuigi õiguslik analüüs ja õiguslike soovitude andmine jääb käesoleva magistritöö käsitlusalast välja siis ARVA puhul võiks autori hinnangul keegi siiski analüüsida andmekogu loomise võimalust. ARVA käsitlemine andmekoguna tähendaks ühtlasi vastava põhimääruse koostamist ja seadustamist, mida saaks omakorda kasutada õigusliku aluse ning raamistikuna ARVA liidestumiseks teabevaldajatega. ARVA põhimääruses saaks seada tingimuse edastada koopia avalikust teabest andmekogusse, sõnastada täpsemaks teabeomanike õigused ja kohustused jne.

¹ Autori isiklikul kogemusel põhinev info.

5 Lahendus

Autori hinnangul on ilmnunud probleemidele lahendamiseks ning strateegia- ja motivatsioonimudelid (Joonis 5) sõnastatud tegevussuuna „ARVA süsteemi ajakohastamine“ realiseerimiseks vajalik alustada ARVA visiooni ümber hindamisest. Autor pakub sõnastusena välja järgmise visiooni:

**ARVA on avaliku sektori teabehaldussüsteem,
mis tagab ajakohase info riigi kõigis digikanalites.**

Sellest visioonist lähtuvalt on ARVA primaarne eesmärk parandada avaliku sektori poolt kodanikule kättesaadavaks tehtava informatsiooni ajakohasust ja ühtsust üle erinevate digitaalsete kanalite ning avalikustamise viiside.

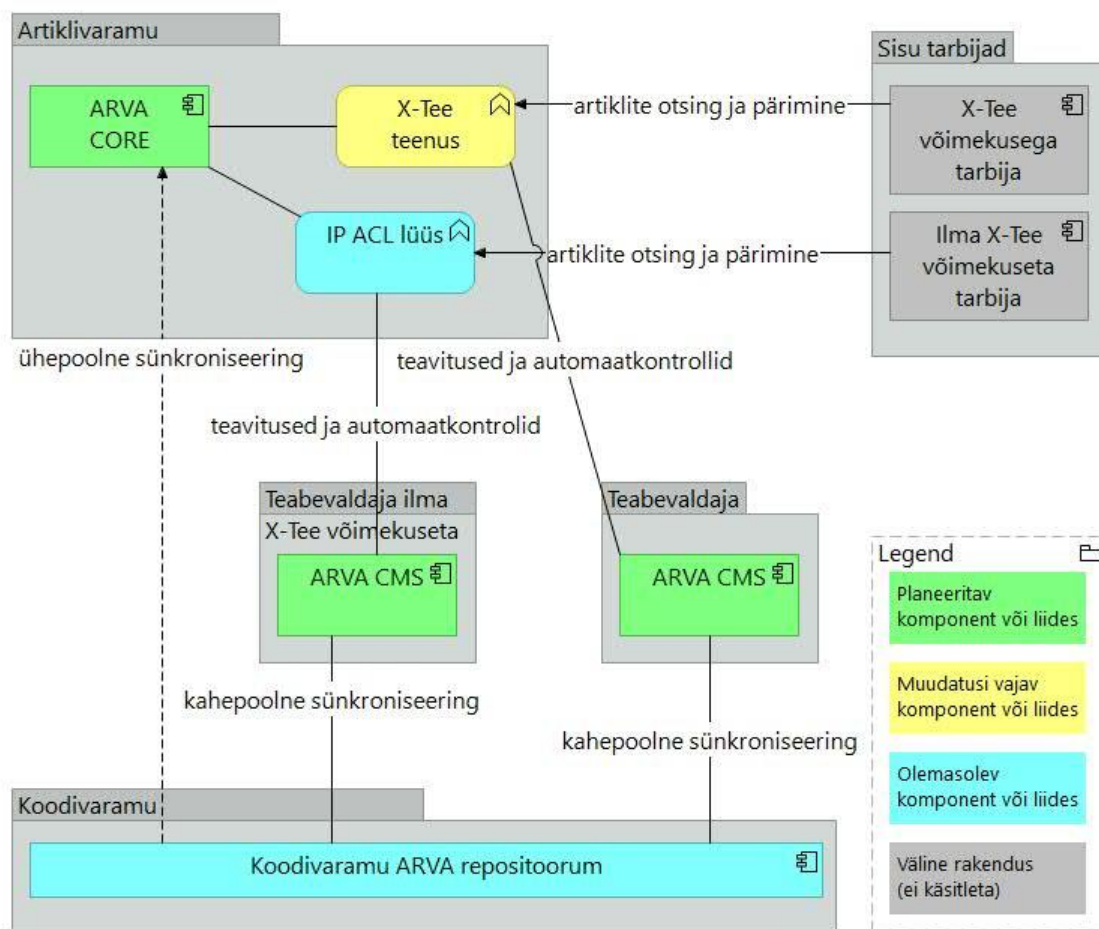
Selle eesmärgi saavutamiseks peab ARVA omama järgmisi üldisi funktsionaalsusi:

1. Pakub masinliidestatavaid otspunkte teabe edastamiseks, ajakohasuse kontrollimiseks ning teabe küsimiseks;
2. Salvestab, organiseerib ning analüüsib koondatud teavet.
3. Käitub proaktiivselt duplikaatinfo ja võimaliku valeinfo tuvastamisel.

Selliselt sõnastatud visioon, eesmärk ja defineeritud funktsionaalsused.

Oluline on autori hinnangul selliselt defineeritud visiooni ja eesmärgi juures see, et ARVA't ei käsitletaks sisuhaldusplatvormina

5.1 ARVA tehniline ülevaade (TO-BE)



Joonis 8. ARVA rakenduse TO-BE kontseptvaade (autori koostatud)

ARVA CORE – keskne rakendus mida arendab, haldab ja evitab ARVA meeskond. Keskne rakendus ei sisalda endas sisuhalduse komponenti. ARVA CORE sünkroniseerib artiklite HTML sisu Koodivaramus asuvast Artiklivaramu GIT repositooriumist ühepoolselt (ainult sisse lugemine).

ARVA CMS – avatud lähtekoodiga rakendus mille abil teabevaldaja saab hallata enda vastutusalas olevat sisu. Rakenduse arendamisega tegeleb ARVA meeskond ent evitus ja käitlemine on teabevaldaja vastutusalas. Juhul kui teabevaldajal vastav võimekus puudub on paigalduse ja käitlemise osas võimalik sõlmida kokkulepe ka RIA'ga. ARVA CMS rakendus sisaldab endas sisuhalduskomponenti ning kasutab artiklite HTML sisu sünkroniseerimiseks, kahepoolset sidet Koodivaramus asuva Artiklivaramu GIT repositooriumi.

Mõlema rakenduse puhul aga eriti ARVA CMS puhul tuleb lähtuda doomenjuhitud disaini põhimõtetest ja mitte segada omavahel erinevate domeenide kontekste. See tähendab, et juhul kui üks asutus (n. RIA) paigaldab ja käitleb mitme erineva omaniku CMS rakendust tuleb igale omanikule paigaldada eraldiseisev instants.

Sarnaselt ei tohiks üritada luua keskses rakenduses universaalset taksonoomiat artiklite ja nende vastutavate kasutajate struktureerimiseks. See annab igale teabevaldajale võimaluse ise privileegid seadistada ning kontekstid luua vastavalt asutuse ja teabe eripärale. Tulevikus võimaldab see teabevaldajal defineerida enda vastutusala teabevalduse rollid keskses volituste halduses [30] ja siduda need privileegidega ARVA's.

Mistahes artiklite grupeerimine või organiseerimine kaustapuud kasutades ei tohi olla artikli kuuluvuse määramise või ligipääsuõiguste aluseks. Hierarhiliste struktuuride ja semantiliste andmekogude loomise võimekus peab olema lahendatud metaandmete ja siltide lisamise abil mis lubab teabevaldajal ise enda vastutusallas olevaid artikleid grupeerida.

Ainus erand sellele reeglile ja vaike privileeg mis on seadistatav ARVA meeskonna poolt on artikli universaalse identifikaatori sidumine teabe eest vastutava asutuse identifikaatoriga.

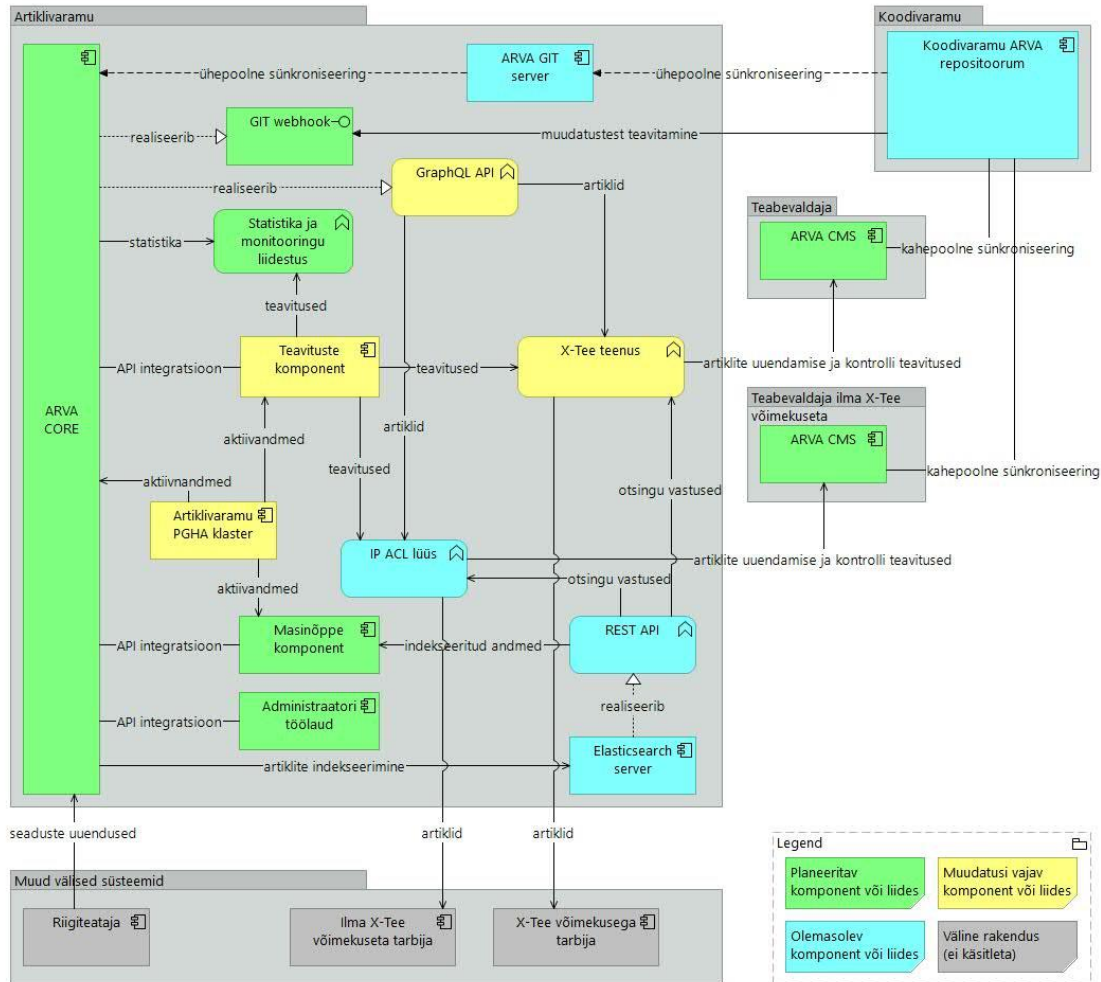
ARVA X-Tee teenuseid on vaja täiendada osas mis puudutab teabevaldaja seost artikli uuendamise protsessidega (kirjeldatud järgnevas peatükkides) ning tema vastutusallas oleva info duplikaatide esinemisest ülevaate andmist ja regulaarsete teabe ajakohastamise teavituste edastamist.

RIA vastutusallas olevaid artiklite sisu tarbijaid või omanikke (nn. Riigiportaal või Sündmusteenused) võib liidestada artiklivaramuga asutuse siseselt ainult ulatuses kus nimetatud tarbijad kasutavad samu teenuseid mis muidu on suunatud väliskasutajatele läbi X-Tee turvaserveri või IP ACL lüüsi.

RIA vastutusallas olevaid rakendusi ei või liidestada ARVA'ga luues neile teistest tarbijatest või info allikatest erinevaid otspunkte.

5.2 ARVA CORE

Alljärgnevalt (Joonis 9) kirjeldab autor ARVA tuumikrakenduse TO-BE liidestusi üldisel tasemel.



Joonis 9. ARVA CORE liidestuste komponentjoonis, TO-BE, C1 tasemel (autori koostatud)

ARVA CORE komponendi põhiline funktsioon on üle võtta Wiki.js need funktsionaalsused mis puudutavad GraphQL API otspunktide abil realiseeritud liidestusi väliste süsteemidega (sh. X-Tee teenused), artiklite indekseerimist Elasticsearch serverisse, Riigiteataja sisendit ning statistika ja tagasiside edastamine RIA statistika moodulile.

Märgiline erinevus varasema implementatsiooni ja Wiki.js'iga on selles, et kontseptuaalselt ei ole ARVA CORE enam sisuhalduskomponent, selles puudub nii funktsionaalsus kui kasutajaliides artiklite loomiseks, muutmiseks või kustutamiseks.

ARVA CORE ei sisalda funktsionaalsust väliste kasutajate loomiseks või sisse logimiseks muul viisil kui RIA AD abil. Ainus kasutajaliides on administraatori töölaud, mille abil saab juhtida kesk komponendi tööd. Administraatori töölaua täpsed funktsionaalsused ja kasutajaliides jäävad väljaspoole käesoleva magistr töö käsitusala ning tuleb täpsustada täiendava detailanalüüsi käigus.

Kui varasemalt ARVA varundas sisemiselt muudetud HTML lehti ARVA GIT serveris ning sünkroniseeris neid Koodivaramus asuva repositooriumiga siis väljapakutavas lahenduses selle suunalist suhtlust ei korraldata. ARVA CORE pakub otspunkti GIT uuenduste kuulamiseks Koodivaramus (*i.k. webhook*) ning reageerib teabevaldajate muudatustele Koodivaramu ARVA repositooriumis, ühepoolse andmete sünkroniseerimise ning duplikaatide automaatkontrolli käivitamisega (protsessi on lähemalt kirjeldatud peatükis 6.4.).

Joonisel 9 kujutatud masinõppe komponendi eesmärk on duplikaatide automaatkontrolli läbi viimine ning komponendi sisu on kirjeldatud peatükis 6.5.

Luu tuleb uus statistika ja monitooringu liidestus, mis võimaldaks ARVA CORE lahendusel liidestuda olemasoleva RIA taristuga ennekõike artiklite kasutusstatistika edastamisel sarnaselt varasemale funktsionaalsusele (artiklid mida päritakse GraphQL liidese abil) ning edastada statistikat duplikaatinfo ja välja saadetud teavituste kohta.

Olemasolevasse teavituste komponenti tuleb sisse viia muudatused mis võimaldavad varasema e-kirja saatmise asemel edastada teavitusi otse ARVA CMS infosüsteemidele. Teavituste edastamise mehhanismi kirjeldab autor lähemalt peatükis 4.2.1. Lisaks varasemale, regulaarsele ajatatud info üle vaatamise palvele, peab komponent suutma edastada ka teavitusi duplikaatinfo (või duplikaadi originaalallika) muudatustest, peatükis 6.4. kirjeldatud protsessi osalisena.

PostgreSQL kõrgkäideldava klasteri (komponent Artiklivaramu PGHA Klaster Joonisel 9) osas tuleb viia sisse muudatusi mis on seotud andmeolemite ja nende suhetega. PGHA klasteri kasutamise jätkamine on autori hinnangul mõttekas ning see tehnoloogia toetab kindlasti ARVA CORE muudetud funktsionaalsust. Olemite ja nende suhete kirjeldamine jääb väljaspoole käesoleva magistr töö käsitusala ja tuleb täpsustada täiendava detailanalüüsi käigus.

Olemasolevas X-Tee liideses tehtavad muudatused on seotud teavituste edastamiseks vajaliku teenuse loomisega ning automaattuvastuse tagasisidestamise teenuse loomisega.

Varasemalt Wiki.js osaks olnud GraphQL API liidese funktsionaalsus mida ARVA kontekstis masinliidestuseks kasutatakse tuleb reprodutseerida uue lahenduse raames viisil mis ei tingi välistes ja seni liidestatud süsteemides muudatusvajadusi. Sellegipoolest võib olla mõistlik kasutada võimalust GraphQL API liidese läbi võimaldatud funktsionaalsuste täiendavaks detailanalüüsiks.

Sisulisi muudatusi ei ole vaja sisse viia ARVA GIT serveri, IP ACL lüüsi ja Elasticsearch serveri ning selle REST API osas.

5.2.1 Teavituste edastamine ja tagasiside küsimine X-Tee või IP ACL lüüsi vahendusel

Järgnevalt kirjeldab autor mehhanismi kuidas peaks lahendama teavituste edastamise, senise e-kirjaga saatmise asemel, sõnumina X-Tee vahendusel.

RIA poolt keskselt hallatud ARVA CORE infosüsteem koos oma teavituste komponendiga saab pakkuda RIA turvaserveris juba olemasolevalt registreeritud „Artiklivaramu“ alamsüsteemi raames teenust, teavituste pärimiseks. Autori visioonis töötab teavituste süsteem seega info allikana mille poole teavituste saajad pöörduvad.

Alternatiiv oleks spetsiaalse X-Tee teenuse registreerimine kõigi teabevaldajate turvaserveris kes ARVA CMS rakendust kasutavad ning siis nende X-Tee liikmete, alamsüsteemide ja teenuste salvestamine ARVA CORE osana ja seostamine teavituste komponendis. Kus teavituste komponent saadab iga teavituse, proaktiivselt, spetsiifilisele ARVA CMS instantsile. Autor ei pea seda alternatiivi mõistlikuks.

Autori hinnangul lihtsustab teavituste RIA turvaserverist küsimine oluliselt kogu ARVA ökosüsteemi juurutamist ning on ühtlasi kooskõlas tehnoloogilise suunisega, kasutada sündmuspõhise arhitektuuri lahendusi [28]. Autori välja pakutud lahenduse puhul on teavituse tekitaja, sündmuse allikaks ning sündmusest huvitatud osapooled küsivad ise (vastavalt vajadusele ja sisemisele loogikale seadistatud korduvpärimise) endaga seotud teavitusi.

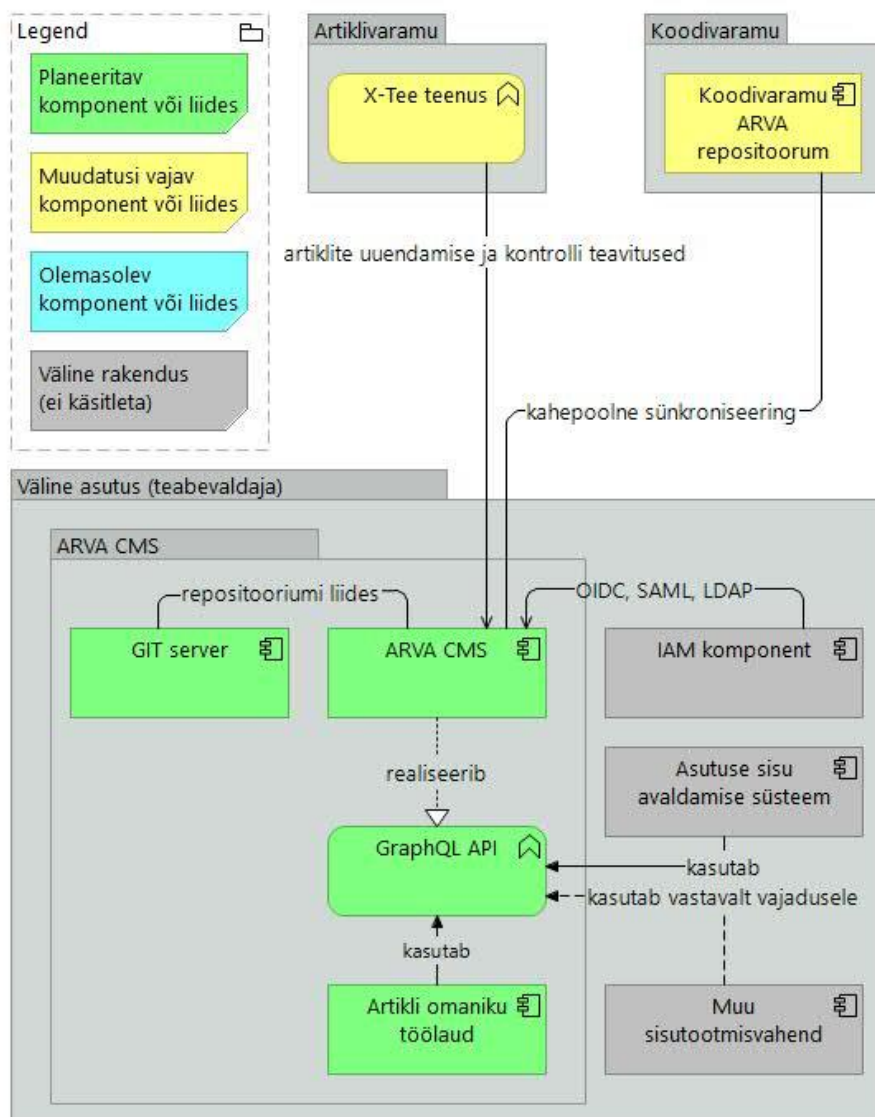
Lisaks teavituste teenusele tuleb luua teenus mille abil ARVA CMS rakenduse töölaualt antud tagasiside masinõppe põhisele duplikaatinfo automaattuvastusele jõuab ARVA CORE masinõppe komponenti ning seda saab kasutada mudeli efektiivsuse hindamisel ja kaalude korrigeerimisel.

Teabevaldajad kellel X-Tee võimekus puudub saavad samu funktsionaalsusi kasutada IP ACL lüüsi abil. Ka seal saab rakendada sama põhimõtet kus liidestatud ARVA CMS lahendused pöörduvad tunneli kaudu ARVA CORE teavituste komponendi poole ja teavituste komponent ei pea olema teadlik kõigi loodud tunnelite otspunktidest.

Mõlemal juhul vastusena tagastatava teavituste objekti formaat jääb käesoleva magistritöö käsitlusalast välja ja tuleb täpsustada täiendava detailanalüüsi käigus.

5.3 ARVA CMS

Alljärgnevalt (Joonis X) kirjeldab autor ARVA sisuhaldus ja integratsioonirakenduse TO-BE liidestusi üldisel tasemel.



Joonis 10. Pakutava lahenduse sisuhalduskomponendi joonis C1 tasemel (autori koostatud)

Autor ei too Joonisel X eraldi välja olukorda kus teabevaldajal puudub X-Tee võimekus, kuna ainus erinevus komponentides on artiklivaramuga liidestumise mehhanism, milleks on X-Tee asemel IP ACL ligipääs.

ARVA CMS komponendi põhiline funktsioon on võtta üle Wiki.js need funktsionaalsused mis puudutavad sisuhaldust.

Iga teabevaldaja kes ARVA CMS instantsi kasutab, säilitab enda kloonid Artiklivaramu repositooriumist, mis tagab info kättesaadavuse isegi siis kui ARVA keskne süsteem ei ole kättesaadav.

Asutused saavad luua ja kuvada sisust enda harusid ning ise kohandada või prioriteerida sisu täiendusi vastavalt vajadusele. Tänu GIT omadustele sünkroniseeritakse sellised harud automaatselt ka koodivaramusse mis aitab parandada ülevaadet riigi avaliku teabe kasutusest, asutuste lõikes.

Kogu sisuhalduse lahendus tuleks autori hinnangul hoida kompaktne ja võimalusel olekuvaba (*i.k. stateless*). See lihtsustab lahenduse evitamist mistahes keskkonda sh. pilvetaristud ning joondub tehnoloogilise suunisega hoida teenused kaasaegselt pilvekõlbulikud [28].

GIT serveriga peab ARVA CMS ühenduma läbi protokollipõhise liidese (implementatsioonid nagu näiteks JGit Java rakenduse puhul). See teeb võimalikuks ARVA CMS'i paigaldamise ja käitlemise ilma andmebaasi kasutamata. Artiklite HTML sisu muudatused salvestatakse otse GIT repositooriumisse.

Kui asutusel on juba olemasolev GIT server siis peab ARVA CMS komponendi saama liidestada sellega. ARVA CMS komponent peab olema GIT liidestuse osas seadistatav. Lisaks peab ARVA CMS seadistuses olema võimalik määrata mis nimelist repositooriumi haru kasutatakse.

Koodivaramuga liidestumiseks tuleb igale evitatud ARVA CMS infosüsteemile luua masinkasutaja ligipääs Koodivaramus asuvale Artiklivaramu repositooriumile. Iga muudatuse korral toimub automaatne mestimine lokaalses harus ning haru uuendamine Koodivaramus (GIT *commit* ja *push*).

ARVA CMS peab lubama teabevaldajal muuta ainult artikleid mis on tema vastutusallas (määramine toimub metaandmetes sisalduva artikli omaniku ID järgi). Kõik ARVA CMS abiga loodud uued artiklid omistatakse automaatselt asutusele kelle instantsis artikkel loodi. See välistab muudatused mis põhjustaksid mestimise konflikte GIT repositooriumis.

Lisaks sisuhaldusele peab ARVA CMS olema võimeline sooritama päringuid ARVA CORE teavituste komponendilt teavituste saamiseks ning neid enda kasutajaliideses kuvama. Teavituste edastamise mehhanism on kirjeldatud peatükis 6.2.1. Teavituste genereerimise protsessid on kirjeldatud peatükis 6.4.

ARVA CMS peab suutma edasta ARVA CORE masinõppe komponendile teabevaldaja tagasisidet duplikaatide tuvastamise mehhanismi efektiivsuse osas.

Mis ulatuses ja vormis, ARVA CMS ja seda toetav kasutajaliidesekomponent „Artikli omaniku töölaud“ peaks Wiki.js sisuhaldusfunktsioone üle võtma või välja nägema, jääb käesoleva magistritöö käsitlusalast välja ning tuleb täpsustada täiendava detailanalüüsi raames.

ARVA CMS peab võimaldama liidestumist asutuse siseste IAM süsteemidega, enamlevinud protokolle kasutades. See võimaldab teabevaldajatel lahendust kergemini oma töövoogu integreerida.

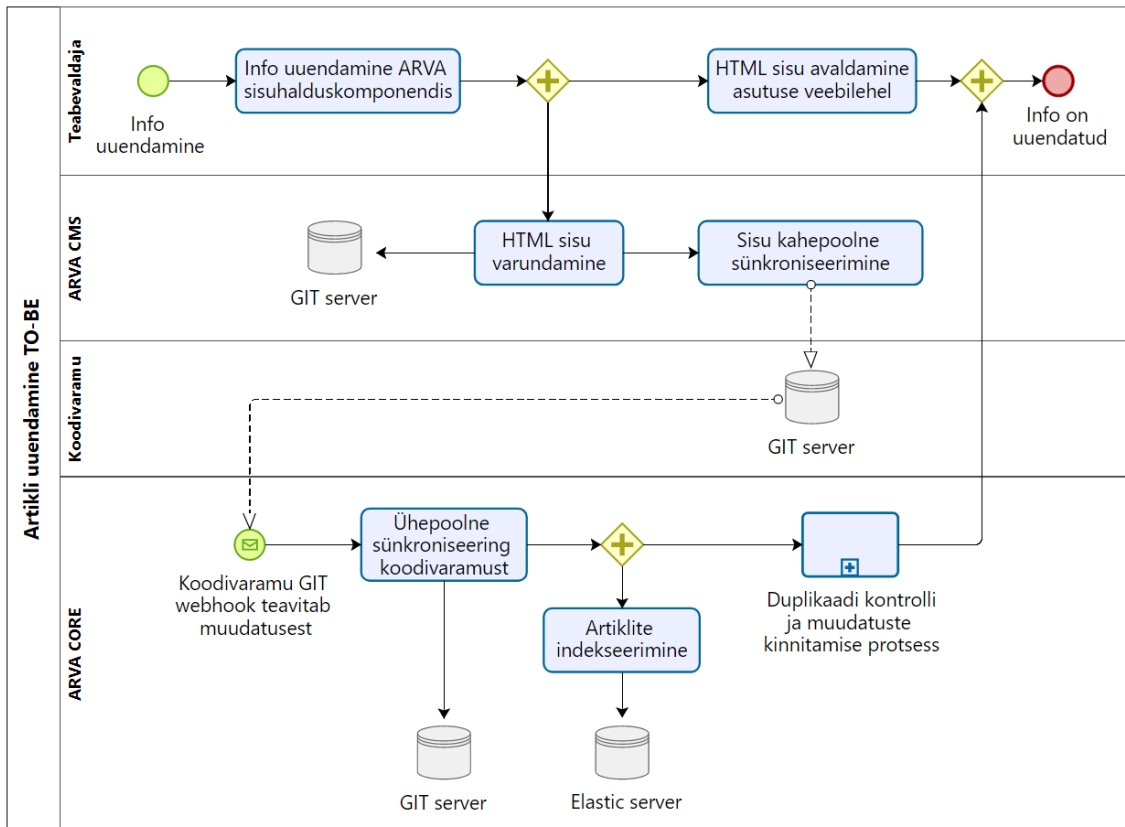
Artiklite HTML sisu kuvamiseks asutuse sisu avaldamise süsteemis peab ARVA CMS pakkuma GraphQL API liidest sisu päringuks ja otsinguks lokaalsest repositooriumist ning võimaldama sisuhaldusoperatsioone üle masinliidese. Selleks otstarbeks on mõistlik taaskasutada ARVA CMS artikli omaniku töölaua tarbeks loodud otspunkte.

5.4 Artiklite uuendamise protsess (TO-BE)

Artikli uuendamise protsess välja pakutavas lahenduses (ülevaatlikult kirjeldatud Joonisel 11), toimub teabevaldaja keskkonda (või tema tarbeks RIA keskkonda) evitatud ja teabevaldajaga seotud ARVA CMS instantsis.

Protsessi käigus toimub automaatselt HTML sisu varundamine, viiakse läbi duplikaatinfo kontrolli alamprotsess ja vajadusel muudatuste kinnitamise protsess.

Kui teabevaldaja liidestab oma asutuse sisu avaldamise süsteemi ARVA CMS masinliideselega on lõppkasutaja vaatest kogu protsess sujuv ning toimub tervikuna ühest kasutajaliideseist.



Joonis 11. Artikli uuendamise protsess ülevahtlikult, TO-BE (autori koostatud)

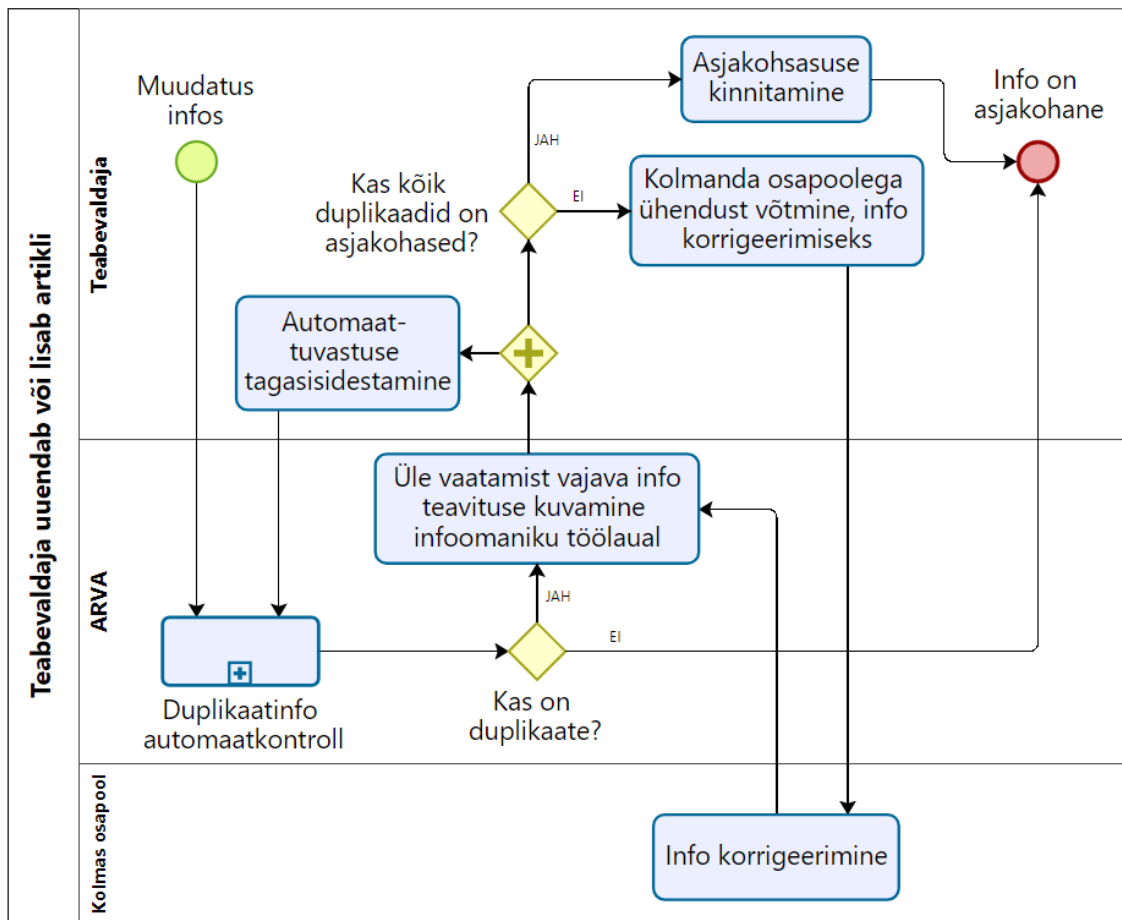
Järgnevalt (Joonis 12) kirjeldab autor olukorda kus teabevaldaja sooritab muutuse endale kuuluvas infos. Protsessi käigus viiakse läbi duplikaatinfo automaatkontrolli alamprotsess ja duplikaatide tuvastamisel suunatakse need teabevaldajale üle vaatamiseks (kasutades ARVA teavituste süsteemi).

Teabevaldaja saab duplikaadid kas kinnitada (juhul kui info on asjakohane) või võtta ühendust duplikaadi avaldanud kolmanda osapoolega, korrektuuride sisse viimiseks. Korrektuurid suunatakse läbi vaatamisele ilma täiendavalt automaatset duplikaadikontrolli sooritamata.

Olukord kus teabevaldaja ei ole nõus duplikaadi asjakohasust kinnitama aga ei soovi ka kolmanda osapoolega ühendust võtta või olukord kus kolmas osapool keeldub sobivaid muudatusi sisse viimast on autori hinnangul ootuspärased kuid kuna selliste olukordade lahendus vajab õiguslikku analüüsi jääb see käesoleva magistritöö käsituselast välja.

Protsess peatub sellisel juhul asjakohasuse kinnitamise sammus ning jääb ootele kuni osapooled jõuavad kokkuleppele ARVA väliseid protsesse kasutades. Protsessi

toppamine teabevaldaja liinis on taotluslik kuna teabevaldajal lasuvate õiguste ja kohustuste tõttu on lahenduse leidmine teabevaldaja vastutusalas.



Joonis 12. Artikli uuendamise protsess teabevaldaja vaatest, TO-BE (autori koostatud)

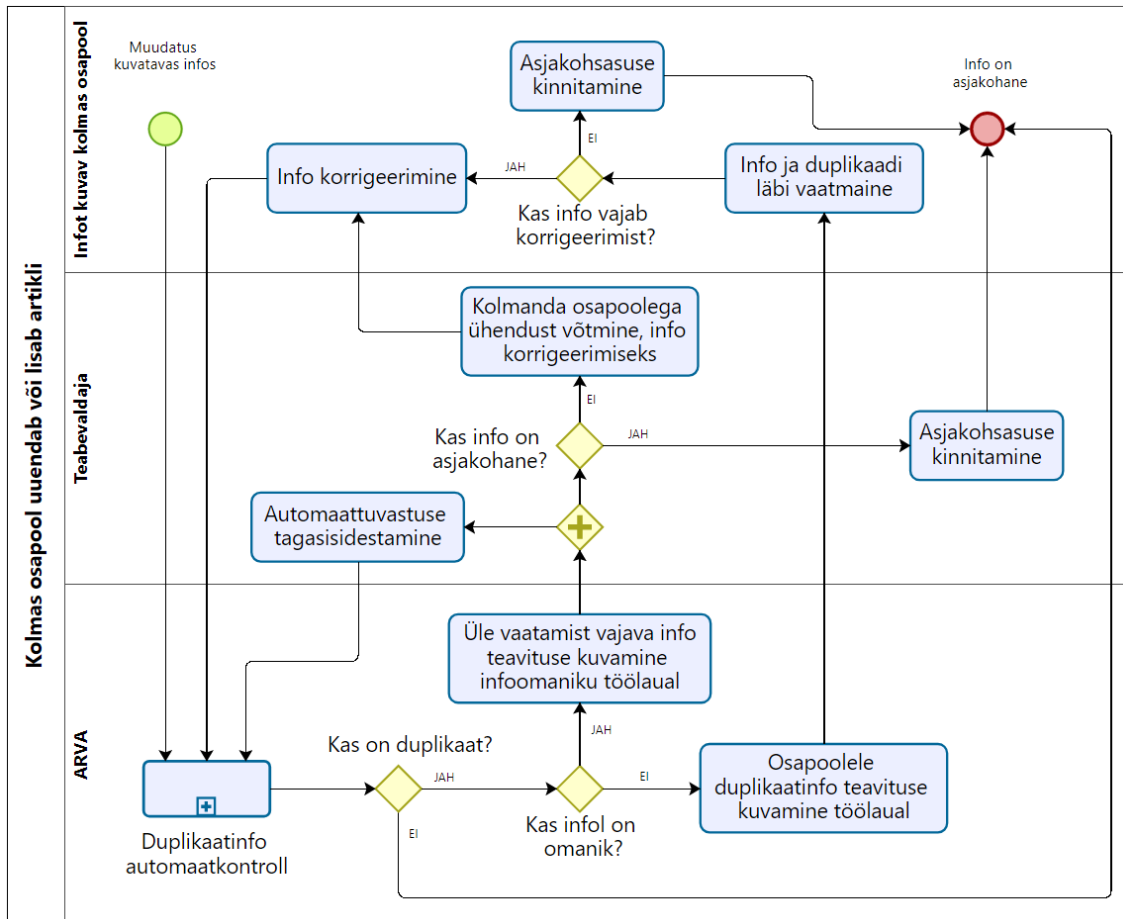
Järgnevalt (Joonis 13) kirjeldab autor olukorda kus infot muudab seda enda keskkonnas kuvav kolmas osapool kes võib aga ei pruugi olla teadlik, et infol on olemas teabevaldaja.

Sarnaselt teabevaldaja poolt algatatud protsessile viiakse esmalt läbi duplikaadikontroll ARVA poolt. Duplikaadi tuvastamisel hargneb protsess kahel moel. Kui duplikaadil on olemas teabevaldaja, suunatakse töövoog asjakohasuse kinnitamise protsessi mis on sarnane teabevaldaja poolt algatatud muudatusega. Kohustus ja õigus asjakohasust kinnitada on teabevaldajal.

Kui duplikaat tuvastati ent tegemist on infoga millel ei eksisteeri teabevaldajat siis suunatakse protsess tagasi muudatuse teinud kolmanda osapoole ARVA CMS töölauale (koos tuvastatud duplikaatidega) ning palutakse tal info kas kinnitada või seda korrigeerida. Kuna teabevaldajat ei eksisteeri siis puudub protsessi algatajal kohustus

tekkinud ebakõla lahendada ning ta saab igal juhul protsessi lõpetada info asjakohasuse kinnitamisega.

Kui duplikaate ei tuvastata, siis konteksti puudumise tõttu, lähtutakse eeldusest et info on asjakohane.



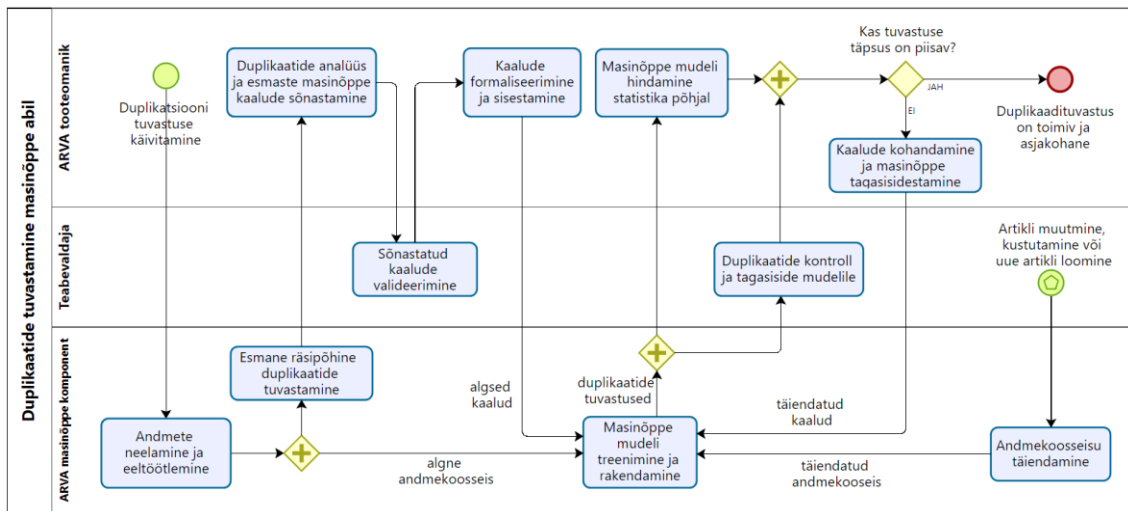
Joonis 13. Info muutmise TO-BE protsess kolmanda osapoole vaatest (autori koostatud)

5.5 Duplikaatide automaatkontroll

Masinõppe areng on viimasel aastakümnel hoogustunud ning selle tehnoloogiaid on praeguseks laialt levinud ja hästi mõistetud. Masinõppe lahenduste loomiseks eksisteerivad nii kommertslikud kui ka avatud lähtekoodiga aga kommertstoega tooted.

Kuigi konkreetsete tehnoloogiatega valik ja soovitamise jääb käesoleva töö käsitlusel välja ning tuleb täpsustada täiendava detailanalüüsi käigus siis indikaatselt võib autor, isikliku kogemuse põhjal, mainida tehnoloogiaid nagu Apache Spark andmete

neelamiseks, Simhash algoritm esmaseks räsi põhiseks duplikaatide tuvastuseks ning TensorFlow masinõppe mudeli treenimiseks ja käitlemiseks.



Joonis 14. Duplikaatinfo automaatkontrolli ja masinõppe protsess minimalistlikult (autori koostatud)

Kuna tegemist oleks, autorile teadaolevalt, esimese deentraliseeritud riiklike infoallikate duplikaadituvastuslahendusega siis lisaks DÜAK 2030 [11] sõnastatud ja RIA 21-25 [13] aluseks oleva eesmärgiga „Eesti täis digiväge“ joendumisega, aitaks see ühtlasi saavutada ka laiemalt EESTI 2035 [10] sõnastatud eesmärki „Eesti on edasipürgiv ja demokraatlik riik“, tugevdades Eesti Vabariigi kui digiriigi lahenduste novaatori kuvandit rahvusvahelisel areenil.

Keeletötluse algoritmide analüüs on võimekus mis hetkel RIA’l puudub ning selleks on vajalik keeleteadlaste kaasamine. Masinõppe kaalude sõnastamine ja teksti sisulise (mitte ainult vormilise) samasuse tuvastamiseks sobivate algoritmide valik on midagi mis jääb keeleteaduse pädevusse [40].

Samuti vajab täiendavat detailanalüüsi masinõppe algoritmile, ARVA CMS rakenduse töölaual teabevaldajalt tagasiside küsimise sisu ning vorm ja ARVA CORE rakendusse edastatava tagasiside sõnumi täpne formaat. Minimaalselt peaks olema võimalik vastata jah-ei vormis kas teabevaldaja arvates oli tuvastus asjakohane.

Kokkuvõte

Autor jõudis analüüsi tulemusel järeldusele, et avaliku info korrektsus ja ajakohasus on riiklike digitaalvaldkonna strateegia eesmärkide saavutamiseks läbivalt tähtsad ning infoomanike senisest suurem kaasatus ja nende töö lihtsustamine sobivate digitaalsete vahenditega võimaldab seatud eesmärkide saavutamise. Praegune lahendus kaasab infoomanikke ainult Riigiportaalis avaldatava info kontekstis ja ei vasta tuvastatud nõuetele hallatavuse parandamises ning info ajakohasuse tagamises. Ka on olemasoleval lahendusel ebasobivad ilmingud pärandüsteemide hulga vähendamise nõude osas millest autor räägib lähemalt süsteemi AS-IS kirjeldamisel.

Autori hinnangul on RIA olemasolevas struktuuris ja töökorralduses olemas enamik vajalikke ressursse ja võimekusi nõuete täitmiseks ja eesmärkide saavutamiseks. Ainsa puuduva ressursi ja võimekusena näeb autor keeletötluse (NLP e. *Natural Language Processing i.k.*) algoritmide analüüsivõimekust mille tekitamiseks on vajalik keeleteadlas(t)e kaasamine. Olemasoleva aga oluliselt täiendamist vajava võimekusena näeb autor info sünkroniseerimise võimekust ent seda on võimalik saavutada olemasolevate ressursside rakendamise teel.

ARVA uuenduse visioon on läbi süsteemi ajakohastamise ning sellele täiendavate võimekuste lisamise tagada joondumine riiklike strateegiatega.

Töö eesmärk sai täidetud ning selle tulemus on:

- ARVA hetkeseis on kirjeldatud ja sõnastatud on analüüsist tulenevad järeldused.
- Huvitatud osapoolte seos nõuete ning strateegiliste eesmärkidega on kaardistatud.
- Riigiasutuste digikanalite avaliku teabe haldamise protsess ja võimekused vajavad täiendamist, et tagada vastavus rahvusvaheliste nõuete, riiklike strateegiate ja RIA asutusesiseste strateegiatega;
- Olemasolev karbitoodetel põhinev ARVA lahendus ei võimalda realiseerida vajalikke funktsionaalsusi ning teatud toodete asemel on vaja on hankida uusi tarkvaralisi lahendusi, mille eskiisid töötati välja käesoleva töö raames;

- Sõnastatud on ARVA uus visioon ja eesmärk ning loetletud põhilised tegevused selle saavutamiseks;

Magistritöö tulem on vajalik ARVA laiema juurutamise strateegia rakendamiseks ning sisendina detailanalüüsi tellimisele.

Kasutatud kirjandus

- [1] Euroopa Parlament, „Euroopa Parlamendi ja nõukogu määrus (EL) 2018/1724,“ 2018. [Võrgumaterjal]. Loetud aadressil: <https://eur-lex.europa.eu/eli/reg/2018/1724/oj>.
- [2] M. Ader, „Artiklivaramu (tutvustavad slaidid),“ 13 04 2022.
- [3] Riigi Infosüsteemi Amet, „Riigiportaali eesti.ee,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/riigi-infosusteem/personaaliik/riigiportal-eestiee>. [Kasutatud 16 05 2024].
- [4] Majandus- ja kommunikatsiooniministeerium, „Sündmusteenused,“ [Võrgumaterjal]. Loetud aadressil: <https://www.mkm.ee/digiriik-ja-uhenduvus/digiteenused/sundmusteenused>. [Kasutatud 16 05 2024].
- [5] Riigi Infosüsteemi Amet, „Bürokratt,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/riigi-infosusteem/personaaliik/burokratt>. [Kasutatud 16 05 2024].
- [6] Majandus - ja Kommunikatsiooniministeerium, „Avaliku sektori kodulehtede visioon,“ 02 2015. [Võrgumaterjal]. Loetud aadressil: <https://www.mkm.ee/media/7375/download>. [Kasutatud 16 05 2024].
- [7] Riigi Teataja, „Avaliku Teabe Seadus,“ 07 03 2023. [Võrgumaterjal]. Loetud aadressil: <https://www.riigiteataja.ee/akt/107032023011>. [Kasutatud 16 05 2024].
- [8] Postimees Grupp, „Sünnitoetusest ilma jäänud pere: poleks oodanud, et kodukant sellisel rõõmu täis ajal hingepiina valmistab,“ 16 11 2022. [Võrgumaterjal]. Loetud aadressil: <https://elu24.postimees.ee/7647068/sunnitoetusest-ilma-jaanud-pere-poleks-oodanud-et-kodukant-sellisel-roomu-tais-ajal-hingepiina-valmistab>. [Kasutatud 16 05 2024].
- [9] Vabariigi Valitsus, „Eesti 2035,“ 12 05 2021. [Võrgumaterjal]. Loetud aadressil: https://www.valitsus.ee/sites/default/files/documents/2021-06/Eesti%202035_PUHTAND%20%C3%9CLDOSA_210512_1.pdf. [Kasutatud 16 05 2024].
- [10] Vabariigi Valitsus, „Eesti 2035 Vabariigi Valitsuse tegevuskava,“ 25 04 2024. [Võrgumaterjal]. Loetud aadressil: https://www.valitsus.ee/sites/default/files/documents/2024-04/Eesti%202035%20tegevuskava_25.04.2024.pdf. [Kasutatud 16 05 2024].
- [11] Majandus- ja Kommunikatsiooniministeerium, „Arengukava Eesti digiühiskond 2030,“ 13 12 2021. [Võrgumaterjal]. Loetud aadressil: https://www.mkm.ee/sites/default/files/documents/2022-04/Digi%C3%BChiskonnna%20ARENGUKAVA_13.12.2021.pdf. [Kasutatud 16 05 2024].

- [12] Riigi Infosüsteemi Amet, „Ameti ülesanded ja struktuur,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/amet-uudised-ja-kontakt/amet-ja-juhtkond/amenti-ulesanded-ja-struktuur>. [Kasutatud 16 05 2024].
- [13] Riigi Infosüsteemi Amet, „Riigi Infosüsteemi Ameti strateegia 2021-2025,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/sites/default/files/documents/2022-11/RIA-strateegia-2021%E2%80%932025.pdf>. [Kasutatud 16 05 2024].
- [14] Wikipedia, „Riigi Infosüsteemi Amet,“ [Võrgumaterjal]. Loetud aadressil: https://et.wikipedia.org/wiki/Riigi_Infos%C3%BCsteemi_Amet. [Kasutatud 16 05 2024].
- [15] N. Giard, „Wiki.js dokumentatsioon,“ Requirks.io, [Võrgumaterjal]. Loetud aadressil: <https://docs.requirks.io/>. [Kasutatud 16 05 2024].
- [16] Wikipedia, „Wiki.js,“ [Võrgumaterjal]. Loetud aadressil: <https://en.wikipedia.org/wiki/Wiki.js>. [Kasutatud 16 05 2024].
- [17] Riigi Infosüsteemi Amet, „Artiklivaramu X-tee alamsüsteemi kirjeldus,“ [Võrgumaterjal]. Loetud aadressil: <https://www.x-tee.ee/service-catalog/organizations/70006317/subsystems/artiklivaramu>. [Kasutatud 02 04 2024].
- [18] Elastic NV, „What is Elasticsearch?,“ Elastic NV, [Võrgumaterjal]. Loetud aadressil: <https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html>. [Kasutatud 16 05 2024].
- [19] BuiltWith PTY LTD, „builtwith.com,“ BuiltWith PTY LTD, 01 04 2024. [Võrgumaterjal]. Loetud aadressil: <https://trends.builtwith.com/websitelist/Wiki.js/Long-Term>. [Kasutatud 01 04 2024].
- [20] Wiki.js, „Wiki.js developers,“ Wiki.js, 03 2024. [Võrgumaterjal]. Loetud aadressil: <https://js.wiki/about#developers>. [Kasutatud 02 04 2024].
- [21] GitHub, „GitHub Wiki.js repositooriumisse panustajad,“ GitHub, 01 04 2024. [Võrgumaterjal]. Loetud aadressil: <https://github.com/requirks/wiki/graphs/contributors>. [Kasutatud 01 04 2024].
- [22] A. Mockus, R. T. Fielding ja J. D. Herbsleb, „A Case Study of Open Source Software Development: The Apache Server,“ *Proceedings of the 22nd International Conference on Software Engineering*, pp. 263-272, 2000.
- [23] Requirks.io, „Wiki.js evitusmärkmed,“ Requirks.io, 20 02 2022. [Võrgumaterjal]. Loetud aadressil: <https://docs.requirks.io/releases/2p5>. [Kasutatud 16 05 2024].
- [24] Requirks.io, „Wiki.js projekti infoleht,“ Requirks.io, 19 01 2021. [Võrgumaterjal]. Loetud aadressil: <https://docs.requirks.io/releases/about>. [Kasutatud 16 05 2024].
- [25] T. Cornford, „Total cost of ownership of open source software: a report for the UK Cabinet Office supported by OpenForum Europe,“ London School of Economics, London, 2011.
- [26] Wikipedia, „Wiki definitioon,“ [Võrgumaterjal]. Loetud aadressil: <https://en.wikipedia.org/wiki/Wiki>. [Kasutatud 16 05 2024].
- [27] S. Robinson, „What is a content management system?,“ TechTarget, 04 2024. [Võrgumaterjal]. Loetud aadressil:

- <https://www.techtarget.com/searchcontentmanagement/definition/content-management-system-CMS>. [Kasutatud 16 05 2024].
- [28] America's Cyber Defense Agency, „Content Management Systems Security and Associated Risks,“ 19 10 2016. [Võrgumaterjal]. Loetud aadressil: <https://www.cisa.gov/news-events/alerts/2013/01/24/content-management-systems-security-and-associated-risks>. [Kasutatud 16 05 2024].
- [29] K. Vaher ja teised, „Digiriigi tulevikuarhitektuuri suunad,“ 24 05 2021. [Võrgumaterjal]. Loetud aadressil: <https://docs.google.com/document/d/1ka4PM-IzpvquGVX10CTgDuTotGv6g47pe1MfQybyq3w>. [Kasutatud 16 05 2024].
- [30] E. Evans, Domain-Driven Design: Tackling Complexity in the Heart of Software, Addison-Wesley, 2004.
- [31] Riigi Infosüsteemi Amet, „Keskne volituste haldamise infosüsteem Pääsuke,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/riigi-infosusteem/kesksed-platvormid-avalike-e-teenuste-pakkumiseks/paasuke>. [Kasutatud 16 05 2024].
- [32] S. Potter, „The Architecture of Open Source Applications (Volume 2),“ [Võrgumaterjal]. Loetud aadressil: <https://aosabook.org/en/v2/git.html>. [Kasutatud 16 05 2024].
- [33] S. K. Devieni, „Version Control Systems (VCS) the Pillars of Modern Software Development: Analyzing the Past, Present, and Anticipating Future Trends.,“ *International Journal of Science and Research (IJSR)*, kd. 9, pp. 1816-1829, 2020.
- [34] Riigi Infosüsteemi Amet, „Andmevahetuskiht X-Tee,“ [Võrgumaterjal]. Loetud aadressil: <https://www.ria.ee/riigi-infosusteem/andmevahetuse-platvormid/andmevahetuskiht-x-tee>. [Kasutatud 16 05 2024].
- [35] Sherlock, „The Sherlock Plagiarism Detector,“ [Võrgumaterjal]. Loetud aadressil: <https://github.com/diogocabral/sherlock?tab=readme-ov-file>. [Kasutatud 16 05 2024].
- [36] Plagiarism-Checker, „Plagiarism-Checker,“ [Võrgumaterjal]. Loetud aadressil: <https://github.com/architshukla/Plagiarism-Checker>. [Kasutatud 16 05 2024].
- [37] Bloomfield Media, „WCOPYFIND,“ [Võrgumaterjal]. Loetud aadressil: <https://plagiarism.bloomfieldmedia.com/software/wcopyfind/>. [Kasutatud 16 05 2024].
- [38] JPlag, „JPlag - Detecting Software Plagiarism,“ [Võrgumaterjal]. Loetud aadressil: <https://github.com/jplag/jplag>. [Kasutatud 16 05 2024].
- [39] M. Conway, „CONWAY'S LAW,“ [Võrgumaterjal]. Loetud aadressil: https://www.melconway.com/Home/Conways_Law.html. [Kasutatud 16 05 2024].
- [40] M. Fowler, „Conway's Law,“ 20 10 2022. [Võrgumaterjal]. Loetud aadressil: <https://martinfowler.com/bliki/ConwaysLaw.html>. [Kasutatud 16 05 2024].
- [41] A. Lohk ja K. Ross, „Joachim Rossihniuse ja Heinrich Stahli perikoopide võrdlus,“ *Emakeele Seltsi aastaraamat*, kd. 64, 2019.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Allan Bernard

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „GIT põhise avaliku sektori sisuhaldussüsteemi arhitektuuri innovatsioon“, mille juhendaja on Alari Krist
 - 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.

16.05.2024

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 tähtjaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Artiklivaramu olemasoleva lahenduse SWOT analüüs

Tabel 2. Artiklivaramu AS-IS lahenduse SWOT analüüs (autori koostatud).

#	Tugevused (<i>Strengths</i>)	#	Võimalused (<i>Opportunities</i>)
S1	Eksisteeriv ja evitatud platvorm ning põhjalik arusaam artiklivaramu ja avaliku info halduse domeenispetsiifilisest ärioloogikast ning aspektidest.	O1	Innovatsioon ning strateegiline vastavus on võimalik saavutada kasutades eksisteerivaid, laialt levinud ja hästi mõistetud tehnoloogiaid ning tuginedes olemasolevatele kompetentsidele.
S2	Pühendunud ja uuendusmeelne meeskond kellel on kõik vajalikud kompetentsid, et innovatsiooniprojekte ellu viia.	O2	Puuduvad kompetentsid on tavapärased ning neid on võimalik hankida nii sisse ostmise kui värbamise teel.
S3	Üldisel tasandil eksisteeriv strateegiline ja seadusandlik tahe. Avaliku teabe halduse ja ajakohasena hoidmise olulisus on sõnastatud nii strateegiadokumentides [11] kui seaduses [7] ¹ .	O3	Keeruline majandusolukord on suurendanud erasektori ettevõtete huvi avaliku sektori hangete vastu. Tihenenud konkurents suurendab hangete õnnestumise tõenäosust ning pärsib pakkumiste maksumust.
S4	RIA kui organisatsiooni muljetavaldav võimekus ja pikaajaline kogemus infosüsteemide hankimisel, arendamisel, evitamisel ja käitlemisel.	O4	Avalikku teavet masinloetaval kujul kasutavate teenuste arvu suurenedes on võimalik nende initsiatiivide raames või koostöös nendega täiendada artiklivaramu kui platvormi võimekusi.
-	-	O5	Võimalus asutusesiseselt itereerida ja katsetada lahendusi Riigiportaali kui ühe suurima teiste teabevaldajate avaliku teabe avalikustajaga. (vt. ka oht T5)

¹ 2. Jagu, §32 lõige 1, punktid 2 ja 3

#	Nõrkused (<i>Weaknesses</i>)	#	Ohud (<i>Threats</i>)
W1	Evitatud lahendus ei ole sisulises kooskõlas seatud strateegiatega ning ei ole piisavalt efektiivne vahend seadusega määratud ootuste täitmiseks.	T1	Ebasobivate ilmingutega pärandvara juurutatakse sügavamalt ja riigiasutuste üleselt mis muudab innovatsiooni ning seatud eesmärkidega vastavuse saavutamise veelgi keerulisemaks.
W2	Hoolimata avaliku teabe riigiüleselt ajakohasena hoidmise vajalikkuse teadvustamisest on selle saavutamiseks vajalikud tegevused ning lahendamist vajavad probleemid jätkuvalt alatähtsustatud ja alarahastatud.	T2	Usaldusväärse statistilise sisendi puudumine duplikaatinfo ja võimaliku väärinfo esinemise ulatusest, põhjustab olukorra kus strateegilisi otsuseid võetakse vastu tuginedes anekdootlikele argumentidele (nn. „kõhutunne“) või üksikutele tuvastatud juhtumitele.
W3	Süsteemi puuduste tõttu ei ole teabe mittevastavuse tegelikku ulatust võimalik täielikult hinnata.	T3	Info haldamise keerukus põhjustab suurema silostumise kus avaliku teabe valdaja keskendub ainult sellele koopiaale avalikust teabest mis asub otseselt tema vastutusalas olevates infosüsteemides.
W4	Avaliku teabe ajakohasena hoidmise protsessid on kas defineerimata või erinevad asutuste üleselt, mistõttu on raske hinnata protsesside tegelikku maksumust ja mõju maksumaksjale ning protsesside optimeerimisest saadavat rahalist kasu.	T4	Suurenevad riiklikud vajadused või rahvusvahelised regulatsioonid, info ajakohasena hoidmise osas, seavad nõuded mille realiseerimine eksisteeriva pärandvara abil ei ole seatud ootuste, tähtaegade ning eelarvete mahus lihtsalt võimalik.
-	-	T5	Riigiportaali kui suurema ja avalikkusele suunatud projekti arhitektuur ja vajadused mõjutavad liigselt Artiklivaramu arhitektuuri ja arendust, muutes asutuseväliste süsteemidega liidestumise keerulisemaks. (vt. ka võimalus O5)
-	-	T6	Süsteemi ühel aluskomponendil (Wiki.js) puudub nii usaldusväärne jätkuarenduste plaan kui ka kommertstugi.

Lisa 3 – Töö tulemuste kokkuvõte

Tabel 3. Töö tulemuste kokkuvõte (autori koostatud).

nr	tüüp	Kirjeldus
J1	järeldus	Avaliku teabe halduse vastavus Avaliku Teabe Seadusega ei ole tagatud ning protsessid vajavad parandamist.
J2	järeldus	Sisuhalduseks mõeldud vabavariante või kommertslike valmistoodete kasutamine ei ole avaliku teabe, asutuste üleselt, haldamiseks optimaalne.
J3	järeldus	GIT kasutamine riigi artiklite tehnilise salvestus- ja sünkroniseerimismehhanismina on jätkuvalt asjakohane.
J4	järeldus	Wiki.js tarkvara ei oma suurt kogukonna tuge täiendarendusteks ning tarkvara omaniku motivatsioon seda hallata on madal. Wiki.js kasutamine ARVA pikaajalise põhikomponendina ei ole mõistlik.
J5	järeldus	ARVA mandaat nn. “tõe allikana” info osas mille tekkekohaks ning omanikuks on RIA välised asutused ei ole mõttekas ning oleks vastuolus riigi andmekogude ja andmehalduse põhimõtetega.
J6	järeldus	ARVA eraldamine riigiportaalist ning riigiportaali käsitlemine sarnasena teistele liidestajatele oli asjakohane.
E1	ettepanek	Mitte lubada ARVA’ga liidestumisel erisusi Riigiportaalile. Kogu Riigiportaalile võimaldatud funktsionaalsust peaks olema võimalik võimaldada ka välisele osapoolle.
E2	ettepanek	ARVA infosüsteemi roll peaks olema artiklite infot sünkroniseeriv ning agregeeriv. ARVA keskne komponent ei peaks sisaldama sisuhalduse aspekti.
E3	ettepanek	Tellida detailanalüüs ARVA keskse komponendi hankimiseks.
E4	ettepanek	Tellida detailanalüüs masinõppel põhineva duplikaadituvastustööriista hankimiseks ning integreerida see ARVA keskse komponendiga.
E5	ettepanek	Mitte juurutada Wiki.js’i RIA välistes asutustes ARVA integratsioonikomponendina.
E6	ettepanek	Tellida detailanalüüs ARVA sisuhaldus- ja integratsioonikomponendi hankimiseks ning avalikustada see avatud lähtekoodiga tarkvarana, riiklikuks kasutamiseks RIA välistes asutustes.
E7	ettepanek	Juurutada punkti (<eelmine>) tulemina hangitud integratsioonikomponent riigiportaalis.

T1	tulemus	Kirjeldatud on vajamineva tarkvara komponente, nende funktsionaalsust ja omavahelisi seoseid, tasemel mis on piisav detailanalüüsi tellimiseks.
T2	tulemus	Kirjeldatud on duplikaatinfo tuvastamise protsessi ja selleks vajaminevaid komponente, tasemel mis on piisav detailanalüüsi tellimiseks.