

Euroopa Liit
Euroopa Sotsiaalfond



Eesti tuleviku heaks

Uurimisprojekti „Andmeaitade (teiseste andmekogude) loomise põhimõtete väljatöötamine“ lühikokkuvõte

Uuringu tellija: Riigikantselei, tarkade otsuste fond

Uuringu partnerid: Riigi Infosüsteemi Amet ja Sotsiaalministeerium

TTÜ Küberneetika Instituut

Tallinna Tehnikaülikool

Data Capital OÜ

Vahur Kotkas
Hele-Mai Haav
Jaak Tepandi
Enn Õunapuu
Jaanus Grauberg

TTÜ Küberneetika Instituut
Tallinn 2013



Andmeaitade uuring

Soovitav viide: V. Kotkas, H-M. Haav, J. Tepandi, E. Õunapuu, J. Grauberg, Uurimisprojekti "Andmeaitade (teiseste andmekogude) loomise põhimõtete väljatöötamine" lõpparuanne, TTÜ Küberneetika Instituut, 2013

Tellija: Riigikantselei ja Riigikantselei tarkade otsuste fond

Partnerid: Riigi Infosüsteemide Amet ja Sotsiaalministeerium

ISBN: 978-9949-430-65-9 (võrguväljaanne, pdf)

TTÜ Küberneetika Instituut

Akadeemia tee 21

12618 Tallinn

Tel. 6204150

Fax 6204151

E-post dir@ioc.ee

www.ioc.ee

Käesoleva töö valmimisse on andnud olulise panuse ka Tellijapoolne töörühm, kuhu kuulusid

Ahto Kalja, Uuno Vallner, Anti Urm, Heiko Vainsalu, Rein Murakas, Siim Sikkut ja Kaisa-Maarja Jagula.

Uuringut rahastasid Euroopa Liidu

Sotsiaalfond ja Riigikantselei läbi Riigikantselei tarkade

otsuste fondi.

SISUKORD

1	SISSEJUHATUS	4
2	PROBLEEMI KIRJELDUS JA UURINGU EESMÄRK	5
3	UURINGU METOODIKA	8
4	UURINGU TULEMUSED	11
4.1	ANKEETIDE JA INTERVJUUDE ANALÜÜSI TULEMUSED	11
4.2	RAHVUSVAHELISE PRAKTIKA JA TEHNOLOOGIATRENDIDE ANALÜÜSI TULEMUSED	21
5	KOKKUVÕTE	28
6	POLIITIKASOOVITUSED	34
7	LÜHENDID JA MATERJALIDE LOETELU	35
7.1	LÜHENDID, MÕISTED JA SÕNASELETUSED.....	35
7.2	KASUTATUD MATERJALID	35

1 SISSEJUHATUS

Andmeait on kindlale valdkonnale (või probleemile) orienteeritud, teisene, integreeritud, ajast sõltuv, püsiv või loogiliselt integreeritud andmekogum, mille eesmärgiks on toetada otsuste tegemist. Andmeaidad koostatakse juba olemasolevatest andmekogudest spetsiifiliste andmetöötlusülesannete (näiteks andmeanalüüsi) ja aruannete koostamiseks. Tänapäeval kasutatakse üha enam ka laiemat määratlust, mille kohaselt andmeait on andmete kasutamise meetodite, tehnoloogiate ja praktikate kompleks, mille eesmärk on teha paremaid otsustusi ning pakkuda paremaid teenuseid ja mida võib realiseerida väga mitmesuguste vahenditega. Eesti riiklikus sektoris loodavad andmeaidad aitaksid toetada poliitikakujundajate ja elluvijate otsuste kvaliteeti, kuna otsuste aluseks saaks võtta erinevatest andmekogudest agregeeritud informatsiooni.

Uuring analüüsib andmeaitade poliitilisi, sotsiaalseid, organisatoorseid, infotehnoloogilisi, metodoloogilisi ja juriidilisi aspekte ning pakub välja kasutatava tehnoloogia ja reeglistiku ühtlustatud andmeaitade loomiseks riigis. Uuringu põhitulemuseks on metoodilised soovitused ja juhised andmeaitade (ka andmeanalüüsi) abil lahendatavate riiklike probleemide (sh poliitiliste, aruandluse, teaduslike jms) ja tehtavate otsuste osas, nende probleemide lahendamiseks andmeaitades ladustatavate andmete kogumise, haldamise, privaatsuse, koosvõime ja analüüsi läbiviimiseks ning kolmandatele osapooltele kättesaadavaks tegemiseks.

Uuringu läbiviimisel on lähtutud „Eesti infoühiskonna arengukava 2013“ suunistest, nende hulgas eriti p. 4.3.1 „Avaliku sektori toimimise tõhustamine“ toodud ülesanded tegevussuundades „Avaliku sektori haldustoimingute ja menetlusloogikate ümberkujundamine vastavalt IKT rakendamisest tulenevate eeliste ja võimaluste ärakasutamiseks“ ja „Poliitikakujundamise tõhustamine parema andmekasutuse ning infoühiskonna mõju ja väljakutseid käsitlevate uuringute läbiviimise kaudu“.

2 PROBLEEMI KIRJELDUS JA UURINGU EESMÄRK

Praeguseks on mitmed ministeeriumid ja nende haldusalade asutused (Sotsiaalministeerium, Siseministeerium jt.) oma tööülesannetest lähtuvalt asunud andmeaitasid välja töötama, kuid ei teosta seda koordineeritult, kuna riigis puudub üldtunnustatud andmeaitade loomise ja kasutamise poliitika. Ka valdkonnasiseselt lahendab iga töörühm oma probleeme iseseisvalt. Samas näiteks on Sotsiaalministeeriumi haldusalasse planeeritavate andmeaitade eeldatav kasutatavus seotud äärmiselt laia otsuste tegemise tasandiga ja ulatub väljapoole sotsiaalvaldkonna piire (teadusuuringud, statistika, finantsvaldkonna planeerimine jne.).

Käesoleva uuringu põhieesmärgiks on lahenduste välja pakkumine andmeaitade haldamise protsessi korrastamiseks ja harmoneerimiseks riigi infosüsteemis (IS) tervikuna, lahendades sellega ka erinevate ametkondade andmete ladustamise ja analüüsi probleemid. Uuringul on nii poliitiline, sotsiaalne, organisatoorne, infotehnoloogiline, metodoloogiline kui ka juriidiline aspekt.

Lähtudes nimetatud põhieesmärgist on käesoleva uuringu ülesanded järgmised:

1. Riigi infosüsteemis rakendatavate andmeaitade tehnoloogia ja kogutavate andmete analüüs, selgitamaks välja riigis kujunenud andmete ladustamise ja analüüsi hetkeolukord.
2. Uurida välja, millised nii infotehnoloogilised, organisatoorsed, poliitilised, seadusandlikud kui sotsiaalsed probleemid on tekkinud andmeaitade rakendamisel riigi erinevates haldusalades.
3. Uurida maailmas levinud uusimaid suundi andmete ladustamise ja suurte andmekogumite analüüsi tehnoloogias, teha kindlaks kuidas andmeaitade tehnoloogiat on kasutatud teiste riikide IS-des ning vaadelda ka Eesti erasektori praktikaid nimetud valdkondades.
4. Uurida andmeaitade võimalikku (semantilist) koosvõimet teiste andmekogudega (ka andmeaitadega), kuna erinevates haldusalades kasutatavaid andmeaitu ei saa vaadelda isoleeritult teistest andmekogudest.

Andmeaitade uuring

Ülalpool nimetud uuringulist iseloomu ülesannete tulemite alusel pakutakse uuringus välja lahendusi ja meetodilisi soovitusi Eesti riigi infosüsteemi andmeaitade arendamise, haldamise ja andmeanalüüsi parendamiseks ning harmoniseerimiseks selleks, et riigile ja selle kodanikele tähtsate otsuste langetajad saaksid võtta vastu teadmispõhiseid otsuseid. Seega on projekti lõppeesmärgiks tagada otsuste vastuvõtmise kvaliteedi parendamine.

Lisaks vaadeldakse käesolevas uuringus ka kõigi ülalpool mainitud probleemiga kaasnevaid juriidilisi aspekte, eriti seost pakutud lahenduste sobivusega meie riigi seadusandlusega; vajadusel pakutakse vastavaid seadusandluse muudatusi. Juriidilised aspektid omavad erilist tähtsust isikuandmete privaatsuse säilitamisel ja seega pööratakse erilist tähelepanu andmete anonüümimise või pseudonüümimise probleemidele andmete ladustamisel ja hilisemal andmete analüüsil.

Uuringu raames otsiti muuhulgas vastuseid järgmistele uuringuküsimustele:

1. Kui palju on tänase seisuga avalikus sektoris andmeaitasid olemas, kui palju on loomisel ning planeerimisel?
2. Milliseid andmeid, millistest valdkondadest ja milliste otsustustasandite jaoks üldse andmeaitadesse kogutakse ja mida perspektiivis (nii olemasolevate konkreetsete plaanide järgi kui ka kaugemas tulevikus) koguda tahetakse?
3. Millised andmed on vaja andmeaita sisestamiseks kodeerida (eelkõige isikuandmete kaitsest lähtuvalt) ja millised mitte?
4. Kuidas on korraldatud eri ministeeriumide ja ametkondade koostöö andmeaitadesse kogutavate andmete kogumisel, laadimisel ja kasutamisel? Kuidas on korraldatud asutuste vahel andmeaitade andmekoosseisu muudatuste teostamise protsess?
5. Millised on andmeaitade loomise ja haldusega seotud seadusandlikud probleemid? Milliseid regulatsioone oleks vaja muuta, täiendada või lisada?
6. Kuidas kujunevad praktikas välja vajadused andmete hoidmiseks ja töötlemiseks andmeaitades?
7. Kuivõrd on andmeaitade andmete koosseis ajutine ja peale töötlushi kustutatav ning kuivõrd on vaja aasta aastalt üha uusi andmeid andmeaita lisada, et tekiks võimalused andmete kumuleerimiseks ja aegridade analüüsiks?
8. Kellele ja mis alusel andmeaitadest andmeid väljastatakse?

Andmeaitade uuring

9. Kuidas jaguneb vastutus andmeaidas olevate andmete kvaliteedi osas? Kuidas andmeaidad omavahel edaspidi suhtlevad? Kas selleks on seadusandlusele vaja lisada uusi regulatsioone?

10. Millised on lahendused ja regulatsioonid avaliku sektori andmeaitade osas teistes riikides? Kuhu suurte andmehulkade töötlus tänapäeva parimaid praktikaid silmas pidades suundub?

11. Mida on andmeaitade loomisel ja halduses õppida Eesti erasektori praktikatest?

12. Mida peaks eelnevast lähtudes Eestis andmeaitade tehnoloogia, seadusandluse ja kasutamise alal edasi tegema?

Uuringüküsimustele vastamiseks kasutati andmeaitade valdkonna tehnoloogilist arengut käsitlevaid teabematerjale (näiteks analoogiliste rahvusvaheliste uuringute tulemused, tehnoloogiatrendide uuringute tulemused jms), teiste riikide ja Eesti erasektori praktikat andmeaitade rakendamisel otsustusprotsesside toetuseks tehtavaks andmeanalüüsiks ning andmeaitade kasutamise, loomise, haldamise ja õigusliku reguleerimisega seotud huvigruppide hinnanguid andmeaitadega seotud erinevatele aspektidele (intervjuude ja ankeetide põhjal).

3 UURINGU METOODIKA

Uuringu läbiviimise metoodika tuleneb uuringu eesmärgist, ülesannetest ja saadavatest lähteandmetest.

Uuringu põhihüpoteesiks oli, et riigi infosüsteemi andmeaitade valdkonnas on vajalikud poliitilised, seadusandlikud, organisatoorsed ja tehnoloogilised muutused.

Vastavalt sellele oli uuringu metoodika kavandatud järgmiselt:

- **Olemasoleva teabe kogumine ja analüüs**, mis põhineb analoogiliste rahvusvaheliste uuringute tulemustel, andmeaitade ja andmeanalüüsi alastel tehnilistel artiklidel ja raamatutel, vastaval seadusandlusel, tooteuuringutel jms.
- **Teabe ja hinnangute kogumine intervjuude ja ankeetküsitluste abil** selgitamaks välja andmeaitade kasutamise, loomise, haldamise ja õigusliku reguleerimisega seotud huvigruppide arvamusi, kogemusi, vajadusi, ettepanekuid ja tulevikuvisionoone.
- **Vaheseminar** arutamaks diskussiooni vormis vastavate huvigruppidega uurimuse vahetulemusi.
- **Töökoosolekud** Tellija-poolsete partnerite esindajatega ja uuringu juhtkomitee liikmetega uurimise käigu täpsustamiseks ning vahe- ja lõpptulemuste arutamiseks.
- **Uuringu tulemuste analüüs ja esitamine.** Vahe- ja lõpparuande koostamine uuringu tulemuste esitamiseks.
- **Uurimuse lõpptulemusi tutvustav seminar** uuringu tulemuste tutvustamiseks kõigile huvigruppidele.

Uuringu metoodika **teabe ja hinnangute kogumiseks intervjuude ja ankeetküsitluste abil** nägi ette, et kõigepealt viiakse läbi intervjuud ettenähtud sihtgruppides selleks, et kaardistada andmeaitade probleemistikku, katsetada küsimustikku ja häälestada valikvastuste skaalasad ning vajadusel püstitada täiendavaid hüpoteese. Anketeerimine oli ette nähtud analüüsi läbiviimiseks ja hüpoteeside kontrollimiseks ning vastavate järelduste tegemiseks. Nii intervjuude kui ankeetide küsimustiku struktureerimisel kasutati eBCM (The E-Business Community Model, e-äri kogukonna mudel) metoodikat, mille järgi määrati küsimustike temablokid järgnevalt:

Andmeaitade uuring

1. **Põhivalmisolek** iseloomustab seda, kui võrd ühiskond on valmis andmeaitade rakendamiseks. Andmeaitade rakendamine eeldab kokkuleppeid - seadusi ja regulatsioone, standardeid ning lepinguid, mis toetavad andmeaitade juurutamist ja kasutamist. Samuti on hädavajalik kindlustunne suuremate riskide suhtes, mille annab infoturbe tagamine, usaldusväärsete andmete kindlustamine ning kasutajate identifitseerimine. Lõpuks on oluline vastav mõtteviis - andmeaitade kohta käivad teadmised, inimeste motivatsioon ning asutuste vaheline koostöö.
2. **Võimaldajad.** Ainult põhimõttelisest valmisolekust ei piisa, on vaja ka võimaldajaid - taristut, eriti IT osas, aga samuti kvalifitseeritud inimressurssi ning andmeaitade realiseerimise kogemusi.
3. **Liikumapanevad jõud**, mille panevad tööle valmisolek ja võimaldajad. Eelkõige loovad andmeaitad uusi võimalusi info kättesaamiseks ning motiveerivad maandama kaasnevaid ohte. Kogu andmeaitade suuna on muutnud võimalikuks tehnoloogia ja innovatsioon. Turud ja firmadevaheline konkurents võimendab tehnoloogilist arengut. Ühiskond ja poliitilised jõud tunnetavad vajadust agregeeritud informatsiooni põhjal tehtavate otsuste järele, mis motiveerib andmeaitade arengut.

Uurimismetoodika nägi ette intervjuude läbiviimise järgmistes sihtrühmades:

- IT-ala koordineerijad riigis,
- IT-alase seadusandluse spetsialistid,
- ministriumide ja ametkondade andmeaitade spetsialistid,
- andmeaitade spetsialistid erasektoris,
- andmete kasutajad (ametnikud, poliitikud, teadlased, statistikud jt.).

Intervjuude küsimused olid jagatud järgmistesse gruppidesse: üldinfo, andmeaitadega seonduv infopoliitika ja seadusandlus, andmeaitade organisatoorsed küsimused, andmeaitade infotehnoloogiline aspekt ja intervjuueeritava taustainfo.

Intervjuusid oli kavas läbi viia vähemalt üks iga sihtgrupi kohta va riigisektori andmeaitade spetsialistid, kus planeeriti viia läbi 3 intervjuud. Intervjuud olid ette nähtud viia läbi poolstruktureeritud intervjuudena, mis salvestatakse hilisema järelkuulamise võimaldamiseks. Intervjuude analüüsi tulemusi oli plaanis tutvustada huvigruppidele projekti vaheseminaril, et saada tagasidet ja sisendit ankeetküsitluse läbi viimiseks.

Andmeaitade uuring

Peale selle oli kavas teha 3 lühiintervjuud e-posti vahendusel 3 välisriigi andmeaitade spetsialistiga, et koguda rahvusvahelist kogemust ja head tava.

Anketeerimise üldine metoodika jälgis Massachusetts Institute of Technology (MIT) anketeerimise põhimõtteid (<http://web.mit.edu/surveys/survey-guidelines.pdf>).

Ankeetide küsimustiku koostamise aluseks olid intervjuude küsimustikud ja intervjuude läbiviimisel saadud tagasiside. Näiteks osa intervjuude küsimusi ei töötanud, osade küsimuste skaalad olid kas ebatäpsed või vajasis laiendamist jms. Samuti võeti arvesse juhtgrupi töökoosolekul ja vaheseminaril kõlanud arvamused ja ettepanekud. Lisaks arvestati, et ankeetide küsimustik ei oleks ülepaisutatud. Et kasutati veebipõhist anketeerimist, siis arvestati ka anketeerimise vahendi Google Drive Form võimalusi ja piiranguid.

Põhiliste välitööde käigus viidi läbi 8 intervjuud järgmistes sihtrühmades:

- IT-ala koordineerijad riigis (RISO-1 intervjuu),
- IT-alase seadusandluse spetsialistid (AKI-1 intervjuu),
- ministriumide ja ametkondade andmeaitade spetsialistid (Sotsiaalministeerium, Statistikaamet, Eesti Haigekassa, Politsei ja Piirivalveamet-kokku 4 intervjuud)
- andmeaitade spetsialistid erasektoris (1 intervjuu),
- andmete kasutajad (erasektor- 1 intervjuu).

Intervjuud salvestati ja nende kirjalikud kokkuvõtted esitati koos uuringu vahearuandega.

Peale selle viidi läbi 3 lühiintervjuud vastavalt Ameerika Ühendriikide, Suurbritannia ja Hollandi andmeaitade ekspertidega. Need intervjuud tehti e-posti vahendusel.

Anketeerimine viidi läbi elektroonselt veebiküsimustiku abil kusjuures anketeeriti ka juba intervjuueeritud inimesi. Anketeerimise sihtrühmad olid järgmised: andmeaitade riigisektori spetsialistid, andmeaitade erasektori spetsialistid ja andmete kasutajad.

Sihtrühmade anketeerimiseks koostati eraldi ankeetid spetsialistide ja kasutajate anketeerimiseks. Saadeti laiali 105 ankeeti, neist 11 korduvalt. Kokku laekus 37 vastust, sellest 27 vastust spetsialisti ankeedile (sellest 17 riigi- ja 10 erasektorist) ja 10 vastust kasutaja ankeedile.

4 UURINGU TULEMUSED

Käesoleva uuringu tulemused baseeruvad põhiliselt ankeetide ja intervjuude tulemuste ning teabematerjali analüüsil. Teabematerjalina kasutati nii seadusandlike akte kui tehnoloogiatrende käsitlevaid ning analoogiliste rahvusvaheliste uurimuste materjale.

4.1 ANKEETIDE JA INTERVJUUDE ANALÜÜSI TULEMUSED

Allpool anname lühiülevaate ankeetide ja intervjuude analüüsi põhitulemustest, põhjalikum analüüs on esitatud uuringu lõpparuandes.

4.1.1 Andmeaitade seadusandlus ja standardid

Andmeaitade rakendamine eeldab seadusi ja regulatsioone, standardeid ning lepinguid, mis soodustavad andmeaitade juurutamist ja kasutamist. Näiteks pakuvad andmekaitseadused (isikuandmete kaitse seadus, avaliku teabe seadus, riikliku statistika seadus jm) küll teatud regulatsiooni andmeaitadega tööks, kuid valdkondlikud seadused võivad käia nende andmete hoidmise ja töötusega halvasti kokku või olla kohati vastuolulised.

Selles ja järgnevates jaotistes vastatakse küsimusele, milliseid regulatsioone oleks andmeaitade loomise, ülalpidamise ja omavahelise suhtlemisega seoses vaja muuta, täiendada või lisada ning mida peaks eelnevast lähtudes Eestis andmeaitade seadusandluse alal edasi tegema.

4.1.1.1 Regulatsioonide täiendamise vajadused.

Intervjuude ja ankeetide analüüs näitas, et küsimustele uute regulatsioonide vajalikkuse kohta vastas jaatavalt üldjuhul vähem ning eitavalt - rohkem vastajaid. Muutmise peamised valdkonnad on seotud andmeaitade määratlemisega, andmete privaatsusega andmeaitades ning andmeaitade loomise ja haldamisega. Riigi esindajad leiavad intervjuudes, et seadusandlike regulatsioonide muudatused on vajalikud, erasektori esindajad seevastu ei näe vajadust muudatusteks.

Intervjuude ja ankeetküsitluste hinnangud regulatsioonide muutmise vajadusele ei ole kaugeltki ühesed.

Andmeaitade uuring

Muudatused võivad esmajoones olla vajalikud järgmistes seadusandlikes aktides.

- Avaliku teabe seadus (sh andmekogu määratlus § 43¹ ning keeld kasutada ühtede ja samade andmete kogumiseks eraldi andmekogusid § 43³).
- Isikuandmete kaitse seadus (sh privaatsuse nõuded).
- Vabariigi Valitsuse määrus "Riigi infosüsteemi haldussüsteem" (sh andmeaitade registreerimine RIHAs).
- Riikliku statistika seadus (sh § 34 lõige 3).
- Muud seadusandlikud aktid vastavalt vajadusele.

Analüüsid regulatsioonide täiendamise vajadusi, tuleks arvestada, et ei saa luua kitsalt andmeaitadest lähtuvaid regulatsioone, vaid lähtuda tuleks üldistest eesmärkidest ja põhimõtetest, sealhulgas nii soovist arendada riigi pakutavaid teenuseid ja teha paremaid otsustusi kui ka vajadusest kaitsta inimeste privaaatsust.

Lisanduvad regulatsioonid võivad tuua selgust andmeaitade kasutamise osas. Samas on võimalik, et regulatsioonide kavandamise ja kehtestamise käigus tekitatakse hoopis täiendavaid kitsendusi ja bürokraatiat.

Tehnoloogilise arengu tingimustes on andmeaitade määratlus pidevas muutumises. Kui sätestada andmeaidad seadusandluses kui teatud eri klass andmekogusid, millel on rida spetsiifilisi omadusi ("*taking the data to the analysis*"), siis selline definitsioon võib hoopis takistada tehnoloogilist arengut koosvõime, andmete virtualiseerimise, pilvetehnoloogia, mobiilsete tehnoloogiate ja suurte toorete andmete baaside suunas, mida vastavalt vajadusele kasutatakse ja järk-järgult kustutatakse ("*taking the analysis to the data*").

Tuleks arvestada ka võimalikke eelseisvaid muudatusi seadusandluses, eelkõige käimasolevat Euroopa Komisjoni isikuandmete kaitse reformi (isikuandmete kaitse üldmäärust ja direktiivi) ning selle arutelu, sealhulgas Eesti seisukohti isikuandmete kaitset puudutavate Euroopa Komisjoni algatuste suhtes.

Kui Euroopa Komisjoni isikuandmete kaitse reform annab tulemusi, võib see oluliselt muuta andmeaitadega seotud regulatsioone. Kuna vastava arutelu tulemused ei ole teada, võiks sellega seotud regulatsioonide tekitamisega olla pigem ettevaatlik.

Andmeaitade uuring

Arvestades küsitletute erinevaid arvamusi, käimasolevat kohati vastuolulist diskussiooni ning erinevaid poolt- ja vastuargumente, esitame ülaltoodud andmeaitade regulatsioonide täiendamise ettepaneku SWOT (tugevused, nõrkused, võimalused, ohud) analüüsi abil.

Andmeaitade regulatsioonide täiendamise eesmärgid on järgmised.

- Arendada riigi pakutavaid teenuseid.
- Teha paremaid otsustusi.
- Jätkuvalt kaitsta inimeste privaatsust.
- Luua selgus selles, millisel määral on võimalik ja otstarbekas luua ühtede ja samade andmete kogumiseks eraldi andmekogusid.
- Luua selgus isikuandmete kasutamise osas andmeaitades.

Järgnevas tabelis on toodud andmeaitade regulatsioonide täiendamise võimalikud tugevused, nõrkused, võimalused ja ohud seoses püstitatud eesmärkidega.

Tabel 1 Regulatsioonide täiendamise SWOT analüüs

	Kasulikud eesmärkide suhtes	Kahjulikud eesmärkide suhtes
Sisemised	<u>Tugevused</u> Andmeaitade lihtsam loomine. Selgus isikuandmete kasutamise osas. Parem ülevaade olemasolevatest andmeaitadest.	<u>Nõrkused</u> Regulatsioon osutub lõpptulemusena kitsendavaks. Tekivad lisatöö, -kulutused ja -bürokratia andmeaitade loomisel ja haldamisel.
Välised	<u>Võimalused</u> Paremad teenused ja otsustused. OIOO metoodika kasutamine kulutuste kompenseerimiseks.	<u>Ohud</u> Oht privaatsusele, isikuandmete leke suurtest andmeaitadest. Euroopa Komisjoni isikuandmete kaitse reform tühistab regulatsioonide muudatused. Uued regulatsioonid osutuvad kitsendavaks uute tehnoloogiate suhtes.

4.1.1.2 Andmete dubleerimine andmeaitades ja avaliku teabe seadus

Vaatame eraldi andmete dubleerimise probleemi andmeaitades. Vastavalt avaliku teabe seaduse § 43³ lõikele 2 on keelatud asutada ühtede ja samade andmete kogumiseks eraldi

Andmeaitade uuring

andmekogusid. See säte võib põhjustada seda, et loodav andmeait on mõnikord ajutise iseloomuga ja kuulub likvideerimisele pärast aruannete ja andmetöötlusülesannete valmimist.

Käesoleva uuringu hinnangul ei tohiks avaliku teabe seaduse § 43³ lõikes 2 sätestatu takistada andmeaitade loomist ja pikaajalist haldamist. Muuhulgas näitab käesolev analüüs, et avaliku teabe seadus (sh § 43¹ lõige 2) hõlmab ka andmeaidad kui ühe andmekogude liigi. Kui dubleerimisega seotud regulatsioone kaaluda, tuleks seda teha andmekogude jaoks laiemalt, mitte vaid andmeaitade jaoks.

4.1.1.3 Andmeturbe probleemid andmeaitades ning isikuandmete kaitse seadus

Andmeturbe probleeme käsitletakse eelkõige järgmistes seadustes ja määrustes.

- Avaliku teabe seadus.
- Isikuandmete kaitse seadus.
- Vabariigi Valitsuse määrus "Infosüsteemide turvameetmete süsteem".

Eelmises jaotises toodud analüüs näitas, et avaliku teabe seadus hõlmab ka teiseseid andmekogusid, sealhulgas andmeaitasid. Küsimus pole siis niivõrd selles, kas oleks vaja eraldi regulatsioone andmeaitade jaoks, kuivõrd selles, kas on vaja eraldi regulatsioone teistest andmekogude jaoks. Vastavalt avaliku teabe seaduse § 43⁶ lõikele 2 ning kehtivale praktikale peavad väga paljud andmekogud kasutama teistes andmekogudes kogutavaid põhiandmeid. Konkreetse andmekogu puhul võib kogutavate põhiandmete osakaal töödeldavatest andmetest olla väike või puududa. Teiseses andmekogus põhiandmeid ei koguta - kõik töödeldavad andmed saadakse välistest andmekogudest.

Kas põhiandmete mittekogumine tekitab vajaduse eraldi regulatsiooni järele? Antud uuringu eesmärkide kohaselt peaks uued regulatsioonid lihtsustama andmeaitade asutamist ja pidamist, mitte tegema neid keerukamaks. Selles kontekstis tuleb küsida, kas põhiandmete mittekogumine võimaldab mingeid seadusandlusest tulenevaid nõudeid lihtsustada või vähendada.

Teine oluline lihtsustus andmeaitade pidamisel oleks võimalik andmeaitade puhul, mis sisaldavad vaid agregeeritud andmeid ning ei võimalda tuvastada konkreetset isikut. Isikuandmete kaitse seadus § 4 ütleb: "Isikuandmed on mis tahes andmed tuvastatud või tuvastatava füüsilise isiku kohta, sõltumata sellest, millisel kujul või millises vormis need andmed on". Seega kui andmeaidad andmed on agregeeritud ja ei võimalda isikut tuvastada,

Andmeaitade uuring

pole enamikku isikuandmete kaitse seaduse sätteid vaja rakendada ning eraldi andmeaitade kohta käivat regulatsiooni ei ole vaja.

Käesoleva uuringu hinnangul ei tohiks põhiandmete mittekogumine andmeaitades tekitada vajadust uue regulatsiooni järele. Kui sellist regulatsiooni kaaluda, tuleb seda teha andmekogude jaoks laiemalt, mitte vaid andmeaitade jaoks. Tuleb vältida muudatusi regulatsioonides, mis tekitavad uusi kitsendusi.

4.1.1.4 Andmeaitade asutamine, registreerimine ja pidamine

Ankeetides jagunesid vastused andmeaitade registreerimise vajalikkuse küsimusele RIHAs ligikaudu võrdselt (8 registreerimise poolt, 10 vastu, 9 ei osanud öelda või ei vastanud). Seejuures 44% riigisektori spetsialistidest leidis, et andmeaitade registreerimine RIHAs on vajalik.

Samas on praegu raskendatud avaliku teabe seaduse § 43⁶ lõike 2 täitmine, sest RIHAs ei ole hetkel lihtsalt välja loetav, millised andmekogud milliseid põhiandmeid koguvad ja kuidas tuleks vastava andmekogu poole pöörduda. Sellel on mitmeid põhjuseid: semantilise koosvõime raamistiku ebapiisav rakendamine (näiteks, puuduvad valdkondade andmete kirjeldused ja andmesõnastikud), RIHAs olevate andmete ebatäpsus (näiteks, paljude andmekogude puhul puudub andmete koosseis) ning RIHA päringute ebapiisavus (pole võimalik teha päringuid põhiandmete kohta).

Käesoleva uuringu hinnangul ei tohiks põhiandmete mittekogumine tekitada vajadust uute üldiste andmeaitade asutamise, registreerimise ja pidamise regulatsioonide järele. Kui selliseid regulatsioone kaaluda, tuleks seda teha andmekogude jaoks laiemalt, mitte vaid andmeaitade jaoks. Samas tuleb olemasolevad ja loodavad andmeaitad viia kooskõlla õigusaktidega.

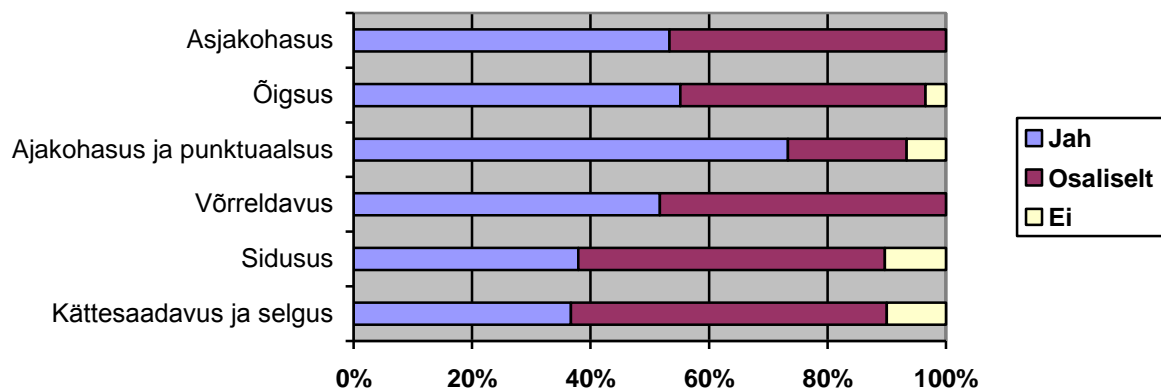
4.1.1.5 Andmete kvaliteet ja andmeaitade omavaheline suhtlus

Ankeetide küsimusele "Kas on vaja uusi regulatsioone andmeid ja nende kvaliteeti puudutava vastutuse sätestamise osas andmeaitade tasandil?" vastas 7 küsitletut jaatavalt, 11 - eitavalt, 15 ei osanud vastata.

Küsimusele "Kas väliste andmekogude kvaliteet on üldiselt rahuldav?" vastas 16 ankeeteeritut jaatavalt ("Jah" või "Pigem jah"), eitavaid vastuseid oli 11.

Andmeaitade uuring

Joonis 1 esitab rahulolu andmeaita laetavate andmete kvaliteediga, mida käsitles täiendav küsimus, mis põhines EUROSTAT andmekvaliteedi parameetritel.



Joonis 1 Rahulolu andmeaita laetavate andmete kvaliteediga

Intervjuude ja ankeetide järgi otsustades hinnatakse rahulolu andmete kvaliteediga andmeaitades paremaks, kui seda enne uuringut oleks võinud oletada. Suhteliselt väiksem oli seejuures rahulolu andmete sidususe, kättesaadavuse ja selgusega. Ankeeteeritute arvates peaksid kvaliteedi eest vastutama vastavate andmekogude vastutavad ja volitatud töötajad.

Arvestades ankeeteeritute hinnanguid, dubleerimise kohta eelpool tehtud ettepanekuid, andmete töötlemise toimingute laia valikut (millest andmekogude vaheline suhtlemine on vaid üks komponent) ning seda, et põhiandmete mittekogumine andmeaitades ei tohiks tekitada vajadust uute regulatsioonide järele, ei soovita käesolev uuring luua uusi regulatsioone spetsiifiliselt andmeaitade omavahelise suhtlemise jaoks. Kui selliseid regulatsioone kaaluda, tuleks neid teha andmekogude ja andmetega tehtavate toimingute jaoks laiemalt, mitte vaid andmeaitade ja andmeaitade omavahelise suhtlemise jaoks.

4.1.1.6 Statistilise üksuse tuvastamine ja riikliku statistika seadus

Riikliku statistika seaduse §34 lõige 3 ütleb: "Statistiline üksus käesoleva seaduse tähenduses on andmete alusel kaudselt tuvastatav, kui otsest tuvastamist võimaldavate tunnuste puudumisel on võimalik statistilist üksust tuvastada muude andmete alusel. Et otsustada, kas statistiline üksus on tuvastatav, võetakse arvesse kõik võimalused, mida kolmas isik võib eeldatavasti kasutada nimetatud statistilise üksuse tuvastamiseks".

Toodud viide kõikidele võimalustele, mida kolmas isik võib eeldatavasti kasutada, võib töötleja või riikliku statistika tegija ette seada raskesti lahendatava ülesande. Võimalusi võib

Andmeaitade uuring

olla väga palju, eriti seoses andmekaevandamise ja sotsiaalse meedia andmete analüüsiga. Otstarbekas on võimaluste ringi kitsendada.

4.1.1.7 Standardid

Enamus ankeetidele vastanud spetsialistidest leidis, et Eesti peaks kasutama rahvusvahelisi standardeid avaliku sektori andmeaitade osas. Põhjendusena toodi seda, et rahvusvaheliste kokkulepete järgimine tuleb üldiselt kasuks, et saab üle võtta häid praktikaid, et saab vältida liigset tööd ("jalgratta leiutamist") ning et standardite kasutamine loob aluse andmete ristkasutuseks.

Tervishoiu valdkond on üks olulisemaid andmeaitade rakendusalasid. Selle valdkonna standardid võivad olla kasulikud ka teiste valdkondade andmeaitade arendamiseks. Arvestades samuti ankeetide vastajate toetust rahvusvaheliste standardite kasutamisele avaliku sektori andmeaitade osas, võiks kaaluda mõnede rahvusvahelise standardite kasutamist andmeaitade loomisel, vajadusel ka tõlkimist eesti keelde ning ülevõttu eesti standardiks.

4.1.2 Andmeaitade kasutamine ja tehnoloogia

4.1.2.1 Andmeaitade kasutusvaldkonnad

Käesolevas uurimuses küsiti anketeeritavatelt, milliste valdkondade andmeid kogutakse andmeaitadesse praegu ja milliste omi planeeritakse koguda tulevikus. Kõigi küsitletute vastustest selgus, et praegu kogutakse peamiselt tervishoiu, info ja side, hariduse ning veonduse-laonduse valdkondade andmeid. Tulevikus planeeritakse peamiselt koguda lisaks hariduse, töötleva tööstuse ja põllumajanduse andmeid. Kui aga vaadelda eraldi riigisektori spetsialistide vastuseid, siis selgub, et riigisektori andmeaitadesse kogutakse eelkõige tervishoiu, info ja side, finants- ja kindlustustegevuse, hariduse ning halduse valdkondade andmeid. Valdkondi, millest andmeid üldse ei kogutud, ei esinenud.

4.1.2.2 Teadlikkus

Intervjueeritavatest ja ankeetidele vastanutest olid 92% kõrgharidusega. Seega on käesolevat valimit arvestades andmeaitadega tegelejatel olemas üldine laiem vaade probleemide lahendamisele, sealhulgas andmetöötlemisele ja infotehnoloogia vahendite kasutamisele.

Andmeaitade uuring

Vastused ankeetides esitatud tehnilist laadi küsimustele olid enamasti konkreetseid ("Ei oska öelda" tüüpi vastused olid vähemuses), mis näitab vastajate orienteerumist andmeaitade tehnilistes aspektides.

Küsimusele "Kas olete tutvunud (kursis) väga suurte, variatiivsete andmetüüpidega ja erinevatest allikatest kogutud andmete töötlemise ning analüüsi (nn big data analytics) tehnoloogiatega?" vastas jaatavalt alla 11% küsitletuist. See näitab, et andmeaitade edasiste perspektiivide osas on teadlikkus väike.

Kokkuvõttes võib järeldada, et vastajate üldine IT-alane teadlikkus ning tehnilised teadmised andmeaitade valdkonnast on paremad kui teadlikkus selle valdkonna suundumustest. Kohati võib olla probleeme andmelao projektide hankimise ja läbiviimisega. Vastanutel on soov täiendada oma andmeaitade alaseid teadmisi ja oskusi (83 % soovis koolitusi).

4.1.2.3 Motivatsioon

Küsimusele "Kas andmeaidad võiksid aidata kaasa otsuste vastuvõtmise kvaliteedi parendamisele?" vastas 92% anketeerituist jaatavalt. Märkiti muuhulgas, et andmeaidad võimaldavad välja tuua suuremat pilti ja seoseid valdkondade vahel, kus ilma andmekaeveveta see praegu võimalik ei ole; et suur osa äriotsustest tehakse seni tunnetuse, mitte faktide baasil.

Kokkuvõttes hindavad vastanud kõrgelt andmeaitade poolt pakutavaid uusi võimalusi. Algatus andmeaitadega tegelemiseks tuleneb enamasti kasutajate või ettevõtte vajadustest (83 % vastajatest), luues seega tugeva motivatsiooni andmeaitadega tegelejate poolel.

4.1.2.4 Avaandmed

Avaliku sektori andmete avamine soodustab andmeaitade loomist, andmeanalüüsi ja selle tulemuste kasutamist otsuste vastuvõtmisel sõltumatult andmete omanikest. See annab ka erasektorile võimaluse kasutada avaliku sektori andmeid oma andmeaitades ja muudes rakendustes. Võimalust, et konkreetse valdkonna andmed võiksid olla avaandmed (küsimus "Kas teie valdkonna andmeaida andmed võiksid olla avaandmed või lingitud avaandmed?"), toetas täiel või osalisel määral 63% küsimusele vastanutest. Eitavalt vastanud (17% küsimusele vastanutest) põhjendasid otsust sellega, et tegu on äriettevõtete andmete, delikaatsete isikuandmetega või teaduslikel eesmärkidel kasutatavate andmetega, mis ei ole mõeldud üldkasutuseks.

Andmeaitade uuring

4.1.2.5 Metoodilised soovitusel andmete kogumiseks, töötlemiseks ja väljastamiseks

Küsimusele "Kas avaliku sektori asutused võiksid andmeaitade loomisel ja halduses õppida Eesti erasektori praktikatest?" vastas kaks kolmandikku anketeerituist positiivselt, eitava vastuse andis vaid üks vastaja.

Ankeetides ja intervjuudes ning välisriikide praktika analüüsis uuriti parimaid andmelao algatamise, loomise ja haldamise praktikaid. Tulemuseks saime palju häid soovitusi, mis on esitatud selle uuringu lõpparuandes. Nimetatud soovitude põhjal on tehtud vastavad ettepanekud, mis on esitatud selle aruande kokkuvõttes.

4.1.2.6 Isikuandmete kaitse ja andmete kodeerimine

Andmeaitade juurutamisel riigisektori infosüsteemides on oluline, millised andmed on vaja andmeaita sisestamiseks kodeerida (eelkõige isikuandmete kaitsest lähtuvalt) ja millised mitte.

Avaliku teabe seadus, Vabariigi Valitsuse määrus "Riigi infosüsteemi haldussüsteem" ja riikliku statistika seadus ei sätesta kodeerimise vajadust.

Isikuandmete kaitse seadus iseenesest ei keela isikuandmete töötlemist teistes andmekogudes, sealhulgas andmeaitades. Vastavalt isikuandmete kaitse seaduse § 5 on isikuandmete töötlemine iga isikuandmetega tehtav toiming, sealhulgas isikuandmete kogumine, salvestamine, korrastamine, säilitamine, muutmine ja avalikustamine, juurdepääsu võimaldamine isikuandmetele, päringute teostamine ja väljavõtete tegemine, isikuandmete kasutamine, edastamine, ristkasutamine, ühendamine, sulgemine, kustutamine või hävitamine.

Isikuandmete töötlemisel tuleb jälgida vastavaid põhimõtteid (isikuandmete kaitse seaduse § 6, sealhulgas turvalisuse põhimõte) ja isikuandmete töötlemise lubatavust (sh isikuandmete kaitse seaduse § 10 kuni 14 sätestatu, nt andmesubjekti nõusolek või töötlemine avaliku ülesande täitmise käigus). Kui isikuandmete töötlemise nõuded on rahuldatud, siis pole vastavaid isikuandmeid vaja andmeaita sisestamiseks kodeerida.

Kodeerimise nõude sätestab isikuandmete kaitse seaduse § 16 lõige 1 isikuandmete töötlemisel teadusuuringu või riikliku statistika vajadusteks: "Andmesubjekti nõusolekuta võib teadusuuringu või riikliku statistika vajadusteks töödelda andmesubjekti kohta käivaid andmeid üksnes kodeeritud kujul. Enne isikuandmete üleandmist teadusuuringu või riikliku

Andmeaitade uuring

statistika vajadustel töötlemiseks asendatakse isiku tuvastamist võimaldavad andmed koodiga. Tagasikodeerimine ja selle võimalus on lubatud ainult täiendavate teadusuuringute või riikliku statistika vajadusteks. Isikuandmete töötleja määrab nimeliselt isiku, kellel on ligipääs tagasikodeerimist võimaldavatele andmetele".

Sama paragrahvi lõige 2 räägib eelmisele lõikele vastu, tuues sisse ka olukorra, kus teadusuuringu või riikliku statistika vajadusteks on lubatud andmesubjekti nõusolekuta tema kohta käivate andmete töötlemine andmesubjekti tuvastamist võimaldaval kujul.

Sama paragrahvi lõige 4 sätestab: "Kogutud isikuandmeid on lubatud töödelda teadusuuringu või riikliku statistika vajadusteks, olenemata sellest, millisel eesmärgil neid isikuandmeid algselt koguti", täpsustamata, kas jutt on kodeeritud või kodeerimata andmetest, töötlemisest andmesubjekti nõusolekul või ilma nõusolekuta.

Need sõnastused on otstarbekas täpsustada.

4.1.2.7 Andmeaitade tehnoloogia

Küsimusele "Kas erinevate andmekogude andmete ühendamiseks ühte andmeaita on vaja eraldi tehnoloogilisi lahendusi?" vastati kolmandikul juhtudel eitavalt ning ligi pooltel juhtudel positiivselt (jah või mõnikord). Seejuures märgitakse, et lähteandmebaasid on erinevate andmebaasimootorite peale ehitatud ning on vaja andmete erinevaid struktuure ühendavaid/tõlgendavaid liideseid. Kolmel korral mainitakse X-tee lahendusi, kahel korral - Extract, Transform and Load (ETL) tööriistu / raamistikke. Võib seega järeldada, et andmeaitade loomine nõuab paljudel juhtudel eraldi tehnoloogilisi lahendusi. Erasektori spetsialistidest vaid 10% märgivad sellist vajadust, millest võib järeldada, et erasektoris on sellisele vajadusele mõeldud aegsasti ja kogu andmemajandus on ehitatud ühilduvatele platvormidele.

Tootjatest märgiti intervjuudes enamasti Sybase, Oracle, SAP ja Microsofti tarkvara, samuti mitmesuguseid kombinatsioone nendest ja muudest tarkvaraplatvormidest. Andmeaitade halduseks kasutatava tarkvara platvormi omadustega ollakse valdavalt kas täiesti või osaliselt rahul.

Perspektiivselt hindasid anketeeritud kõige olulisemateks andmeaitade tarkvara omadusteks päringute jõudlust, häid administreerimisvahendeid, tõrketaluvust, integreeruvust olemasolevasse IT keskkonda ja ärianalüüsi süsteemide toetust. Kõige vähemtähtsamaks hinnati andmete kompressiooni, pilvearvutuse toetust ja suurte andmemahtude toetust.

Andmeaitade uuring

Arvestades tendentsi andmebaasidest võetud andmete osatähtsuse vähenemisele ning erinevatest andmeallikatest pärinevate osaliselt mittestruktureeritud andmete suuremale kasutusele, tuleks andmeaida tarkvara valikul siiski kaaluda ka neid omadusi. On märgata kavatsusi keerukama ärianalüüsi (andmekaevandus, voogandmete reaajas analüüs jt) tegemiseks tulevikus võrreldes praegu valdavalt tehtava aegriidade, ennustava ja on-line analüüsi ning summeeriva analüütilise töötlusega.

4.2 RAHVUSVAHELISE PRAKTIKA JA TEHNOLOOGIATRENDIDE ANALÜÜSI TULEMUSED

Peamised trendid, mis tekitavad praeguse revolutsiooni andmeaitades ja andmekaevanduses on spetsialistide arvates järgmised [McKinsey 2011]:

1. Läbilaskevõime tähtsuse oluline tõus.
2. Reaalaja töötlus ja andmeanalüüs saab reaalsuseks.
3. Uute andmekeskuste (Data center) kasutamine on tugevalt lihtsustunud.
4. Andmekaevandamine on lihtsustunud ja tema kasutamine laienenud.
5. Võime kasutada ülisuuri andmemahte (nn Big Data) andmeanalüüsiks hakkab organisatsioonide efektiivsust mõjutama.
6. Operatsioonilisi andmeid ja andmelao andmeid hakatakse koos töötlemata.
7. Süsteemide läbilaskevõime suureneb hüppeliselt seoses uute mälutüüpide kasutuselevõetuga (Flash ja DRAM).

Andmaks käesolevas uuringus läbiviidud intervjuude ja ankeetide tulemustele teatud rahvusvahelist tausta ja osadele küsimustele ka võrdlusvõimalust kasutasime mõningate andmeaitade alaste rahvusvaheliste uuringute tulemusi [McKendrick 2011, Russom 2009, Butler 2012, Gartner 2013] ja analüüsisime 3 välisriigi Suurbritannia, Hollandi ja Ameerika Ühendriikide andmelaondust puudutavaid regulatsioone, andmete andmeaita integreerimise praktikaid, nn Big Data mõju andmeaitade arengule tulevikus, andmelaonduse häid praktikaid ja avaandmete seisukorda. Lisaks teabematerjali analüüsile nimetatud valdkondades intervjuueerisime nendes küsimustes ka igast nimetatud riigist 1 andmeaitade eksperti.

Tabelis 2 võrdleme rahvusvaheliste uuringute, välismaise praktika analüüsi ja meie uuringu tulemusi põhilistes andmelaonduse küsimustes ja trendides. Samas esitame ka soovitusid Eestile.

Tabel 2 Rahvusvaheliste uuringute ja Eesti uuringu tulemuste võrdlus

Rahvusvaheline andmeaitade uuring või praktika	Eesti andmeaitade uuring	Soovitused Eestile
<p>Paljude juhtudel jäävad ärianalüüsi rakendused isoleerituks ja pole hästi seostatud äriprotsessidega [McKendrick 2011].</p>	<p>Andmeaidad pole loodud toetama äriprotsesse ja andmeanalüüs pole sinna integreeritud [intervjuu erasektori spetsialistid].</p> <p>Üle 90% anketeeritutest arvab, et andmeaidad võivad aidata kaasa otsuste vastuvõtmise kvaliteedi parendamisele [lõpparuanne ptk 4.3.2].</p>	<p>Ärianalüüsi rakendused peaks juba projekti algatamisel planeerima toetamaks äri- ja otsustusprotsesse.</p>
<p>Paljud rakendused on loodud organisatsiooni siseselt põhiliselt analüütikute ja tippjuhtide jaoks [McKendrick 2011].</p>	<p>Sageli on andmelao kasutajate ring väga kitsas.</p> <p>Meie arvates, peab andmelaol olema taga ikkagi kogu kasutajate fännklubi, vastutajad ja protsessid [intervjuu erasektori spetsialistid].</p>	<p>Tuleks oluliselt laiendada ärianalüüsi rakenduste kasutajate gruppe nii organisatsiooni sees kui väljaspool. Näiteks erineva tasemega juhid, analüütikud, statistikud, teadlased jt huvigrupid võiksid andmeloa analüüsi kasutada vastava kasutusõigusega või vabalt (sõltub andmetest).</p>
<p>Umbes pooled vastajatest kavatsevad oluliselt suurendada mittestruktureeritud andmete analüüsi lähema 5 aasta jooksul [McKendrick 2011].</p>	<p>Nähakse relatsiooniliste andmebaaside ning struktureeritud pärandandmebaaside andmete osatähtsuse vähenemist (46% ja 27%) ja XML andmete ning veebi logide osatähtsuse suurenemist (vastavalt 46% ja 22%). Suureneb ka video/audio andmete kasutamine [lõpparuanne ptk 4.5.1].</p>	<p>Mitte-struktureeritud andmete töötlemiseks võiks kasutada vajadusel noSQL liidestamist olemasoleva andmeaida platvormiga.</p>
<p>40% küsitlusest soovivad näha rohkem andmeladude sulandumist olemasolevasse infrastruktuuri [McKendrick 2011].</p>	<p>Andmeaida platvormi integreeritavusega olemasolevasse IT keskkonda oli rahul 57% vastajatest [lõpparuanne ptk 4.5.2].</p>	<p>Projekti ja tellimuse planeerimisel oleks vaja pöörata tähelepanu ka andmeaida platvormi integreeritavusele olemasolevasse infrastruktuuri.</p>

Andmeaitade uuring

<p>Kasvav vajadus reaalaja või reaalajalähedaste andmeaitade järele [Russom 2009].</p>	<p>Eesti Statistikaamet on ühinemas EL SensusHub projektiga [SensusHub 2013], mille tulemusena tuleb meie statistika-ametil esitada rahvastikustatistika andmeid EL nõudmisel (on-demand) ja vajadusel vähemalt 10 min värskusega (nn pull technology). See on reaalajalähedane nõue andmete järele [intervjuu riigisektori spetsialistiga].</p> <p>Anketeeritud märkisid, et andmeaitade kasutamine võimaldaks liikuda rohkem reaalajas info põhjal tehtavate ja terviklikumate otsusteni [lõpparuanne ptk 4.3.2].</p>	<p>Vajadus kiiresti ja andmetepõhiselt otsustada tingib ka vajaduse reaalaja andmeaitade järele. Taoline vajadus on Eesti asutustel juba tekkinud ja sellest sõltuvalt tuleb üle minna uuele mõtteviisile andmeanalüüsi läbiviimisel ja vastavatele nn Big Data tehnoloogiatele.</p>
<p>Veebiteenuste üha suurem kasutamine erinevate andmete ja rakendusplatvormide liidestamisel [Russom 2009].</p>	<p>X-tee teenuseid kasutatakse andmete integreerimiseks juba praegu, küll mitte veel andmeaitade endi liidestamiseks.</p>	<p>Veebiteenuste kasutamine andmeaitade ja teiste andmeallikate liidestamisel ning andmete integreerimisel.</p>
<p>On oodata andmemahutude kiiret kasvu [Russom 2009, McKendrick 2011, Gartner 2013].</p>	<p>Eestis pole eriti asutusi, kus oleks üle paari TB andmeid. Andmemahud pole tähtsad, pigem on tähtis, et päringud jookseksid talutava kiirusega.</p> <p>Ülioluline on kompressioonitehnika. Vastupidi üldlevinud ettekujutusele, et andmeladu, mis koondab mitu operatiivandmebaasi muutub mahult väga suureks võimaldab just kompressiooni tehnikate kasutamine luua andmelao, mis on väiksem kui tema lähteallikate summa [intervjuu erasektori spetsialistiga].</p> <p>Ligi neli viiendikku anketeerituist vastas, et andmeaida maht jääb alla 10 TB piiri [lõpparuanne ptk 4.3.3].</p> <p>Alla 11% küsitletuist oli kursis nn Big Data analüüsi võimalustega.</p>	<p>Vajadusel andmeaida platvorm välja vahetada või kasutada Hadoop jt liideseid.</p>

Andmeaitade uuring

<p>Eelda, et andmeanalüüs saab järgmise põlvkonna andmeaitade platvormide prioriteediks [Russom 2009].</p>	<p>Keerukamaid andmeanalüüse nagu näiteks andmekaevandus (30%), ennustav analüüs (60%), on-line analüüs (70%) teevad juba praegu suhteliselt paljud erasektori firmad. Riigisektoris vastavalt (6, 41 ja 41 %) teevad neid praegu ja planeerivad tulevikus teha vastavalt 59, 29 ja 29% vastanutest (riigisektori spetsialistidest) [Lõpparuanne ptk 4.3.3].</p>	<p>Tõsta teadlikkust andmeanalüüsi võimalustest ja uutest trendidest riigisektoris.</p>
<p>Andmeaitade loogiline sidumine on megatrend [Gartner 2013].</p>	<p>Intervjuude ja ankeetide põhjal selgus, et Eestis on valdav andmeait, mis füüsiliselt integreerib andmed lähteallikatest ühte hoidlasse analüüsi tarbeks.</p>	<p>Võiks laiendada praegust Eestis põhiliselt levinud klassikalist andmeaida käsitlust analüütilise platvormi ja virtuaalsete jt andmeaitade definitsioonide suunas.</p> <p>Soovitame ka andmeaitade/andmestike linkimist kasutades lingitud andmete standardeid (RDF, SPARQL jt).</p>
<p>Suurbritannia valitsus on väga vähe süstemaatiliselt kasutanud ärianalüüsi võimalusi ja vahendeid selleks, et toetada oma kõrgetasemelisi otsuseid [McKenna 2011].</p>	<p>Põhiprobleem, miks andmeaitade kasutamise juhtimisel on nõrk seisneb selles, et andmeait pole loodud juhtimisprotsessi toetavana [intervjuu erasektori spetsialistiga].</p>	<p>Soovitame laiemat ja paremat ärianalüüsi kasutamist kõigis valdkondades kaasa arvatud riigi juhtimine (selleks sobib näiteks andmetepõhine otsuste vastuvõtmine).</p>
<p>Tuleks luua selged reeglid andmeaida loomise projekti tellimiseks ja läbiviimiseks [intervjuueeritav Inglismaalt].</p>	<p>Riigisektoris on vähene teadlikkus andmeaitade valdkonna kohta. Ei ole reeglina selget ettekujutust sellest, et mis on andmeladu, kuidas seda ehitatakse, kuidas projekti läbi viia, kuidas vastavaid hankeid korraldada. Andmelao lahenduste hanked peaksid olema mõnevõrra spetsiifilised. Hangetel kohtab pealisehitusena tihti nõudeid nt objektorienteerituse vmt enda arendatud tarkvaralahenduste kohta. Tundub, et hanke tegijad ei tunne andmeaitade tehnoloogiat. Sageli tehakse ühe hankega analüüs ja</p>	<p>Tõsta teadlikkust andmeanalüüsi ja seda võimaldavate tehnoloogiate osas. Lihtsustada riigihanke protsessi tehes selle paindlikumaks ja kohandades andmeaida loomise tsüklitega.</p> <p>Erasektoris on andmeaitade tellimused palju dünaamilisemad, seega võiks riik õppida andmeaida projektide tellimist erasektorilt.</p> <p>Peaks selgelt sätestama tellija</p>

Andmeaitade uuring

	<p>realisatsioon. Analüüsi käigus aga alles selgub tegelik skoop ja tööde maht [intervjuu erasektori spetsialistid].</p> <p>Kaks kolmandikku anketeerituist vastas, et avaliku sektori asutused võiksid andme-aitade loomisel ja halduses õppida Eesti erasektori praktikatest [lõpparuanne ptk 4.3.3].</p>	<p>suhte andmete algallikate omanike suhtes.</p> <p>Tellida väiksematelt ja keskmise suurusega teenusepakkujailt, sest see peaks võimaldama kiiremini kasutusele võtta uusimaid tehnoloogiad.</p>
<p>Peaks looma kõigile andmelao projekti osapooltele üheselt mõistetava sõnastiku [intervjueeritav Inglismaalt].</p>	<p>Sellega (st nn Big Data analüütikaga) seonduvad ka andmete sisu kvaliteet ja andmete tähenduse kvaliteet ehk andmete semantika, mis vajavad korrastamist ja suurt tähelepanu [intervjuu erasektori spetsialistid].</p> <p>23% anketeeritustest on olnud probleeme oma valdkonna andmeaitade semantilise koosvõimega teiste andmekogudega (ka andmeaitadega). Märgitakse, et kui andmeaida koosseis ja semantika on ette antud, siis on suur töö leida semantiline kooskõla andmekogude andmetega, mis on andmeaidale andmete andjateks [lõpparuanne ptk 4.3.3].</p>	<p>Vastavate inim- ja masinloetavate sõnastike (ontoloogiate) loomine.</p>
<p>Hoia fookus andmekvaliteedil isegi kui andmemahud kasvavad, sellega tagad kasutajate usalduse [Butler 2012].</p>	<p>Teades milline on andmete kvaliteet saab anda usaldatavuse protsendi aruannete jaoks. Andmeaida põhjal tehtud aruandluse andmekvaliteet on alati parem kui operatiivsüsteemide põhjal tehtud</p> <p>Andmeaida vastutavad ja volitatud töötajad saavad takistada valede andmete andmeaita sattumist [intervjuu erasektori spetsialist].</p> <p>Suhteliselt väiksem oli rahulolu andmete sidususe, kättesaadavuse ja selgusega. Andmeaitade andmekvaliteedi probleemide põhilised allikad on kvaliteediprobleemid andmeaitade aluseks olevates andmekogudes [lõpparuanne ptk 4.2.2].</p>	<p>Eriliselt tähtsustada andmekvaliteedi rolli andmeaitade andmeanalüüsi usaldusväärsete tulemuste saamiseks. Töötada välja meetmed/reeglid andmekvaliteedi tagamiseks andmekogudes ja muutuste haldamiseks.</p>

Andmeaitade uuring

<p>Avaandmestike publitseerimine: Suurbritannia 9234, Holland 5193 ja Ameerika Ühleriigid 373029 andmestikku.</p>	<p>Publitseeritud on 6 avaandmestikku, nendest osa näidised (vt opendata.riik.ee). Alates 1. aprillist 2013 jõustub Eestis EL PSI-direktiiv [PSI directive 2003].</p>	<p>Alustada avaandmestike publitseerimist kõigis riigiasutustes. Töötada välja meetmed huvi tekitamiseks andmete avalikustamise vastu. Luu veel pilootrakendusi avaandmete baasil ja publitseerida parimad praktikad. Alustada lingitud avaandmete loomist.</p>
<p>Suurbritannia eksperdid (50%) tunnetavad, et erasektori juurdepääs andmestandarditele ja riigi andmekogumitele on vajalik uute töökohade ja paremate teenuste pakkumiseks. Samas suur osa Suurbritannia riigiametnikke ei näe, et andmete avamine erasektorile oleks prioriteet [Listpoint 2012].</p>	<p>Kaks kolmandikku küsitletutest vastasid, et nende valdkonna andmed võiksid olla kas täielikult või osaliselt avaandmed [lõpparuanne ptk 4.3.3].</p>	<p>Tõsta riigiametnike teadlikkust avaandmetest, riigi plaanidest selles valdkonnas, nende rollist avaandmete tagamisel/kasutamisel ja avaandmete vajalikkusest.</p>

Selle tabeli andmetest võime järeldada, et välismaa ja Eesti praktika on paljuski sarnased. Näiteks on sarnasused järgmistes valdkondades: ärianalüüsi isoleeritus äri ja juhtimisprotsessidest, rakenduste kitsas kasutajaskond organisatsioonis ja väljaspool, mittestruktureeritud andmete üha suurenev kasutamine, suurenev nõudmine reaallajalahenduste järele, andmekvaliteedi tähtsustamine, andmete üheselt mõistetava sõnastiku vajadus, andmeaitade projekti tellimine ja läbi viimine.

Erinevused aga on järgmistes valdkondades: andmeaitade ja analüütilise platvormi käsitlus, suurte või ülisuurte andmemahutuste analüüsimine, erinevate (eriti keerukamate) andmeanalüüsi meetodite kasutamine, loogiliste andmeaitade trendid, avaandmete publitseerimine.

Andmeaitade uuring

Huvitav on see, et vaid kaks anketeeritud arvas otseselt, et teiste riikide lahendusi ja regulatsioone peaks kasutama Eesti riigisektori andmeaitade loomisel. Samas enamus (74%) vastas sellele küsimusele, et „Ei oska öelda“, mis näitab rahvusvahelisse praktika mittetundmist. Küsimusele „Kas andmeaitade tehnoloogia areng teistes riikides sunnib meie asutustes ka sellega tegelema?“ vastasid 41% riigi- ja erasektori andmeaitade spetsialistidest eitavalt, kusjuures riigisektori spetsialistidest vastas eitavalt 53 %. Nagu aga nägime uuringu tulemustest, on Eestil teiste riikidega palju sarnasusi andmelaonduse valdkonnas, seda nii regulatsioonide kui lahenduste ning tulevikuvisionide osas.

5 KOKKUVÕTE

Uuringu põhihüpoteesiks oli, et riigi infosüsteemi andmeaitade valdkonnas on vajalikud nii poliitilised, seadusandlikud, organisatoorsed kui tehnoloogilised muutused.

Uuringu tulemused näitasid, et põhihüpotees pidas paika poliitiliste, seadusandlike ja organisatoorsete muutuste vajalikkuse osas. Tehnoloogiliste muutuste osas on pigem vaja jälgida tehnoloogia arengutrende äriprotsesside ja otsustusprotsesside muutuvate vajaduste rahuldamiseks.

Seoses muutuste vajadusega pakume välja rea lahendusi ja ettepanekuid nii poliitika, seadusandluse kui andmeladude organiseerimise valdkonnas. Ettepanekute juures on toodud ka soovitused selle kohta, kes vastavaid ettepanekuid peaks ellu viima. Üldine põhimõte on, et ettepanekud viib ellu vastava tegevuse teostaja, jälgides seadusi, standardeid, häid praktikaid jne. Seadusandlike regulatsioonidega seotud soovitused peaksid läbi viima Majandus- ja Kommunikatsiooniministeerium, AKI, Riigikantselei, Justiitsministeerium ning vastavalt vajadusele muud osapooled.

Allpool esitatud soovituste ja ettepanekute koostamisel lähtuti anketeerimise analüüsi tulemustest, intervjuudel räägitust, teabematerjalide analüüsist, välismaisest kogemusest (ka välisintervjuude tulemustest) ja isiklikest ekspertteadmistest ning kogemustest.

Uuringus pakutavad lahendused ja ettepanekud on järgnevad.

- Arvestada iga konkreetse regulatsiooni muutmisel ja tekitamisel andmeaitade jaoks nii sellise tegevuse tugevusi ja võimalusi kui ka nõrkusi ja ohte, muuhulgas tehnoloogilisi arenguid, eelseisvaid muudatusi seadusandluses, võimalikke lisakulutusi ja privaatsuse probleeme. Kindlasti tuleks SWOT analüüs läbi viia võimalike andmeaitade temaatikast lähtuvate seadusemuudatuste puhul, mis lisanduvad käesoleva uuringu ettepanekutele.
- Määratleda avaliku teabe seaduse tekstis ilmutatult andmete kogumise mõiste, tekitamaks ühese arusaamise sellest, et avaliku teabe seaduse § 43³ lõige 2 ei puuduta andmete ülekannet teistest andmekogudest ning ei takista seega andmeaitade loomist ja kasutamist.
- Vältida võimalust mööda lisanduvaid kitsendusi isikuandmete kasutamises, näiteks seoses käimasoleva Euroopa Komisjoni isikuandmete kaitse reformiga (isikuandmete kaitse üldmäärus ja direktiiv).

Andmeaitade uuring

- Täpsustada avaliku teabe seaduses, millised õigusaktid on andmekogu (sh andmeaida) asutamiseks lubatavad (muuhulgas, korrigeerida Riigi Teatajas avaliku teabe seaduse § 43³ lõikes 1 toodud õigusakti linki). Andmeaitade puhul tuleks lihtsustava tegurina kasuks võimalikult madala taseme õigusakti kasutamine. Seega võiks õigusakti tase olla minimaalselt selline, mis hõlmab andmeaita kuuluvate andmeallikate valdkondi.
- Peale eelmise ettepaneku elluviimist viia olemasolevad andmeaidad kooskõlla õigusaktidega, asutades need vastavalt avaliku teabe seaduse § 43³.
- Võimaldamaks saada andmeaitade jaoks infot riigi infosüsteemi põhiandmete kohta, luua RIHAs põhiandmete kohta käivad päringud, täpsustada RIHAs olevaid andmeid ning rakendada järjekindlamalt semantilise koosvõime raamistikku. Soovitame inim- ja masinloetavate sõnastike (ontoloogiate) loomist andmeaita integreeritavate andmete tähendusest arusaamiseks. Neid tegevusi peaks läbi viima RIA.
- Arvestades vajadust andmeanalüüsi usaldusväärsete tulemuste saamiseks, tuleks andmeaitade andmekvaliteedi rolli eriliselt tähtsustada nii seadusandlikul kui ka iga konkreetse projekti tasemel.
- Sätestada avaliku teabe seaduses ilmutatult vastutava ja volitatud töötaja vastutus andmekogu andmete kvaliteedi eest ja nõue kehtestada ning kooskõlastada kriteeriumid, mille alusel hinnatakse andmete kvaliteeti. Töötada välja meetmed/reeglid andmekogude andmekoosseisude muutuste haldamiseks ja nendest teavitamiseks.
- Lisada Vabariigi Valitsuse määrusse "Riigi infosüsteemi haldussüsteem" RIHA andmekogude alamregistrisse kantavate andmete koosseisu kriteeriumid, mille alusel hinnatakse andmete kvaliteeti.
- Määratleda avaliku teabe seaduses või muudes õigusaktides andmete töötlus.
- Piiritleda riikliku statistika seaduse §34 lõikes 3 kasutatavad võimalused, näiteks kas kasutatavate võimaluste otstarbekuse kriteeriumi abil või loetledes tuvastamise võimaluste klassid.
- Kaaluda konkreetsete andmeaitade loomisel rahvusvaheliste, sh valdkondlike standardite kasutamist. Võimalusel tõlkida eesti keelde ning võtta üle eesti standardiks olemasolevad tervishoiu valdkonna standardid ISO/TS 29585:2010 ja ISO/TS 29585:2010.
- Soovitame tõsta spetsialistide teadlikkust andmelaonduse valdkonnas ja selle tehnoloogia perspektiivide osas, korraldades koolituskursusi ja temaatilisi seminare,

Andmeaitade uuring

töötades välja ja tehes vabalt kättesaadavaks andmeaitade teemalisi juhendmaterjale jne. Koolituste raames võiks ühtlasi laiendada praegust Eestis põhiliselt levinud klassikalist andmeaida käsitlust analüütilise platvormi ja loogiliste/virtuaalsete jt andmeaitade käsitlustega. Koolituste läbiviimiseks võiks RIA korraldada riigihanke.

- Et andmeaitu on lihtsam luua hästi kättesaadavate avaandmete baasil, siis soovitame alustada avaandmestike (eelkõige andmekogude, seejärel ka andmeaitade) publitseerimist kõigis riigiasutustes. Töötada välja meetmed huvi tekitamiseks andmete avalikustamise vastu. Luua pilootrakendusi avaandmete baasil ja publitseerida parimad praktikad. Alustada lingitud avaandmete loomist. Selle soodustamiseks võiksid RISO ja RIA korraldada infopäevi ja seminare. Seminarid peaksid tõstma riigiametnike teadlikkust riigi plaanidest avaandmete valdkonnas ja avaandmete kasulikkusest (sh seoses andmelaondusega). Riigiasutused (andmete omanikud) peaksid kindlaks määrama erinevate ametnike rollid nende poolt hallatavate andmete avalikustamisel ja ka avaandmete kasutamisel. Andmete avalikustamisele peaks asutustes seadma kõrge prioriteedi. Asutustes tuleks luua pilootrakendusi avaandmete baasil, publitseerida parimad praktikad ning alustada lingitud avaandmete loomist.
- Lähtudes rahvusvahelisest kogemusest tuleks lihtsustada riigihanke protsessi, arvestades pakkumuskutse tehnilises kirjelduses andmeaida loomise tsüklitega. Seda peaksid tegema eelkõige andmeaida projekti osapooled, sätestades selgelt andmeaida projekti tellija suhte andmete algallikate omanikega.
- Tuleks oluliselt laiendada ärianalüüsi rakenduste kasutajate gruppe nii organisatsiooni sees kui väljaspool. Näiteks erineva tasemega juhid, analüütikud, statistikud, teadlased jt huvigrupid võiksid andmela analüüsi kasutada vastava kasutusõigusega või vabalt (sõltub andmetest). Soovitame ärianalüüsi rakendused juba projekti algatamisel planeerida toetamaks erineva tasemega äri- ja otsustusprotsesse ja seega rahuldamiseks erinevate kasutajagruppide vajadusi.
- Vajadus kiiresti ja andmetepõhiselt otsustada tingib ka vajaduse reaalaja andmeaitade järele. Taoline vajadus on Eesti asutustel juba tekkinud ja sellest sõltuvalt tuleb vajadusel üle minna uuele mõtteviisile andmeanalüüsi läbiviimisel ja vastavatele nn Big Data tehnoloogiatele. Soovitame selle teema lülitada koolituste programmi.
- Soovitame andmeaitade/andmestike linkimist kasutades lingitud andmete standardeid (RDF, SPARQL jt) ja veebiteenuste kasutamist andmeaitade ja teiste

andmeallikate liidestamisel ning andmete integreerimisel. See teema peaks olema üks koolitusprogrammi osa.

- Soovitame õppida erasektori ja välismaa praktikatest ning headest tavadest andmelaonduse, ülisuurte andmemahtude töötlemise ja analüüsi, avaandmete ja nende linkimise valdkondades. Soovitame selle teema lülitada koolituste programmi.
- Andmete kogumisel, töötlemisel ja väljastamisel avaliku sektori andmeaitadest on otstarbekas määratleda mingi sobiv elutsükli mudel. Näiteks võib aluseks võtta üldise andmeaida elutsükli, mis koosneb järgmistest etappidest: algatamine, väljatöötamine, kasutamine ja hooldus. Seejuures võib kasutada mitmesuguseid elutsükli mudeli tüüpe, näiteks inkrementarendus, evolutsioonarendus, spiraalmudel jt. Soovitame selle teema lülitada koolituste programmi.
- Andmeaida algatamisel tuleks andmeaida kasutajate ja teiste osapoolte koostöös määratleda ja heaks kiita andmeaida loomise vajadus, sõnastades selle rolli, tähtsuse ja kasu tulevaste kasutajate jaoks ning nende kasutajate kaudu laiemalt loodavad hüved; hinnata tasuvust; läbida koolitus, mille esmane teema on andmelao loomise eesmärk ja meetodika; tutvuda edukate lahendustega; koguda soove, korraldada, ajurünnakud jmt; püstitada eesmärgid; püstitada ülesanne; määratleda vastutavad, aruandekohustuslikud, konsulteeritavad ja informeeritavad osapooled; hoida lahus parenduslikud (kiirem, täpsem, kvaliteetsem, odavam analüüs) ja uuenduslikud (uued teadmised ja võimalused) aspektid; läbi mõelda andmekooseis ja aruanded; analüüsida, kas andmeaidas on delikaatseid isikuandmeid; spetsifitseerida käideldavuse, tervikluse, konfidentsiaalsuse nõuded; võimalusel jälgida kindlat andmete standardit; eeldada, et nende andmekogude andmetega, mida andmeait kasutab, tuleb lahendada kvaliteediprobleeme ning planeerida selleks aega ja ressursse; selgitada välja, kas andmeaitade andmeid saab välistelt osapooltelt ja kui ei saa, siis kas probleem on tehniline (mida saab parandada) või on küsimus pigem usalduses; võimalusel mõelda andmeaida peale juba uue andmekogu loomisel; jälgida seadusandluses ettenähtud nõudeid andmeaida asutamisele, registreerimisele, turbele jne; võimalusel kasutada vastavaid standardeid.
- Andmeaitade väljatöötamisel, kasutamisel ja hooldusel tuleks jälgida mingit praktikas järele proovitud andmeaitade arendamise meetodikat; alustada etapiviisiliselt, väikese mahuga; soodustada kommunikatsiooni IT poole ja äriinimeste vahel, tutvuda andmeaitade tehnoloogiate pakujate ülevaadetega; vaadata nii andmeaida tarkvara platvorme kui ka seda, kas meil pakutakse tuge ja milline on selle toe kvaliteet; soovitatavalt mitte teha andmete teisendusi nende andmeaita laadimise käigus - kõik puuduvate näitajate arvutused tehakse peale andmete ülekandmist andmeaidas;

Andmeaitade uuring

kuna probleemid võivad olla andmete hõive ja kvaliteediga, kaaluda automaatset andmehõivet ja andmepuhastust.

- Andmeaitade puhul tuleks valida sobiv metoodika, sh nt Kimballi metoodika, mille puhul andmeaida andmed koosnevad põhifaktide tabeli(te)st ning nendega seotud kontekstiandmeid sisaldavatest "dimensioonidest"; normaliseeritud lähenemine, mille puhul andmed andmeaidas normaliseeritakse nagu tavapärasel andmebaasis; spetsiaalsed tehnikad struktureerimata andmete kogumiseks ja kasutamiseks; tootjate poolt pakutavaid metoodikaid, eriti kui suund on pigem andmete virtualiseerimisele, suurte töötlemata andmekogumite kasutamisele ning andmete analüüsile. Soovitame selle teema lülitada koolituste programmi.
- Andmeaitade andmehõivet ja uute andmeaitade loomist spetsiaalselt laiendada ei ole vaja, küll aga on mõtet andmeaitade võimalusi ja tehnoloogiat laiemalt tutvustada, et asutused oleksid sellest teadlikud ning oskaksid vajadusel andmeaida projekti algatada. Soovitame selle teema lülitada koolituste programmi.
- Kasutada andmeaitades kodeerimist isikuandmete töötlemisel andmesubjekti nõusolekuta teadusuuringu või riikliku statistika vajadusteks. Muudel juhtudel jälgida isikuandmete töötlemisel vastavaid põhimõtteid (isikuandmete kaitse seaduse § 6, sealhulgas turvalisuse põhimõte) ja isikuandmete töötlemise lubatavust (sh isikuandmete kaitse seaduse § 10 kuni 14, nt andmesubjekti nõusolek või töötlemine avaliku ülesande täitmise käigus).
- Kui andmeaitade jaoks on tihti vaja kasutada kodeerimist ka muul juhul kui isikuandmete töötlemisel andmesubjekti nõusolekuta teadusuuringu või riikliku statistika vajadusteks, kaaluda vastava täienduse tegemist isikuandmete kaitse seadusesse.
- Täpsustada ja muuta mittevasturääkivaks isikuandmete kaitse seaduse § 16, sealhulgas § 16 lõigete 1, 2 ja 3 sõnastused.
- Lihtsamatel juhtudel kaaluda andmeaitade loomisel võimalust läbi ajada olemasolevate tehnoloogiliste vahenditega, näiteks kui kasutatavad andmekogud on samal platvormil.
- Kui lähteandmebaasid on erinevate andmebaasimootorite peale ehitatud või liidestamata, tuleks kaaluda andmete erinevaid struktuure ühendavaid/tõlgendavaid liideseid. Seejuures võib kasutada X-tee lahendusi, spetsiaalset andmeaitade tarkvara, ETL (*Extract, Transform and Load*) tööriistu jm.
- Kuna andmeaida halduseks kasutatava tarkvara platvormi omadustega ollakse valdavalt kas täiesti või osaliselt rahul, siis käesolevas uuringus ei pakuta konkreetse

Andmeaitade uuring

andmeaitade tarkvara tootja eelistust. Uute andmeaitade tarkvara valikul tuleks eelkõige arvestada omadusi, mida küsitlusel hinnati kõige olulisemaks: päringute jõudlust, häid administreerimisvahendeid, tõrketaluvust, integreeruvust olemasolevasse IT keskkonda ja ärianalüüsi süsteemide toetust. Arvestades tendentsi andmebaasidest võetud andmete osatähtsuse vähenemisele ning erinevatest andmeallikatest pärinevate osaliselt mittestruktureeritud andmete suuremale kasutusele, tuleks andmeaitade tarkvara valikul kaaluda ka hetkel vähemtähtsamaks hinnatud andmete kompressiooni, pilvearvutuse toetust ja suurte andmemahutuste toetust.

- Teha valik keskse või virtuaalse andmeaitade tehnoloogiate vahel iga andmeaitade puhul eraldi, sõltuvalt konkreetsest ülesandest, analüüsi vajadustest, lähteandmete kättesaadavusest jne. Soovitame seda teemat käsitleda ka koolitustel.

6 POLIITIKASOOVITUSED

Käesoleva uuringu tulemusena pakume välja järgmised poliitikasoovitused:

- Korrastada seadusandlus vastavalt eelmises jaotises toodud ettepanekutele. Selle tegevuse peaksid läbi viima Majandus- ja Kommunikatsiooniministeerium, AKI, Riigikantselei, Justiitsministeerium, vastavalt vajadusele muud osapooled.
- Tõsta erineva tasemega spetsialistide teadlikkust andmeanalüüsi võimalustest äri- ja otsustusprotsesside toetamisel, andmeida tehnoloogiatest ja nende uutest suundumustest. Sel otstarbel võiks RIA korraldada koolitusi.
- Soodustada süstemaatilist ärianalüüsi võimaluste kasutamist kõigis valdkondades kaasa arvatud riigi juhtimine, poliitika tõhustamine ja kõigi tasandite otsuste kvaliteedi tõstmine. Sellega peaksid tegelema kõik asutused.
- Töötada välja meetmed huvi tekitamiseks andmete avalikustamise vastu. Alustada avaandmestike publitseerimist kõigis riigiasutustes. Selle soodustamiseks võiksid RISO ja RIA korraldada infopäevi, seminare ja koolitusi.

7 LÜHENDID JA MATERJALIDE LOETELU

7.1 LÜHENDID, MÕISTED JA SÕNASELETUSED

1. **Andmeait, andmeladu (kitsam määratlus)** - kindlale valdkonnale (või probleemile) orienteeritud, teisene, integreeritud, ajast sõltuv, püsiv andmekogum, mille eesmärgiks on toetada otsuste tegemist.
2. **Andmeait, andmeladu (laiem määratlus)** - andmete kasutamise meetodite, tehnoloogiate ja praktikate kompleks, mille eesmärk on teha paremaid otsustusi ning pakkuda paremaid teenuseid, säilitades andmesubjektide privaatsuse ning luues võimalused andmete analüüsiks. Sellist andmeaita võib realiseerida väga erinevalt, sealhulgas andmete koondamise, koosvõime, andmete virtualiseerimise, pilvetehnoloogiate, mobiilsete tehnoloogiate, suurte töötlemata andmekogumite kasutamise ja muude vahenditega.

7.2 KASUTATUD MATERJALID

1. Seadused, määrused ja regulatsioonid
 - 1.1. Avaliku teabe seadus
 - 1.2. Isikuandmete kaitse seadus
 - 1.3. Riikliku statistika seadus
 - 1.4. Vabariigi Valitsuse seadus
 - 1.5. Haldusmenetluse seadus
 - 1.6. Riigi Teataja seadus
 - 1.7. Riigi infosüsteemi haldussüsteem. Vabariigi Valitsuse määrus. Vastu võetud 28.02.2008 nr 58.
 - 1.8. Infosüsteemide turvameetmete süsteem. Vabariigi Valitsuse määrus. Vastu võetud 20.12.2007 nr 252.
 - 1.9. Euroopa Parlamendi ja nõukogu määrus üksikisikute kaitse kohta isikuandmete töötlemisel ja selliste andmete vaba liikumise kohta (tekstis "isikuandmete kaitse üldmäärus"), [COM (2012) 11].

Andmeaitade uuring

- 1.10. Euroopa Parlamendi ja nõukogu direktiiv üksikisikute kaitse kohta seoses pädevates asutustes isikuandmete töötlemisega kuritegude tõkestamise, uurimise, avastamise ja nende eest vastutusele võtmise või kriminaalkaristuste täitmisele pööramise eesmärgil ning selliste andmete vaba liikumise kohta (tekstis "isikuandmete kaitse direktiiv"), [COM (2012) 10].
- 1.11. Eesti seisukohad isikuandmete kaitset puudutavate Euroopa Komisjoni algatuste suhtes. Riigikantselei, 29.03.2012 nr 2-5/12-00274-4.
- 1.12. Proposal for an EU Data Protection Regulation - government impact assessment. Ministry of Justice (United Kingdom), 2012. [WWW] <https://consult.justice.gov.uk/digital-communications/data-protection-proposals-cfe> (7.03.2013).

Viidatud kirjanduse loetelu

- Butler 2012. J. Butler, Big Data and Analytics at the IRS, [https://www-950.ibm.com/events/wwc/grp/grp004.nsf/vLookupPDFs/Jeff%20Butler's%20Presentation/\\$file/Jeff%20Butler's%20Presentation.pdf](https://www-950.ibm.com/events/wwc/grp/grp004.nsf/vLookupPDFs/Jeff%20Butler's%20Presentation/$file/Jeff%20Butler's%20Presentation.pdf) (01.03.2013).
- CensusHub 2013. CensusHub projekt, https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Census_Hub (01.03.2013).
- Gartner 2013. Gartner 2013 data warehouse Magic Quadrant, <http://www.zdnet.com/gartner-releases-2013-data-warehouse-magic-quadrant-7000010796/> (21.03.2013).
- Listpoint 2012. Listpoint open data survey, <http://www.publictechnology.net/news/open-datas-potential-remains-closed-public-sector-staff/37572> (08.03.2013).
- McKendrick 2011. J. McKendrick, A New Dimension To Data Warehousing:2011 IOUG Data Warehousing Survey, Unisphere Research, a Division of Information Today, Inc. September 2011, <http://www.oracle.com/us/solutions/datawarehousing/2011-ioug-data-warehousing-survey-522175.pdf> (13.03.2013).
- McKenna 2011. B. McKenna, National Audit Office advocates strategic role for BI in government IT, Computer Weekly, Monday 28 February 2011, <http://www.computerweekly.com/news/2240032829/National-Audit-Office-advocates-strategic-role-for-BI-in-government-IT> (21.03.2013).

Andmeaitade uuring

McKinsey 2011. McKinsey and Company „Big data: The next frontier for innovation, competition, and productivity“ June 2011.

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation retrieved 21 March 2013 (16.03.2013).

PSI-directive 2003. DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information, http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive/psi_directive_en.pdf (16.03.2013).

Russom 2009. P. Russom, Next generation data Warehouse Platforms, TDWI Best Practises report, 2009.