



Euroopa Liit
Euroopa Sotsiaalfond



Eesti tuleviku heaks

Uuringiprojekti „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

Tellijaja: 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 kaTellijapoolne 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	3
2	PROBLEEMI KIRJELDUS JA UURINGU EESMÄRK	3
3	UURINGU METOODIKA	4
4	KOKKUVÕTE	6
5	POLIITIKASOOVITUSED	6
6	LÜHENDID	11

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 rakendamise 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 ministriumid 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, selgitamiseks välja riigis kujunenud andmete ladustamise ja analüüsi hetkeolukord.
2. Uurida välja, millised nii infotehnoloogilised, organisatoorsed, poliitilised, seadusandlikud kui ka 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 nimetatud valdkondades.
4. Uurida andmeaitade võimalikku (semantilist) koosvõimet teiste andmekogudega (ka andmeaitadega), kuna erinevates haldusalades kasutatavaid andmeaitu ei saa vaadelda isoleeritult teistest andmekogudest.

3 UURINGU METOODIKA

Uuringu läbiviimise meetoodika 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 meetoodika kavandatud järgmiselt:

- **Olemasoleva teabe kogumine ja analüüs**, mis põhineb analoogiliste rahvusvaheliste uuringute tulemustel, andmeaitade ja andmeanalüüsi alastel tehnilistel artiklitel ja raamatutel, vastaval seadusandlusel, tooteuuringutel jms.
- **Teabe ja hinnangute kogumine intervjuude ja ankeetküsitluste abil**, selgitamiseks välja andmeaitade kasutamise, loomise, haldamise ja õigusliku reguleerimisega seotud huvigruppide arvamusi, kogemusi, vajadusi, ettepanekuid ja tulevikuvõimalusi.
- **Vaheseminar** arutamaks diskussiooni vormis vastavate huvigruppidega uurimuse vahetulemusi.

Andmeaitade uuring

- **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 meetodika 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 skaalasisid ning vajadusel püstitada täiendavaid hüpoteese. Ankeeteerimine 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) meetodikat, mille järgi määrati küsimustike teemablokid järgnevalt:

1. **Põhivalmisolek** iseloomustab seda, kuivõrd 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 muutunud 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.

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 vahearuanedega.

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 intervjueeritud inimesi. Anketeerimise sihtrühmad olid järgmised: andmeaitade riigisektori spetsialistid, andmeaitade erasektori spetsialistid ja andmete kasutajad.

Sihtrühmade anketeerimiseks koostati eraldi ankeedid 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 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 soovitusel 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 soovitusel peaksid läbi viima Majandus- ja Kommunikatsiooniministeerium, AKI, Riigikantselei, Justiitsministeerium, vastavalt vajadusele muud osapooled.

Allpool esitatud soovitusel 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 andmeaitadesse 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 andmekooseisude 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 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, 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 andmeaitadega). 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 andmeaitade loomise tsüklitega. Seda peaksid tegema eelkõige andmeaitade projekti osapooled, sätestades selgelt andmeaitade projekti tellija suhte andmete algallikate omanikega.
- Tuleks oluliselt laiendada ärianalüüsi rakenduste kasutajate grupe nii organisatsiooni sees kui väljaspool. Näiteks erineva tasemega juhid, analüütikud, statistikud, teadlased jt huvigrupid võiksid andmelao analüüsi kasutada vastava kasutusõigusega või vabalt (sõltub andmetest). Soovitame ärianalüüsi rakendused juba projekti algatamisel planeerida toetamaks erineva tasemega äri- ja otsustusprotsesse ja seega rahuldavamaks 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/andmetike 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 andmelaoanduse, ülisuurte andmemahutude 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 andmeaitade 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.
- Andmeaitade algatamisel tuleks andmeaitade kasutajate ja teiste osapoolte koostöös määratleda ja heaks kiita andmeaitade 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 ajurünnakuid 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 andmekoosseis ja aruanded; analüüsida, kas andmeidas 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 andmeaita peale juba uue andmekogu loomisel; jälgida seadusandluses ettenähtud nõudeid andmeaita asutamisele, registreerimisele, turbele jne; võimalusel kasutada vastavaid standardeid.

- Andmeaitade väljatöötamisel, kasutamisel ja hooldusel tuleks jälgida mingit praktikas järele proovitud andmeaitade arendamise metoodikat; alustada etapiviisiliselt väikese mahuga; soodustada kommunikatsiooni IT poole ja äriinimeste vahel, tutvuda andmeaitade tehnoloogiate pakkujate ülevaadetega; vaadata nii andmeaita tarkvara platvormi 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 andmeidas; 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 andmeaita andmed koosnevad põhifaktide tabeli(te)st ning nendega seotud kontekstiandmeid sisaldavatest "dimensioonidest"; normaliseeritud lähenemine, mille puhul andmed andmeidas 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 andmeaita 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 põhimõtte) 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 andmeaitade halduseks kasutatava tarkvara platvormi omadustega ollakse valdavalt kas täiesti või osaliselt rahul, siis käesolevas uuringus ei pakuta konkreetse 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.

5 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, andmeaitade 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 avaandmetike publitseerimist kõigis riigiasutustes. Selle soodustamiseks võiks RISO ja RIA korraldada infopäevi, seminare ja koolitusi.

6 LÜHENDID

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.
3. **RIHA** - riigi infosüsteemi haldussüsteem.
4. **SWOT** - tugevused (*strengths*), nõrkused (*weaknesses*), võimalused (*opportunities*), ohud (*threats*).