

Põhiandmete olulisuse hindamine

Juhend asutustele

29. veebruar 2024

Lõppversioon



MAJANDUS- JA
KOMMUNIKATSIOONI-
MINISTEERIUM



Rahastanud Euroopa Liit
NextGenerationEU



Eesti
tuleviku heaks



EY

Building a better
working world

Sisukord

Sisukord	1
1. Sissejuhatus	2
1.1 Mõistete ja lühendite sõnastik.....	3
2. Andmete väärtus ja põhiandmete haldamine	4
2.1 Andmete väärtus.....	4
2.2 Avaliku sektori andmete taaskasutamine	5
2.3 Põhiandmed riigi infosüsteemides.....	5
2.4 Põhiandmed asutuse andmevarana.....	5
2.4.1 Põhiandmete liigid	6
2.4.2 Põhiandmete objektid	8
2.4.3 Põhiandmete objektide määratlemine	10
2.4.4 Põhiandmete arhitektuur	11
2.5 Suurbritannia kogemus põhiandmete käsitlemisel.....	12
3. Andmete olulisuse hindamise meetodika	14
3.1 Põhiandmete objektide praktiline määratlemine	14
3.2 Andmete olulisuse hindamise kriteeriumid	15
3.2.1 Äriline kontekst lõppkasutajale.....	15
3.2.2 Äriline kontekst asutusele.....	16
3.2.3 Andmeobjekti praeguse halduse ja kasutatavuse tase.....	16
3.2.4 Andmekogu kvaliteet.....	17
3.3 Andmeobjektide hindamine	17
3.4 Tulemuste analüüs ja visualiseerimine	19
4. Soovitused asutuse põhiandmete haldamise korraldamiseks.....	22
4.1 Põhiandmete haldus	22
4.2 Põhiandmete kvaliteedi haldus.....	22
4.3 Põhiandmete kaardistamine, kirjeldamine ja kirjelduste avalikustamine.....	22
4.4 Põhiandmete ristkasutust võimaldavad andmeteenused.....	23
4.5 Põhiandmete omaniku poolt kasutajatele pakutavad teenustasemed.....	23
4.6 Põhiandmete omaniku kommunikatsioon kasutajatega.....	23
5. Kokkuvõte	24
6. LISAD.....	25
6.1 Lisa 1 – Hindamisvahendi piloteerimise tulemused.....	25

1. Sissejuhatus

Põhiandmete kontseptsioon juured on andmete dubleerivat kogumiste vältimise vajaduses. Avalike ülesannete täitmise käigus riigi andmekogusse kogutud põhiandmeid peab olema võimalik teistel andmekogudel taaskasutada, selle asemel et neid ise uuesti kodaniku või ettevõtte käest koguda [AvTS §436]. Samast põhimõttest oli kantud juba 1990ndate keskel vastu võetud andmekogude seaduse riigi põhiregistrite idee, kuid hoolimata kogu vahepealse perioodi kehtinud põhimõttest ei ole see praktikas veel hästi juurdunud. Algse, peamiselt ökonomika ja efektiivsuse ning kodanike riigi e-teenustega rahulolu vaatenurgast lähtuvate argumentide kõrvale on viimastel aastatel kerkinud uued faktorid, mis väärtustavad suurt taaskasutust leidvaid andmeid, aga ka kriisiolukorras toimetuleku jaoks kriitiliselt hädavajalikke andmeid.

Üha kasvav maht ja sisu rikkus on andmetest teinud võtmetähtsusega ressursi erinevate organisatsioonide tegevuse ja teenuste toetamisel ning ka otsustus- ja juhtimisprotsessides. Andmete haldamine on saanud organisatsioonide igapäevatöök, sest andmete käsitlemine strateegilise varana võimaldab luua uut väärtust nii organisatsioonile endale kui ka ühiskonnale tervikuna.

Andmehalduse praktikas tuleb tihti teha andmetike ja tegevuste vahel valikuid, kuna vahendid ja võimalused on piiratud. Valikute aluseks on arusaam andmete väärtusest ja olulisusest ning nendega seotud riskidest ja piirangutest. Ühe võimalusena andmete eristamiseks nende väärtuse järgi on kasutusele võetud põhiandmete mõiste, mis hõlmab organisatsiooni olulisemaid ja püsivamaid andmeid (vt ka ptk 2).

Põhiandmete määratlemise peamine eesmärk organisatsioonis on luua kindel alus andmete haldamiseks ja tagada, et kõik osakonnad ja süsteemid kasutavad ühtset ja usaldusväärset andmestikku, mis võimaldab tõhusat äriprotsesside ja -teenuste toimimist. Ainult siis, kui põhiandmed on organisatsiooniüleselt määratletud, saab tekkida ka nendest **ülevaade riigiüleselt**, et samad põhimõtted (ühtsed ja usaldusväärsed andmed) säiliks. Tänu ülevaatele riigiüledest andmestikest on võimalik **riigis rakendada andmete *once-only* printsiipi**.

Põhiandmetest tekib kasu asutusele ja riigile vaid siis, kui andmeomanikud selgitavad andmeobjektide kirjeldamise käigus välja äriteenuste üleselt kasutatavad ning tihti dubleeritud andmevarad (näiteks isikute ja asukohtade andmed). Oluline on, et igal andmestikul oleks olemas eraldi andmeomanik, kes koordineerib ja kindlustab andmete kvaliteeti ning tagab, et kõik kasutaksid õiget ja täpset informatsiooni (andmeid).

Kuigi digiühiskonna arengukava ei käsitle eraldi põhiandmeid, on seal probleemina välja toodud andmete ühekordse küsimise põhimõtte läbiv rakendamine, mis tuleneb peaaesjalikult andmete kirjelduse ja kvaliteedi ebapiisavast tasemest. Lisaks on asutuste arvates veel teisigi andmete taaskasutusega seotud probleeme. Üheks selliseks on mitteühilduvad andmete kogumise meetodid, standardid, vormingud, klassifikaatorid ja loendid. See tähendab, et esineb olukordi, kus andmed on olemas, aga praktikas ei ole neid võimalik taaskasutada. Samuti on probleemiks põhiandmete taaskasutamiseks ette nähtud liideste käideldavus, SLA puudumine või selles ette nähtud tingimustest kinni pidamine (nt teenustes tehtavatest uuendustest teavitamine).

Majandus- ja Kommunikatsiooniministerium tellis 2024. aasta jaanuaris küsitluse teemal „Avaandmed ja andmete ühekordse küsimise põhimõte“. Uuringu käigus küsitleti 1010 üle 15-aastaseid Eesti elanikke.

Ligikaudu pool (~50%) uuringus osalejatest teab avaandmete mõistet ning selle tähendust. Vastavalt uuringule peab 71% elanikkonnast avaandmeid ühiskonnale oluliseks, sest nende põhjal on võimalik vastu võtta otsuseid või saada statistilisi andmeid. Lisaks toodi välja, et andmete kättesaadavus on läbipaistvuse ja demokraatia tunnus ning on oluline, et riigi käes olev informatsioon oleks avalikult kättesaadav.

Andmete ühekordse põhimõttega on kursis ligikaudu 41% uuringus osalejatest. Umbes 51% küsitletutest peab aga seda teemat väga oluliseks ning arvab, et see on muutnud riigi teenuste kasutamise lihtsamaks. Umbes 67% vastanutest toetab avaliku sektori omavahelist andmevahetust. Küll aga toodi välja mitmeid olukordi, kus andmeid on küsitud nende käest mitmekordselt. Põhiliselt toodi välja sellised olukorrad tervishoiuteenustes, äriregistris ja maksuametis tuludeklaratsiooni esitamisel. Lisaks on mitmekordselt küsitud andmeid ka aadressi, elukoha ja kontaktide kohta.

Siinne juhised valmis MKM-i tellitud projekti „Põhiandmete olulisuse hindamise metodoloogia ja nõuete koostamise analüüs“ raames, mille eesmärgiks oli välja töötada üldised põhimõtted või kriteeriumid põhiandmete olulisuse hindamiseks ning koostada praktiline metoodika, mille alusel on asutustel võimalik:

- ▶ määratleda oma põhiandmete objektid,
- ▶ neid analüüsida ja hinnata, ning
- ▶ selle käigus eristada nende hulgast olulisemad.

Rakendatav metoodika võimaldab saada ka riigiülese ülevaate põhiandmetest.

Aruanne annab ülevaate teoreetilisest taustast ehk andmete väärtuse hindamisest, põhiandmete määratlemisest ja haldamisest ning Suurbritannia kogemusest (ptk 2), praktilisest juhiseist ehk põhiandmete olulisuse hindamise metoodikast ja selle piloteerimise tulemustest (ptk 3) ning oluliste põhiandmete haldamise soovitud asutustele (ptk 4).

Lisaks aruandele valmisid projekti käigus järgmised tulemid:

- õigusanalüüs, mis keskendub õiguslikele probleemidele ja lahendustele seoses kehtiva põhiandmeid reguleeriva õigusega;
- põhiandmete olulisuse hindamise töövahend, mida piloteeriti asutuste peal;
- soovitused MKM-ile, kuhu on kokku koondatud projekti käigus kogutud informatsioon ning välja pakutud soovitused järgnevateks põhiandmetega seotud tegevusteks;
- kokkuvõttev koondaruanne, mis võtab kokku projekti tegevused ja tulemid.

Aruande koostas MKM-i tellimusel Ernst & Young Baltic AS Eesti meeskond koostöös projekti kaasatud asutuste esindajatega. Aruanne valmis perioodil november 2023 kuni märts 2024.

1.1 Mõistete ja lühendite sõnastik

Tabel 1 annab ülevaate aruandes kasutatud peamistest mõistetest ja lühenditest.

Tabel 1. Mõistete sõnastik

Mõiste	Selgitus
AvTS	Avaliku teabe seadus
MKM	Majandus- ja Kommunikatsiooniministeerium
RIHA	Riigi infosüsteemi haldussüsteem
RIA	Riigi infosüsteemi amet
SLA	Teenusetasemelepe (ingl <i>Service-Level Agreement</i>)

2. Andmete väärtus ja põhiandmete haldamine

2.1 Andmete väärtus

Andmete väärtus väljendub nende kasutamises ja avaldub näiteks paranenud juhtimis- ning otsustusprotsessides. Lisaks avaldub andmete väärtus mõjuna organisatsiooni eesmärkidele ja informatsioonina lõppkasutajale, kus see on küll vaid tajutav väärtus ning kattub kasutatavuse mõistega.

Kasutatavus hõlmab tõhusust (ressursside kasutamise mõju), efektiivsust (soovitud eesmärkide saavutamist) ja rahulolu. Tõhusust ja efektiivsust saab kvantitatiivselt mõõta ja seostada finantsväärtusega, rahulolu on kvalitatiivne.

Tõhusus näitab tegelikku mõju, mida kasutajad andmete abil saavutavad. Kui kasutada olev informatsioon viib otsusest tuleneva (äriilise) väärtuse tekkimiseni, siis on see tõhus. Informatsiooni ja äriilise väärtuse vaheline kaudne suhe teebki andmete ja informatsiooni väärtuse määramise raskeks. Siiski, kui otsustaja sõltub täielikult informatsioonist, võib otsuse ja andmete finantsväärtust lugeda võrdseks, jättes andmete loomise, kasutamise ja hoolduskulud kõrvale.

Efektiivsus keskendub peamiselt aja ja pingutuste minimeerimisele ning seda saab mõõta süsteemis toimivate tegevuste kaudu, näiteks mõõtes päringute kestust ja arvu andmeteenuste kasutamisel. Samas ei pruugi päringute arv ise olla kasutajate jaoks oluline, neile on olulisem tulemus, mida nad otsivad.

Iga organisatsioon omab teatud andmeid, mida on vaja kasutada korraga paljudes valdkondades, protsessides ja süsteemides. Organisatsioon ise ja tema kliendid saavad korraga tarbida lisanduvat väärtust, kui need andmed on koondatud ja kasutamiseks jagatud: kõik kasutavad samu nt kontaktandmeid, loendeid, aadresse, kulu- ja maksukoodide jne. Kasutajad eeldavad, et lahknevate andmete korral kehtib organisatsioonis teatav järjepidevus, kuniks kvaliteediprobleemid saavad lahendatud. Seda protsessi nimetatakse põhiandmete haldamiseks.

Infosüsteeme ja andmestikke arendatakse aga orgaanilisemalt kui kasutajad eeldaksid. Suurtes organisatsioonides põhjustavad erinevad projektid ja algatused, ühinemised ja omandamised ning muud paralleelsed tegevused mitme sarnase, kuid isoleeritud süsteemi kasutamist. See toob paratamatult kaasa ebakõlad eri süsteemide andmestruktuurides ja -väärtustes. Hajusus suurendab kulusid ja riske.

Põhiandmete väärtus seisneb nende olulisuses organisatsiooni- ja riigiüleste otsuste tegemisel. Peamised aspektid, mis määravad põhiandmete väärtuse:

- allika usaldusväärsus ja tõendusväärtus: põhiandmed moodustavad organisatsiooni alustalad, pakkudes ühtset ja usaldusväärset teavet protsesside ja tarbijate osas;
- täpsemad otsused: hallatud põhiandmed võimaldavad paremaid ja teadmispõhiseid otsuseid, mis toetavad eesmärkide saavutamist;
- efektiivsemad protsessid: kvaliteetsed põhiandmed võimaldavad andmete tõhusamat kasutamist ja töötlemist, aidates seeläbi optimeerida protsesse ning vähendada kulusid ja riske;
- kliendirahulolu: täpsed ja täielikud põhiandmed võimaldavad paremat klienditeenindust ja kohandatud lahendusi, suurendades seeläbi teenuste tarbijate rahulolu;
- riskide maandamine: kvaliteetsed põhiandmed aitavad vähendada riske, nagu andmete dubleerimine, ebatäpsus või ebakvaliteetsus, mis võivad kaasa tuua probleeme või õiguslikke tagajärgi.

Põhiandmete haldamise ja jagamise eesmärgiks on luua organisatsiooni andmetest liigiti ühesed ja tõepärased (üleriiklikud) allikad. Kokkuvõttes **võimaldavad põhiandmed organisatsioonil ja riigil paremini mõista asutuste teenuseid ja nende tarbijaid** ning teha teadlikumaid otsuseid, mis viivad kõigi jaoks parema tulemuseni.

Projekti eesmärgiks oli luua mudel, mille abil saaks asutus hinnata põhiandmete olulisust, mis omakorda on vajalik otsuste toetamiseks. Andmevara väärtust teades saab paremini hinnata, kas seda põhiandmete objekti peaks (riiklikus vaates) looma või mitte, st kasutada juba teises asutuses loodud andmevara.

2.2 Avaliku sektori andmete taaskasutamine

Andmete taaskasutamine võimaldab ressursitõhusamat lähenemist andmete majandamises, võimaldades ühekordsuse põhimõttest lähtuvalt tugineda juba olemasolevatele andmetele. Seadus keelab asutada samade andmete kogumiseks eraldi andmekogusid¹, ehk ühed ja samad andmed võivad küll esineda mitmes andmekogus, kuid vaid ühes on nad põhiaandmed. Sellisel juhul teised andmekogud üksnes kasutavad põhiaandmeid, mitte ei kogu neid algallikast uuesti.

Avaliku sektori andmekogude koosseisu analüüs selgitab, millised vajalikud andmed on asutuses juba olemas ning millised on võimalik kätte saada riigi infosüsteemist või teiselt asutuselt.² Et võimendada andmete taaskasutamist, peavad avaliku sektori andmekogude juurdepääsupiiranguta andmed olema tervikuna allalaetavad masinloetaval kujul.

2.3 Põhiaandmed riigi infosüsteemides

Selleks, et riigi infosüsteem saaks efektiivselt toimida ja eesmärgi täita, on vaja ühtseid põhimõtteid, millest riigi infosüsteemi haldamine peab lähtuma. Ühekordsuse põhimõtte tagamiseks peab enne andmekogu asutamist või andmekogus kogutavate andmete koosseisu muutmist andmekogu asutaja või vastutav töötaja kontrollima, kas vajalikke andmeid juba kogutakse mõnes olemasolevas andmekogus. Kui uue andmekogu loomine või olemasolevas kogutavate andmete koosseisu muutmise on andmekvaliteedi tagamise seisukohalt põhjendatud, koostab andmekogu asutaja või vastutav töötaja andmekogu dokumentatsiooni.³

2.4 Põhiaandmed asutuse andmevarana

Asutuste andmevarade kirjeldamise käigus selgub tihti, et paljud mõisted ja nendega seonduvad andmed korduvad mitmes valdkonnas ja sisuteenuses ja on tihti ka dubleeritud, näiteks isikud, kontaktid, teenuslepingud, asukohad jne. Sarnased andmed oleks asutuses mõistlik koondada ühtseteks põhiaandmete objektideks ja luua võimalus nende ühiseks kasutamiseks. Põhiaandmed peavad kirjeldama organisatsiooni andmestike usaldusväärsemat ja täpsemat osa. Põhiaandmete ulatus, vorm ning lubatavad vahemikud määratakse ärireeglite ja rakenduste loogikaga. Asutuse andmekogude kirjeldamisel RIHA-s nõutakse ka põhiaandmeid kirjeldavate andmeväljade märkimist, kuid korrektseks määramiseks puudub nõuete metoodika, mistõttu ei saa seda korrektselt teostada ega hiljem usaldada.

Eri andmestike kasutamisel võib tekkida olukordi, kus asutusel on korraka mitu erinevat vaadet olukorrast ning on vaja otsustada, milline on õigeim. Selleks on vaja teada, kust need andmed pärinevad või kuidas neile juurde pääseb ning millised andmed on vaid kindla kasutusotstarbega. Põhiaandmete tarbeks tuleb määrata usaldusväärsed allikusteemid, kus andmed luuakse/kogutakse ja/või hoitakse vaid kindlate reeglite ja nõuete alusel. Põhiaandmetele määratakse asutuses eraldi omanikud, kes koordineerivad ja kindlustavad andmete kvaliteeti, samal ajal tagades, et kõik kasutavad sama, õiget ja täpset informatsiooni. Põhiaandmete alamliigiks on ka viiteandmed.

Põhiaandmete haldus hõlmab kirjeldatud andmeobjektide väärtuste ja tunnuste kontrolli, mis võimaldaks rakendustel infosüsteemide ja ka teiste asutuste üleselt (X-tee) täpseimate ja ajakohasemate andmete kasutamist. Peamiseks eesmärgiks on siin täpsete ja ajakohaste väärtuste kättesaadavuse tagamine, samal ajal vähendades mitmeti tõlgendatavate tunnuste riske.

Põhiaandmeid ei hallata üldjuhul kindlas rakenduses, vaid inimeste, protsesside ja tehnoloogia kombinatsioonis, mille käigus asutuse sisuteenused ja IT töötavad koos, tagamaks põhiaandmetee ühtsuse, täpsuse, semantilise järjepidevuse ja vastutajad.

¹ AvTS § 43³ lg 2

² Riigi Infosüsteemi Amet „Meespea andmekogude asutamise ning andmekogude põhimääruste koostamise ja kooskõlastamise kohta“.

³ RIHA määrus § 6

Asutuse põhiantmete loomisel on olulised järgmised aspektid:

- Millistele rollidele, teenustele, organisatsioonidele, asukohtadele, asjadele jne viidatakse korduvalt?
- Milliseid andmeid kasutatakse näiteks isikute, organisatsioonide, asukohtade ja asjade kirjeldamiseks?
- Kuidas on andmed kirjeldatud ja struktureeritud, sealhulgas andmete granulaarsus?
- Kus andmeid luuakse, hoitakse ning kättesaadavaks muudetakse?
- Kuidas andmed või formaadid muutuvad, kui need liiguvad läbi asutuse infosüsteemide?
- Kes neid andmeid kasutab ja milleks?
- Milliseid kriteeriume kasutatakse andmete ja nende allikate kvaliteedi ja usaldusväärsuse mõistmiseks?

Põhiantmete loomine ja haldamine illustreerib hästi põhilist andmetega seotud probleemi: inimesed valivad sarnaste mõistete esitamiseks väga erinevaid viise ja nende esitusviiside kooskõlastamine ei ole lihtne. Oluline on ka see, et teadmised andmetest/mõistetest muutuvad aja jooksul ja nende muutustega arvestamine nõuab pidevat energiat, planeerimist, andmete tundmist ja tehnilisi oskusi.

Iga asutus, milles on vajadus põhiantmete järele, omab juba keerulisi/aegunud infosüsteeme, kus on realiseerunud erinevad viisid reaalse maailma objektide kirjeldamiseks. Ajas toimuva orgaanilise kasvu või ühinemiste tulemusena võivad põhiantmetele sisendit andvad süsteemid omada erinevaid mõistete ning andmekvaliteedi standardite määratlusi.

Peamised sammud põhiantmete loomisel:

- Valida lähtesüsteemid, mis annavad kindlale põhiantmete elemendile peamise sisendi.
- Leppida kokku reeglid erinevate instantside andmestike täpseks sobitamiseks ja ühendamiseks.
- Leppida kokku viis mittedobivalt ühendatud andmete tuvastamiseks ja taastamiseks.
- Leppida kokku viis usaldusväärsete andmete levitamiseks kõikidesse asutuse infosüsteemidesse.

Tegevused põhiantmete elutsükli haldamisel:

- Põhiantmetele konteksti määramine, sealhulgas seonduvate atribuutide ja kasutamistingimuste kirjeldamine.
- Andmeallikates dubleeritud olemite tuvastamine, tunnuste ja viidete loomine ning haldamine andmete hilisemaks integreerimiseks.
- Erinevate allikate vahel andmete ühildamine ja koondamine, et tekitada põhiantmetest „tõesem“ versioon. Konsolideeritud kirjed pakuvad ühendatud vaadet infosüsteemidest ja aitavad lahendada atribuutide nimetamise ja andmevälja väärtuste vastuolusid.
- Valesti sobitatud või liidetud andmete ja mõistete tuvastamine, probleemide lahendamine ja õigete tunnuste abil sobivate seoste loomine.
- Rakenduste ja infosüsteemide ligipääsu tagamine põhiantmevaradele otsepöördumise, andmeteenuste või analüütilise andmeida kaudu.
- Põhiantmete kasutamise nõude jõustamine asutuse sees. See protsess nõuab tugevat muudatuste juhtimist.

2.4.1 Põhiantmete liigid

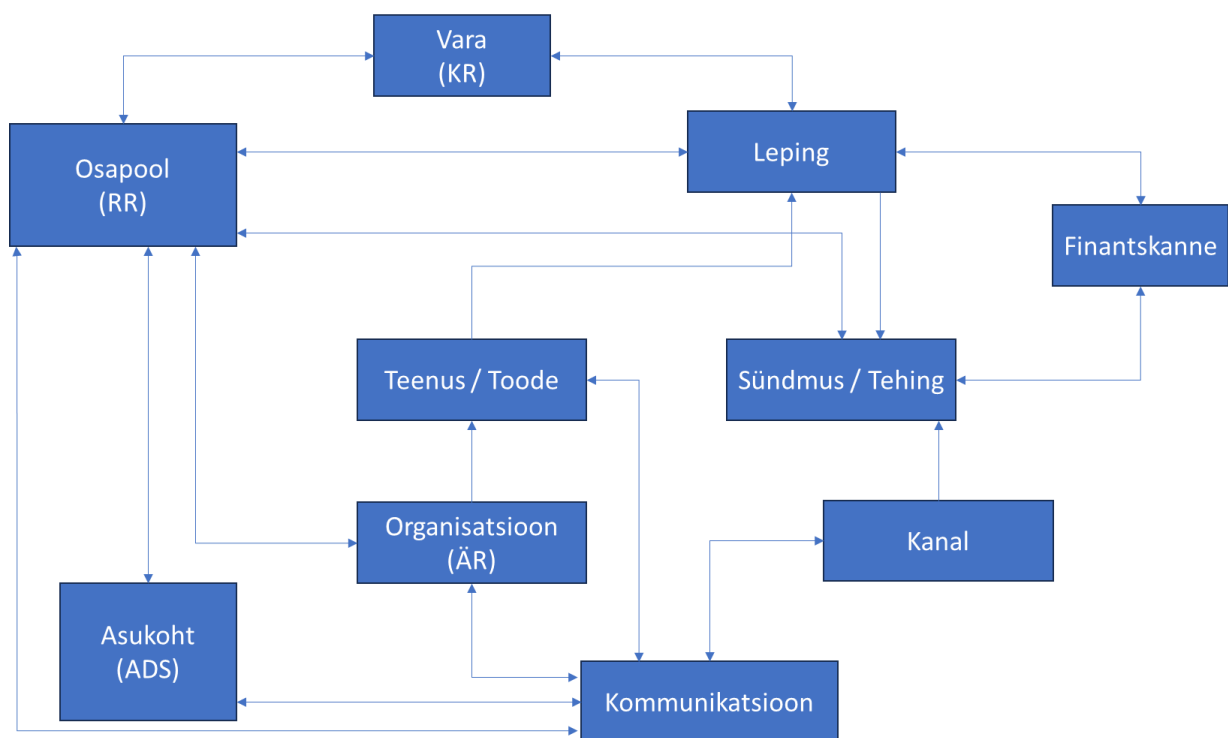
Põhiantmete haldamine eeldab asutuselt järjepidevat ja loogilist andmete liigitamist (grupeerimist) ja kirjeldamist. Kuigi kasutatavad terminid ja andmekirjeldused võivad ühe süsteemi piires olla mõistlikud, ei pruugi need olla kogu asutuse või riigi tasandil enam asjakohased. Seetõttu peaksid asutuse tasemel kasutatavad põhiantmed, terminid ja kirjeldused olema kooskõlas riikliku kontekstiga ning ei tohiks sõltuda ühestki konkreetse asutuse infosüsteemist.

Põhiandmete liigid (ka dimensioonid) on andmehalduse mõiste, mis viitab organisatsiooni peamistele ja püsivatele andmetele, mis kirjeldavad organisatsiooni tegevuste ja teenuste objekte. Liigitamine aitab luua süstematiseeritud ja ühtset vaadet andmevaradele, võimaldades tagada paremat andmekvaliteeti ja ühtset lähenemist erinevates andmeteenustes.

Infosüsteemides kasutatakse parema käideldavuse saavutamiseks relatsioonilisi andmemudeleid, kus kesksel kohal on andmetabel, milles paiknevad andmed ridades ja veergudes ja nendevaheliseks seoseks on n-ö võtmeväljad. Andmeliigipõhises mudelis on kesksel kohal mõõdikuid sisaldavad ärilised faktid (üks keskne tabel), mille ümber on erinevat liiki andmete ehk tunnuste tabelleid. Andmeliikidepõhine mudel on mittetehnolisele kasutajale lihtsam ja intuiitsem, sest vastab juba kindlatele äriküsimustele ning muudab seeläbi andmete mõistmise, omanduse võtmise ja kasutamise lihtsamaks.

Andmeliikidepõhised mudelid on loodud äriliseks aruandluseks ja analüütikaks, aga seda tüüpi mudel on valitud ka põhiandmete kirjeldamiseks, sest taoline mudel koondab erinevate tasandite (äri)protsessidega seonduvad (äri)küsimused ühtseks tervikuks. Seoseid ja küsimusi tunnustega piirates/laiendades saab liikuda üksiku infosüsteemi tasemelt asutuse ja riigi tasandile nii, et äriküsimus säilib. Selliste püsivate seoste järjepidevus ongi projektis põhiandmete objektide loomisel eelduseks ning on seetõttu hiljem üheselt mõistev kõikidel tasanditel.

Joonis 1 kujutab peamiste põhiandmete liikide ja seoste mudelit, mis erinevatel üldistusastmetel sobib nii infosüsteemi, asutuse, kui ka riikliku vaate aluseks. Andmeliikidepõhine mudel on vajalik, kuna see visualiseerib hästi ärilisi seoseid ja võimaldab nii organisatsioonil ning hiljem ka riigil oma andmevarasid mõista.



Joonis 1. Põhiandmete liigid

Peamised põhiandmete liigid on erinevas detailsusastmes olulised nii infosüsteemi, asutuse ja riigi tasandil:

(1) Osapool: asutuse või organisatsiooniga seotud füüsiline või juriidiline isik, kes kannab erinevas kontekstis erinevat rolli (kliendid, partnerid, hankijad, tarnijad, laopidaja, võlglane, esindaja jne). Neid põhiandmeid hoiavad rahvastiku- ja äriregister, aga enamikus asutustes on isik kopeeritud/replikeeritud ka ühte või mitmesse infosüsteemi.

(2) Vara: osapoolega seotud materiaalne ja immateriaalne vara. Näiteks kinnistud (kinnistusraamat), vallasvara (ARK), raha, kaubad, ehitised, litsentsid jne.

(3) Asukoht: osapoolte ja varaga seonduvad aadressid (ADS), piirkonnad, riigid, postiindeksid, kontaktandmed, katastriandmed, koordinaadid jne.

(4) Toode/Teenus: põhiaandmed hõlmavad asutuse sisuteenuste nimetusi, kategooriaid, regioone, subjekte, parameetreid, tarbijaid, kasutamist, piiranguid, tuluseid, kulusid jne. Teenuse põhiaandmelemendid on aluseks sündmuspõhise teenuse tekkimiseks riigis.

(5) Leping: kokkolepped osapoolte ja nende huvide vahel (teenus/toode ja vara), püsivad asutuse sees ja on kättesaadavad vaid järelevalveasutustele.

(6) Sündmus: füüsiliselt registreeritav huvide kokkupuude, mis võib, aga ei pea kaasama osapoolt. Iga sündmus peaks saama registreeritud tehinguna (logina). Väga tihti on asutustes aga märgata, et sündmus on hoopis mingi dokumendi olekumuutus. See on vana ja otseselt paberdokumentidest üle tulnud mõtteviis ja pärsib täna oluliselt põhiaandmete ühtlustamist asutuste vahel ja riigis.

(7) Organisatsioon: osapoolte täiendav vaade, mis kirjeldab põhiaandmete vastutava töötaja organisatsiooni hierarhia, ametnikud, teenistujad, nende volitused ja vastutused, kasutajaõigused jne.

(8) Kanal: digitaalsed rakendused, füüsilise kontorid, mille kaudu osapool saab suhelda ameti/asutusega ja tarbida teenuseid/tooteid.

(9) Kommunikatsioon: viis, kuidas asutus osapooltega suhtleb kindlas asukohas kindla kanali kaudu, et pakkuda või nõuda toote/teenuse tarbimist.

(10) Finantsandmed: eelarve kulud, tulud, arved, maksetingimused, limiidid, arveldusperioodid jne.

Andmeliikidepõhiseid mudeleid on hiljem, näiteks erinevate asutuste vahel, sarnaste liikide järgi lihtsam sobitada ja kombineerida. See võimaldab riigi tasandi põhiaandmete kasutajatel teostada keerukamaid analüüse ja avastada seoseid erinevate põhiaandmeobjektide vahel.

Asutus või organisatsioon võib piirduda vaid oma põhiaandmete koondamise ja korrastamisega (näiteks ministriumid) ilma kindlaid objekte määratlemata. Rahvastikuregister või äriregister on näiteks suur relatsiooniline andmekogu (infosüsteem), mis tsentraliseerib paljude asutuste tarbeks osapoolte dimensiooni. Äriküsimused ehk faktid kooruvad pigem välja hoopis seda dimensiooni kasutavate asutuste põhiaandmete objektidest. Suured registrid aga saavad jälgida, et erinevate põhiaandmeobjektide vahel oleks osapoolte kvaliteet korrektselt lahendatud, andmeid ei dubleeritaks, poleks andmete kadu ning uuendamata ühikuid.

2.4.2 Põhiaandmete objektid

Andmeobjekt on erinevat tüüpi andmete kogum, mis on korraldatud teatud struktuuri või vormingu järgi. Objekt koosneb andmeväljadest (ka elemendid või atribuudid), mida saab andmebaasi koguda, salvestada ja töödelda. Infosüsteemide kontekstis mõistame andmeobjekti kirjete või andmestruktuuride väljadena. Andmeobjektidel on kindlad parameetrid (omadused, meetodid, vaikesätted) ja sellele on määratud omanik. Tihti on see seonduva äriteenuse juht või infosüsteemi/registri peakasutaja.

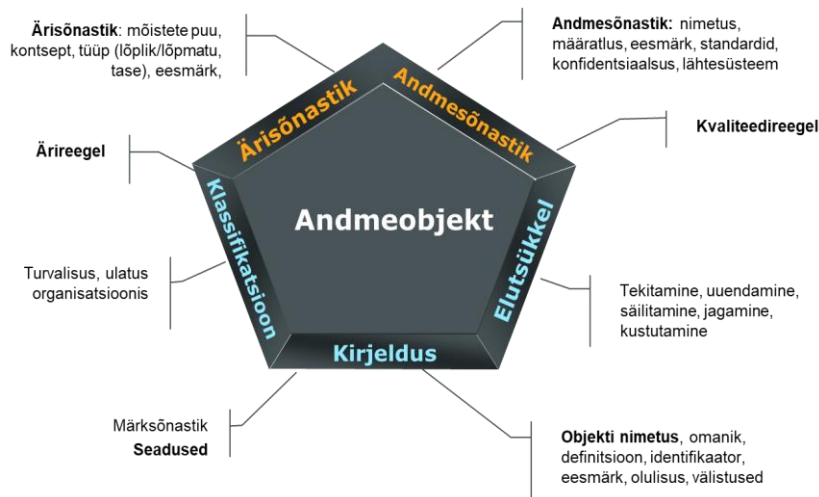
Põhiaandmete objekt on määratletud andmekogu, mis on seotud ühte kindlat tähendust ja kompositsiooni määrava viitega [ISO/IEC 18013-2]. Põhiaandmetest moodustatud objekt on laiapõhjalisem, ameti põhimäärusest tulevaid kohustusi kirjeldav ning see viitab andmetele kui varale või ressursile, millel on äri- või organisatsiooniline väärtus. Andmevara võib hõlmata erinevaid andmetüpe ja nende haldamine on oluline osa asutuse strateegiast.

Põhiaandmete objekti saab kasutada sisemiste ja välimiste kasutajate dokumenteeritud vajaduste järgi, mis suunab objektis sisalduvate andmete edasist kogumist, töötlemist, säilitamist ja kasutamist. Andmeobjekti dokumenteeritud nõuete hulka nimetatakse andmeteenuks ja see on ajas püsiv ning hallatud (säilitab kvaliteedi) ning kindlas kanalis, kindla andmekoosseisu (lähteandmed ja reeglid) pakkumine. Kõige lihtsam andmeteenus on ühekordne raport. Põhiaandmete objekti kasutajal on põhiliselt kolme tüüpi vajadusi:

- vajab teatud vajaduste ja tingimuste alusel koondatud andmehulka, et töödelda seda ise edasi või esitada see mõnele teisele tarbijale sisendiks;
- vajab andmete seoste ja ulatuse alusel vastust äriküsimusele ja tihti vastuse esitamist aegreana mitmes andmeliigis;
- objektis on andmeid puudu või need on madala kvaliteediga, edastatakse asutuse lähtesüsteemidele arendusprojekti tellimusena.

Parameetrite kasutamine võimaldab andmeobjekti üldistatud ja korduvkasutatavat loogikat ning toimimist erinevate sisendiks kirjeldatud andmete korral. Lisaks on parameetrid olulised andmete vahetamisel objektide vahel ja erinevate sisendite manipuleerimisel (andmete muutmisel muu eesmärgi nimel). Joonisel 2 on esitatud andmeobjekt ja parameetrid.

- **Andmeobjekti nimetus:** dokumentatsioon ja kirjeldus, mis selgitab, millist rolli ja eesmärgi andmeobjekt täidab andmeid koguva ja töötleva organisatsiooni protsessides. See sisaldab üldist ülevaadet sellest, miks andmeobjekt on oluline, millist teavet see sisaldab, kuidas seda kasutatakse ja millised on selle mõjud asutuse tegevusele. Kirjeldus tagab laiapõhjalisema arusaadavuse sisemistele ja välimistele kasutajatele ja annab ka üleriikliku vaate.
 - Andmeobjekti unikaalne nimi ja kood, mis identifitseerib andmeobjekti.
 - Andmeobjekti omanik asutuses.
 - Andmeobjekti eesmärk: miks andmeobjekt on loodud ja millist ärilist vajadust see rahuldab.
 - Välistused ja piirangud: millistel tingimustel ei saa või ei tohi objekti kasutada.
 - Andmeobjekti struktuur: millist teavet andmeobjekt sisaldab ja kuidas see on struktureeritud.
 - Andmeobjekti kasutamine: kuidas andmeobjekti asutuse protsessides või süsteemides kasutatakse ning millised on selle eeldatavad tulemused.
 - Andmeobjekti suhted teiste andmeobjektidega: kuidas see andmeobjekt on seotud teiste andmeobjektidega ja kuidas see aitab kaasa asutuse andmete terviklikkusele ja tõhusale kasutamisele.
 - Andmeobjekti haldus: kuidas andmeobjekti muudetakse, hoitakse ja ajakohastatakse, et tagada vastavus asutuse ja väliste tarbijate arenevatele vajadustele ja muudatustele.



Joonis 2. Andmeobjekt ja tema parameetrid

- **Ärisõnastik:** asutuse metaandmestik, mis sisaldab organisatsioonis või organisatsioonide vahel kokku lepitud sõnavara ja kogumina kõiki organisatsiooni tegevustes kasutatavaid mõisteid koos sisu selgituse, andmeobjektide loendi, atribuutide loendi ning neid mõisteid kasutavate osapooltega. Mõisted-terminid

võivad pärineda nii valdkonna sõnastikest kui ka üldisest märksõnastikust, aitab tagada asutustes ühtse arusaamise ja kommunikatsiooni valdkonnas kasutatavate terminite tähenduse üle.

- Andmeobjekti koosseisu kuuluvate terminite valik.
- Valitud terminite kirjeldus, tüüp, andmeomanik.
- Valitud terminite hierarhiline seos asutuses.
- **Märksõnastik:** organisatsiooniväline mõistete või terminite korrastatud loetelu, tuleneb kas seadustest või kindla valdkonna spetsiifikast, kus kirjeldatakse valdkonna terminite ja nende seoste parimat praktikat. Valdkonda hõlmavaid teadmisi ja mõisteid piiritlevad ning nende kokkuleppelisi seletusi esitavad valdkonna märksõnastikud.
- **Andmesõnastik:** asutuse andmestruktuuride sisukirjelduste ja kirjelduste käigus tekkivate terminite loetelu, mis on tähistatud sisu väljendava suhtluskeelse sõna, fraasi, akronüümi, lühendi või muu masinloetava tähisega. Andmesõnastik asub eraldi andmebaasis ning sisaldab lisaks andmete kasutamiseks vajalikke metaandmeid (nt andmetüüp, väljade pikkus, täitmise kohustuslikkus, juurdepääsupiirangud jmt). Andmesõnastiku sisu saadakse andmebaasi füüsilise mudeli ekspordist ja seotakse manuaalselt sõnastiku terminitega (Sõnastike koostamine andmekirjeldustes. Praktiline juhise. Versioon 0.3. (Statistikaamet, Märts 2023)).
 - Objektis sisalduvate andmete kirjeldus: nimetus, kirjeldus, tüüp, standard, tundlikkus, omanik, lähtesüsteem jne.
- **Ärireegel:** andmeobjekti arendusnõue, mis määratleb või piirab äriprotsessi mingit aspekti. Nende kaudu saab kontrollida või mõjutada asutuse pakutavaid teenuseid. Ärireegleid saab teisendada mõõdetavaks andmekvaliteedi reegliks, mille alusel andmekvaliteedist raporteerimine käivitab omakorda erinevaid sisu- ja IT arendusprojekte. Ärireeglid on defineeritud mõistena, mõistete vahelise seosena, mingi tegevuse piirangu või käivitajana, mingi tegevuse tulemina.
 - Andmeobjektiga seonduvate ärireeglite nimetus, kirjeldus, sisu, omanik.
- **Kvaliteedireegel:** ärireeglitest tuletatud tehnilised skriptid või SQL kujul laused, mida on võimalik paigutada lähtesüsteemi või andmebaasi koodi. Reeglid käivituvad lähtesüsteemis automaatselt ja väljastavad andmekvaliteedi raportisse veateated. Andmekvaliteedi reeglite skriptid ei ole tavakasutajale loetavad.
 - Andmeobjektiga seonduvate kvaliteedireeglite nimetus, kirjeldus, sisu, omanik.
- **Elutsükli kirjeldus:** põhiantmete objekti elutsükkel on objekti loomine, säilitamine, kasutamine, -jagamine, kustutamine.
- **Turvalisuse ja käideldavuse nõuded**

Hetkel ei ole teada, kuidas ja millises hoidlas hakata asutuste kirjeldatud põhiantmete objekte hoidma, kuid tõenäoliselt tänasele RIHA-le lähedases keskkonnas. Arhitektuursed kontseptsioonid on kirjeldatud peatükis 2.3.4.

2.4.3 Põhiantmete objektide määratlemine

Määratlemist on mugav alustada ja jätkata asutustel, kus on juba olemas andmehalduse kord, määratud rollid, töökorraldus ja ressurss ning kasutusel tehniline ärisõnastiku- ning andmekataloogitarkvara.

Andmeväljad nimetatakse andmekataloogis arusaadava eestikeelse terminina. Väljade nimetamisel tuleb järgida ühtset stiili ja standardit. Väljade kasutusotstarbe mõistmiseks ning tulevikus andmekvaliteedi parendamiseks lisab andmehaldur andmeväljale juurde ka kokkulepituid staatused. Andmekataloogis säilitatakse ka andmeväljade originaalnimetused, et hiljem oleks võimalik andmekataloogi uuendada. Andmekataloogi loovad vastutavad andmehaldurid, kes koguvad protsessi käigus andmekvaliteedi nõudeid ning toetavad andmeomanikku mõistete ja seoste loomisel.

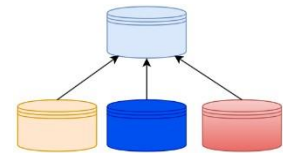
Ärisõnastik võib olla tesauruse struktuuriga märksõnastik, mis võimaldab näidata andmete ja terminite seoseid. Ärisõnastik võib olla ka vaadepõhine, kus andmetega saab siduda lisaks terminitele veel näiteks andmeelemente, teenuseid, rakendusi jne. Ärisõnastiku kaudu saavad andmestikes sisalduvad andmed endale konteksti ja andmeelemendid tähenduse. Ärisõnastiku mõistete kaudu saab andmed ja nende kasutamise siduda asutuse tööprotsesside ja teenustega ning kehtestada andmetele sisulist väärtust loovad kvaliteedireeglid.

2.4.4 Põhiandmete arhitektuur

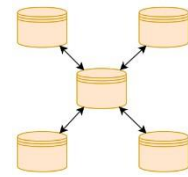
Andmearhitektuur keskendub andmevara strateegiliste nõuete ja integratsioonide detailidele. Andmevara tuleb kirjeldada asutuse erinevatel abstraktsioonitasemetel, luua andmeobjektide disain, määrata standardid, mis reguleerivad andmete kogumist, säilitamist, korrastamist, kasutamist ja eemaldamist.

Põhiandmete tehnilise lahenduse võib jagada arhitektuurselt nelja gruppi. Lähtesüsteemi alusarhitektuuri teave on põhiandmeobjekti kasutamisel oluline näiteks selleks, et kokku ei satuks erineval hetkel uuendatavad andmed, kui mõni andmekogu uueneb vaid kindlal kuupäeval, teine register aga jooksvalt.

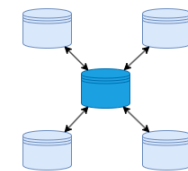
- **Konsolideeriv:** põhiandmed kogutakse kokku kesksesse andmebaasi, kuid neid ei avaldata ega kasutata enam uuesti kanalirakendustes. Lahendust kasutatakse analüüsi ja andmeaitade tarbeks. Põhiandmed pole siin tehingute aluseks ega toeta andmeallikate värskendamist. Andmeida pealt tulenevad andmed oleksid põhiandmeobjektis analüütilised.



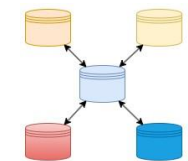
- **Kooseksisteeriv:** sarnane konsolideerivale stiilile, aga põhiandmed on avaldatavad ja ühtlustatakse heterogeensetesse süsteemidesse. Keskset andmebaasi kasutatakse kanalirakenduste põhiandmete replikatsiooniks, koondamiseks, puhastamiseks ja rikastamiseks. Uuendamine võib toimuda partiidena (*batch*) või jooksvalt, kui andmeallikates tuvastatakse muudatusi.



- **Tsentraliseeriv:** keskne liigipõhine andmebaas või register, millele on (kanali)rakendustel ja infosüsteemidel otseligiipääs (Xtee). Asutus ja riik otsustavad, milline andmebaas või register on milliste põhiandmeliikide peamine hoidla. See on täna kehtiv riigi põhiregistrite kontseptsioon.



- **Registreeriv:** põhiandmed asuvad ainult kanalirakendustes ja kasutakse keskset indeksit ja erinevaid integratsioonitehnikaid, et pääseda kiiresti ligi lähtesüsteemides asuvatele põhiandmetele ja neid vajadusel konsolideerida.



Asutuse infosüsteemide arhitektuuri mõjutab muuhulgas füüsiliste instantside ja andmebaaside arv, andmekeskuste ja tarbijate paiknemine, füüsilise viivituse piirangud, põhiandmete määramise ja kinnitamise asukoht, operatiivne või analüütiline andmevajadus jne. Asutuste orgaaniliselt arenenud infosüsteemid (erinevate väliste arenduspartnerite toel) ja haldusvahendite piiratus on pigem tekitanud lahendused, mis on täiesti eraldiseisvad (silod) või laetakse andmed lihtsalt hiljem konsolideerival viisil andmeaita. Need isoleeritud lähtesüsteemid jäävad ikka põhiandmeid dubleerima. Teine asutustes levinud arhitektuuri tüüp on tsentraliseeriv ja siin mängivad olulist rolli ka näiteks RIK-i registrid, kuhu asutuste rakendused üle X-tee otse pöörduvad.

Põhiandmete kogumiseks sobib kõige paremini registreeriv arhitektuur, mis võimaldab põhiandmete kogumist ja hoidmist ühes keskses hoidlas (asutuse või riigi repositoorium), kus andmeid saab kirjeldada, standardiseerida ja hallata ühtsel viisil. Projektis kirjeldatud põhiandmete objekt olekski see keskne viide, mille alusel erinevad

süsteemid ja rakendused pöörduksid läbi põhiandmete hoidla kindla asutuse poole, et kasutada vajalikke andmeid. See lähenemine tagaks andmekvaliteedi ning väldiks kõrge riskiga superandmebaaside tekkimist.

2.5 Suurbritannia kogemus põhiandmete käsitlemisel

Suurbritannia valitsus soovib kasutada andmete maksimaalset potentsiaali, et nende põhjal teha paremaid andmepõhiseid otsuseid, mis kokkuvõttes aitaksid valitsusel parendada avalike teenuste osutamist, mõista poliitikate ja programmide mõju ning tagada ressursside tõhus kasutamine. Riiklikus andmestrategias on sätestatud valitsuse ambitsioonid kogu digitaalmajanduse andmete teemal ning loodud vajalikud sammud selle rakendamiseks. Riiklikus andmestrategias on kindlaks määratud muutused, mis on vajalikud valitsuse andmete visiooni elluviimiseks tehnilisel, tehnoloogilisel ja kultuurilisel maastikul, sealhulgas:

- andmete kvaliteedi parendamine,
- andmestandardite kehtestamine ja nende vastuvõtmise arendamine,
- andmete vabajagamise ja nende riskasutamise lähenemisviisi kasutamine,
- valitsuse kontrolli ja vastutuse suurendamine,
- tõhus andmete jagamine asutuste vahel.

Tuginedes andmestrategia kohustusele valitsuse andmete kasutamiseks, avaldas valitsuse keskne digitaal- ja andmeamet (CDDO – *Central Digital and Data Office*) digitaalse tuleviku jaoks tegevuskava aastateks 2022 – 2025.⁴

Tegevuskava sätestab ühise valitsusülese visiooni aastaks 2025 ning paneb paika konkreetsed tegevused, mida ühiselt visiooni saavutamiseks ette võetakse. Tegevuskava on välja töötatud koos valitsuskabineti liikmetega ja digieksperditidega. Koostöös on välja töötatud konkreetsed, mõõdetavad ja ambitsioonikad eesmärgid, mis võimaldavad jälgida ja toetada progressi. Tegevuskava on mõeldud avaliku sektori asutustele ega kehti otseselt kohalikele omavalitsustele ja haldusasutustele, kuid nad on siiski kaasatud, et toetada kohalike teenuste reformi ja julgustada vajadusel liituma keskvalitsuse teenustega.

Kokku on määratletud kuus eesmärki ning 2023. aasta sügise seisuga on nendes saavutatud juba märkimisväärsed edusammud. Eesmärgid ja senised edusammud on järgmised:

- **Eesmärk 1 – Kujundada ümber avalikud teenused, mis võimaldavad saavutada õigeid tulemusi**
Teenuseeksperdid on välja töötanud raamistiku ning välja valinud 75 Suurbritannia jaoks kõige olulisemat ja tähtsamat teenust, mis sisaldavad põhiandmeid. 2023. aasta lõpuks on 15 teenust saavutanud „suurepärase“ standardi, mis tähendab, et need teenused on kvaliteetsed ja taaskasutatavad. Tegevuskava lõpuks on eesmärk jõuda 50 kvaliteetse teenuseni.
- **Eesmärk 2 – Gov.uk ühine sisselogimine (Gov.uk One Login)**
Tegemist on teenusega, mis pakub kodanikele ühist sissepääsu riigiteenustele kiirelt, turvaliselt ning ühekordse autentimisega. 2023. aasta seisuga on liitunud GOV.uk One Login'iga 23 valitsusteenust ning seda rakendust on kasutatud rohkem kui kolm miljonit korda. Need arengud on võimaldanud sulgeda teised sarnased rakendused ning võtta kasutusele uus ja ühtne lahendus. Eesmärk on parandada kaasatust veelgi, et rohkem inimesi pääseks erinevatele teenustele veebis paremini ligi.
- **Eesmärk 3 – Paremad andmed otsuste tegemiseks**
Andmeomanikel ja nende meeskondadel on loodud andmete omandiõiguse mudel, mis määrab põhikohustused ja -rollid, mis toetavad ühtset andmete haldamist. Välja on töötatud andmete küpsushinnangu mudel, millega kõik valitsuse osakonnad on avaldanud valmisolekut 2023/2024. majandusaasta lõpuks töötama. Samuti on andmeomanikud koos oma meeskondadega välja töötanud

⁴ <https://www.gov.uk/government/publications/essential-shared-data-assets-and-data-ownership-in-government/essential-shared-data-assets-esdas-policy-html>

andmekataloogi ja andmeturu prototüübi, mis võimaldab ametnikel leida vajaminevaid andmeid ja saada neile ligipääs, et nende põhjal teha andmepõhiseid valitsemisotsuseid.

- **Eesmärk 4 – Tõhus, turvaline ja jätkusuutlik tehnoloogia**

Tehnoloogiajuhid ja nende meeskonnad on kasutusele võtnud raamistiku (*Legacy IT Framework*), mis mõõdab aegunud tehnoloogia kasutamise seotud riske ja pingutusi. 26 asutust on seda ka juba kasutanud. Lisaks on loodud teine raamistik (*Secure by Design*), mis võimaldab kavandada ja luua teenuseid, mis tagavad kodanike ja valitsuse turvalisuse.

- **Eesmärk 5 – Digitaalsete oskuse levitamine**

Valitsuse digitaal- ja andmeedastusalade kogumaht kasvas ajavahemikus 2022. aprillist kuni 2023. aasta aprillini 19%, täites sellega kriitilise tähtsusega digioskuse nõuded. 32 asutust on kasutusele võtnud ühise valitsuse digitaalse andmete raamistiku, mis aitab vähendada sõltuvust töövõtjatest ja säästa maksumaksja raha. Rohkem kui 600 riigiteenistajat on ennast täiendanud digitaalsete ja andmetega seotud teemade vallas, aidates luua digitaalsete oskustega juhtide meeskonda.

- **Eesmärk 6 – Digitaalse transformatsiooni süsteem**

Valituse digijuhid leppisid kokku ja avaldasid digitaalse funktsionaalse standardi uue versiooni, mis juhendab kõrgemaid juhte, kuidas hallata digitaalseid andmeid ja tehnoloogiat valitsuses. Tootekeskne raamistik võimaldab meeskondadel teha koostööd kvaliteetsete digitaalsete toodete ja teenuste pakkumisel.

Suurbritannia on pärast teekaardi käivitamist õppinud väga palju. Nad on kasutanud andmeid ja teadmiseid kogu valitsusalast, et aidata tuvastada väljakutseid ja valdkondi, mis vajavad tuge, et meeskonnad saaksid prioriseerida oma tegevusi ning vastavalt sellele oma lähenemist kohandada.⁵

⁵ <https://www.gov.uk/government/publications/roadmap-for-digital-and-data-2022-to-2025/transforming-for-a-digital-future-2022-to-2025-roadmap-for-digital-and-data#annex>

3. Andmete olulisuse hindamise metoodika

Kolmas peatükk kirjeldab projekti käigus välja töötatud ja piloteeritud hindamisvahendit, mis võimaldab asutustel saada ülevaade oma põhiandmetest. Peatükis on kirjeldatud põhiandmete objektide määratlemine (ptk 3.1), andmete olulisuse hindamise kriteeriumid (ptk 3.2), andmeobjektide hindamise loogika (ptk 3.3) ning hinnatud andmeobjektide tulemuste analüüs (ptk 3.4).

Põhiandmete olulisuse hindamine on andmehalduse protsessi kriitiline samm, mis aitab tagada andmete kvaliteeti, usaldusväärsus ja kasutatavust. Selle eesmärk on tuvastada need andmestikud ja andmeobjektid, mille kvaliteedi tõstmine ja andmeteenuste kasutatavuse parandamine annaks enim kasu. See võimaldab asutustel oma sisemisi andmehalduse ja andmete taaskasutusega seotud ressursse rakendada senisest efektiivsemalt. Põhiandmete olulisuse hindamine on vajalik, sest see aitab tuvastada dubleerivat põhiandmete kogumist.

Põhiandmete olulisuse hindamine on kasulik nii asutuse kui ka riigi vaatest. Asutuse vaates võib see anda:

- ülevaate oma asutuse põhiandmetest, mis on täpsed, kättesaadavad ning toetaksid protsesse ja otsuste langetamist;
- praktilise andmehalduse kasu ehk ühised andmekvaliteedi ja juurdepääsupiirangute reeglid;
- teadmise, kas kvaliteedinõuded üle erinevate asutuste vajavad ühtlustamist;
- ülevaate andmete taaskasutatavusest ehk kes ja kui palju asutuse andmeid kasutavad;
- teadmised võimalikest uutest võimalustest ja seostest äriaga;
- ülevaate andmete ühiskasutatavusest ja nende täpsusest ning asutuse kõigi üksuste juurdepääsust ühistele andmetele, mis sisaldavad ka andmestruktuure ja mõisteid, et vältida erinevaid või dubleerivaid tõlgendusi;
- ühised ja täpsemad põhiandmed, mis aitavad kaasa protsesside optimeerimisele ja vähendavad IT-analüüsikulu (efektiivsus);
- teadmised andmepõhiseks juhtimiseks, kus otsuste tegemiseks on oluline kvaliteetsete andmete olemasolu (juhid saavad tugineda andmetele);
- regulatsioonidele vastavad andmed (põhiandmete määratlemine aitab asutusel olla seadusega kooskõlas);
- sujuvama koostöö äriprotsesside ja -teenuste vahel (põhiandmete ühtsus võimaldab paremat infovahetust erinevate süsteemide ja osakondade vahel).

Riigi vaates aitab põhiandmete olulisuse hindamine vähendada põhiandmete kogumise dubleerimist, et ressursse säästa. Samuti aitab see määratleda nõuded, mille järgimine peaks edaspidi olema oluliste põhiandmete omanikele kohustuslik. See aitab parandada andmete kvaliteeti ja usaldusväärsust, mis on oluline riigi teenuste pakkumisel ja otsuste tegemisel.

Riigiülene vaade olulistest põhiandmetest saab tekkida vaid eeldusel, et kõikidel asutustel on olemas asutusesisene ülevaade oma põhiandmetest. Riigiülse vaate puhul on oluline, et igal andmeobjektil/andmekogul oleks määratud vastutav isik või omanik.

3.1 Põhiandmete objektide praktiline määratlemine

Põhiandmeobjektide määratlemisel on abistavaks materjaliks Eesti andmehalduse juhend⁶. Asutuse infosüsteemide loendi alusel tuleb kaardistada andmekogude tehnilised asukohad ja eksportida andmebaaside metaandmestik. Seejärel tuleb kirjeldada asutuse andmekataloog, luua vastavad ärisõnastiku mõisted ja märkida mõistete ning andmeväljade vahelised seosed, grupeerida mõisted ning määrata gruppide põhjal asutuse põhiandmed, nimetada ja kirjeldada põhiandmed nõutud parameetrite abil andmeobjektideks.

Soovitame koostada mõistete hierarhia kolm taset. Esimene on asutuseväliselt ja üleriiklikult mõistetav üldmõiste ehk termin, mis üldjuhul peegeldab reaalselt maailma (isik, kaup, aadress). Teine mõistete tase on

⁶ <https://www.kratid.ee/juhised>

asutus(t)e sisuteenusteülene ühtlustatud mõiste, mis kirjeldab kasutusjuhte ja sisaldab äriküsimusi. Kolmas mõistete tase on infosüsteemide ja andmebaaside tasand.

Parim praktika on kirjeldamisel siduda andmeväljad kolmanda või teise tasandi äriterminitega. Iga andmeväli peab olema seotud vähemalt ühe terminiga. Juba teadaolevad põhiandmeväljad tähistatakse andmehaldustööriistas kohe eraldi märkega.

Parima hinnangu andmekogu olulistele objektidele annab vastava äriteenuse juht või andmeomanik, kes tuvastab oma teadmiste põhjal andmekogu ärimõistete grupid ja määrab andmeobjekti ulatuse ehk tunnused, mida see veel võiks kaasata. Grupeerimine on mõistlik teha teise tasandi mõistetest, siis jääb veel hea võimalus kaasata kolmandat taset tunnusteks. Andmehaldur aitab vormistada põhiandmete objekti ja veendub et kõik nõutud parameetrid on kirjeldatud.

3.2 Andmete olulisuse hindamise kriteeriumid

Andmeobjekti on võimalik varana ja väärtusena kirjeldada ning talle esitavad kindlad ootused nii tarbijad kui ka asutused. Ootuseid ja väärtuseid on võimalik iseloomustada järgmistel alustel:

- andmete realiseeritud ja potentsiaalne väärtus asutuse sees ja väljas,
- hetkel tagatud andmete kasutatavuse ja kvaliteedi tase,
- andmekogu kvaliteeditase.

Käesolevas peatükis kirjeldatava põhiandmete olulisuse hindamise metoodika rakendamine hindamisvahendiga võimaldab asutustel kontrollida, kas andmete kasutatavus ja kvaliteeditase ning andmekogu kvaliteeditase adresseerivad tuvastatud andmete väärtust piisavalt hästi .

Põhiandmete olulisuse hindamise kriteeriumid pakuvad asutusele võimaluse paremini mõista, millised andmed on olulised nii oma asutuse eesmärkide saavutamiseks kui ka laiemalt, andes täiendavaid teadmiseid, kuidas andmeid kasutada ja hallata. Lisaks aitavad need kriteeriumid ja nende hinded identifitseerida võimalikke kitsaskohti, mis võivad mõjutada andmete kvaliteeti ja kasutatavust, võimaldades seeläbi asutusel teostada parendusi. Kriteeriumite kasutamine toetab asutuse otsustusprotsesse ning aitab ressursse paremini planeerida ja koordineerida. Samuti aitavad kriteeriumid tagada, et andmekogud ja nende haldus oleks korraldatud viisil, mis toetab asutuse põhitegevust ning aitaks kaasa strateegiliste eesmärkide saavutamisele.

Järgnevalt on loetletud andmeobjektide hindamiskriteeriumid neljas rühmas:

- äriline kontekst lõppkasutajale,
- äriline kontekst asutusele,
- andmeobjekti praeguse halduse ja kasutatavuse tase,
- andmekogu kvaliteet.

3.2.1 Äriline kontekst lõppkasutajale

Ärilise konteksti lõppkasutajate kriteerium on jagatud alamkriteeriumiteks:

- **Andmeobjekti atraktiivsus sisekasutajate jaoks** - kriteeriumi abil hinnatakse, kui paljud teised sama asutuse kasutajad/andmekogud kasutavad või soovivad võtta kasutusele andmeobjekti, et sellega oma äritegevust arendada. Näiteks võib sisekasutajate jaoks konkreetse andmeobjekti huvi puududa täielikult, olla osaline või täielik huvi. Täielik huvi peegeldub näiteks jätkusuutlikes liidestusvõimalustes (X-tee, RestAPI vm), eraldisesivas raporteerimislahenduses või rakendatud andmehaldusnõuetes.
- **Andmeobjekti atraktiivsus väliskasutajate jaoks** – kriteeriumi abil hinnatakse, kui palju välised kasutajad/registrid/andmekogud kasutavad või soovivad võtta kasutusele andmeobjekti, et sellega on äritegevust arendada. Näiteks võib väliskasutajate jaoks konkreetse andmeobjekti huvi puududa täielikult,

olla osaline või täielik huvi. Täielik huvi peegeldub näiteks jätkusuutlikes liidestusvõimalustes (X-tee, RestAPI vm), eraldisesivas raporteerimislahenduses või rakendatud andmehaldusnõuetes.

- **Andmeobjekti kasutuse maht erinevates teenustes, raportites ja muudes väljundites** – kriteeriumi abil hinnatakse, kui paljudes erinevates teenustes, raportites ja muudes kasutusjuhtudes on andmeobjekt kasutusel. Näiteks iseteeninduskeskkonnas, andmeteenuses või dünaamilises raportis.
- **Andmeobjekti kasutatavus poliitikakujundamises, andmeteaduses või riigi tasandil masinõppe arendamiseks** – kriteeriumi abil hinnatakse, kas andmeobjekti kasutatakse näiteks andmete visualiseerimisel, masinõppe rakendamisel, mõõdikute kirjeldamisel või andmepõhiste otsuste langetamisel. Lisaks on andmeobjekti väärtus kõrgem, kui andmeobjekt on lõppkasutajale orienteeritud, seotud tehisintellektiga või seda kasutatakse statistika koostamisel.
- **Andmeobjekti mõju sotsiaalkeskonnale, inimkeskkonnale või keskkonnale laiemalt** – kriteeriumi abil hinnatakse, kas andmeobjekti mõju võib avalduda näiteks nullbürokracias, digiprügis või kestlikkuses. Mida jätkusuutlikum ja eeskujulikum on andmeobjekt asutuses, seda olulisem on selle mõju keskkonnale.
- **Andmeobjekti huvi avaandmetena laiemale üldsusele** – kriteeriumi abil hinnatakse, kas andmeobjekt pakub väärtust laiemale üldsusele läbi avaandmete (näiteks aadressiandmed, ettevõtte tegutsemise load, spordirajatiste andmed vm). Kriteeriumi mõju on kõrge juhul, kui andmeobjektile on rakendatud avaandmete andmehaldusnõuded, andmed on kättesaadavad asutuse või MKM-i portaalides ning andmetel on analüüsi ja graafilise esitamise võimekus.
- **Andmeobjekti kasutatavus riikliku statistika koostamisel** – kriteeriumi abil hinnatakse, kas andmeobjekti kasutab või soovib kasutada Statistikaamet riikliku statistika koostamiseks.

3.2.2 Äriline kontekst asutusele

Äriline kontekst asutuse kriteerium on jagatud alamkriteeriumiteks:

- **Andmeobjekti tõendusväärtuse olemasolu** – kriteeriumi abil hinnatakse, kas andmeobjektile on olemas tõendusväärtus, mis pärineb näiteks toimingust, deklaratsioonist või määrusest. Tegemist on lihtsa jah/ei väärtusega.
- **Andmeobjekti arhiiviväärtuse olemasolu** – kriteeriumi abil hinnatakse, kas andmeobjektile on olemas arhiiviväärtus, mis pärineb näiteks mõnest toimingust. Tegemist on lihtsa jah/ei väärtusega.
- **Andmeobjekti mõju asutuse strateegiliste eesmärkide saavutamiseks** – kriteeriumi abil hinnatakse, kas andmeobjekti kasutatakse strateegiliste eesmärkide saavutamiseks. Näiteks, kas kasutatakse andmepõhiste otsuste tegemisel, andmete visualiseerimisel või statistika koostamisel. Andmeobjekti kõrgem mõju ja väärtus avaldub siis, kui asutusel rakendatud andmeobjektile andmehaldusnõuded või kui andmeelemendid on defineeritud asutuse juhtimismõõdikutena.

3.2.3 Andmeobjekti praeguse halduse ja kasutatavuse tase

Andmeobjekti praeguse halduse ja kasutatavuse taseme kriteerium on jagatud alamkriteeriumiteks:

- **Andmeobjektide rist- ja erinevas kontekstis kasutamise tõhusus asutuse kui terviku vaatest** – kriteeriumi abil hinnatakse, kas sama andmeobjekti on võimalik kasutada või kasutatakse mitmes erinevas andmeteenuses, raportis, iseteeninduskeskkonnas või mujal. Näiteks kas on olemas mõni sisemine andmeteenus, regulaarne raport, masin-masin liides ning kas andmehaldusnõuded on rakendatud neile.
- **Andmeobjekti dubleerituse tase** – kriteeriumi abil hinnatakse, kas andmeobjekti on dubleeritud ehk kasutatakse samas või mõnes teises andmekogus.

- **Andmeobjekti digitaliseerituse tase** – kriteeriumi abil hinnatakse andmeobjektiga seotud toimingute ja protsesside digitaliseerituse taset. Näiteks andmete kogumine, andmete haldamine või andmete arhiveerimine.
- **Andmeobjektide elutsükli hallatavuse määr asutuse kui terviku vaates (järjepidevus)** – kriteeriumi abil hinnatakse, kas andmeobjekti elutsükli alla kuulub näiteks andmete arhiveerimine, andmehaldusnõuete rakendamine, andmekvaliteedi tööd või ärireeglite haldamine.
- **Andmehaldussüsteemi kasutus andmeobjekti haldamiseks** – kriteeriumi abil hinnatakse, kas andmeobjekti on võimalik hallata selleks spetsiaalselt loodud süsteemis, kus on võimalik andmeid kirjeldada ja hallata. Näiteks RIHAKE.
- **Andmeobjekti võrdlemine välise allikaga andmete erinevuste tuvastamiseks** – kriteeriumi abil hinnatakse, kas andmeobjekt võib olla kasutusel ka mõnes teises välises allikas või andmekogus. Oluline on võrrelda andmeobjekti välises allikas olevatega, et tuvastada erinevusi ning seeläbi tagada andmete ajakohasus.

3.2.4 Andmekogu kvaliteet

Andmekogu kvaliteedi kriteerium on jagatud alamkriteeriumiteks:

- **Andmekogule määratud omaniku/vastutaja olemasolu** - kriteeriumi abil hinnatakse, kas andmekogule on määratud kindel omanik või vastutaja, kellel on määratud selged ja konkreetset ülesanded.
- **Ligipääs andmekogule ja andmete otsene kasutatavus** – kriteeriumi abil hinnatakse väliste osapoolte ligipääsu andmekogule ja seal olevatele andmetele. Näiteks kas ligipääsemiseks on loodud API liidesed, X-tee teenused või on andmed tehtud kättesaadavaks näiteks avaandmetena.
- **Andmekogu kohandatavus vajadustega (kiirus/hind)** – kriteeriumi abil hinnatakse andmekogus muudatuste teostamise kiirust. Kas on olemas meeskond, kes suudab tehnilised ja ärilised muudatused andmekogusse koheselt sisse viia?
- **Andmete integreeritus andmekogusse paljudest sisemistest ja välistest tõendusväärtusega allikatest** – kriteeriumi abil hinnatakse, kui suur hulk andmekogusse kuuluvaid andmeid on pärit teistest sisemistest või välistest allikatest ning kas need allikad on tõendusväärtusega ehk kas seal olevad andmed on kvaliteetsed ja ajakohased.
- **Andmete uuendamise õigeaegsus/kiirus andmekogus (viivitused)** – kriteeriumi abil hinnatakse, kui tihti uuendatakse andmekogus olevaid andmeid (reaalajas, kord päevas kord kuus vm). Samuti hinnatakse, kas andmete ajakohasus on kontrollitud ning kas jälgitakse andmete muutuseid välistes allikates.
- **Andmete ulatuse piisavus avalike ülesannete täitmiseks (täielikkus)** – kriteeriumi abil hinnatakse, kas lõppkasutajate vajadused ja nõuded on rahuldatud ning andmeid on võimalik täiendada vastavalt kasutajate soovidele. Olulisel kohal on andmehalduse rakendatavus ning võimalus andmeobjekte juurde luua.
- **Andmete järjepidevus andmekogus** – kriteeriumi abil hinnatakse, kas andmed on kirjeldatud koos äri- ja kvaliteedireeglitega, toimuvad automaat- või manuaalkontrollid ning kas andmekogule on määratud andmeomanik, kes tagab andmete järjepidevuse ja teostab vajadusel muudatused.

3.3 Andmeobjektide hindamine

Väljapakutud põhiandmete hindamisvahend võimaldab iga andmeobjekti või andmekogu hinnata eelnimetatud kriteeriumite alusel. Kriteeriumite puhul hinnatakse andmeobjekti vastavust ja mõju. Vastavus on asutuse tänase olukorra hinnang ehk kuidas andmeobjekt vastab praegusel hetkel igale kriteeriumile. Mõju on asutuse hinnang sellele, kui oluline hinnatav kriteerium konkreetse andmeobjekti lõikes on. Nii vastavust kui ka mõju hinnatakse skaalal 0–3 ning nende tähendused on järgmised:

Tabel 2. Kriteeriumite hindamise vastavuse ja mõju skaalad

Vastavus	Selgitus	Mõju	Selgitus
0	Asutusel kriteeriumile vastavus puudub täielikult	0	Asutuse arvates kriteerium ei ole asjakohane
1	Asutusel kriteeriumile vastavus vähesel määral	1	Asutuse arvates kriteeriumi asjakohasus väike või ei anna piisavat efekti
2	Asutusel kriteeriumile vastavus suurel määral	2	Asutuse arvates kriteeriumi asjakohasus keskmine, sest mõjutab palju erinevaid huvigruppe
3	Asutusel kriteeriumile vastavus eeskujulik ja jätkusuutlik	3	Asutuse arvates kriteerium kriitiline, sest sellest sõltub otseselt klientide äritegevus ja tulu

Kõikide kriteeriumite puhul arvutatakse välja andmeobjekti kohta käiv kriteeriumi skoor, mis tuleneb vastavuse ja mõju korrutisest. Näiteks kui mõne andmeobjekti kohta hinnatakse „Andmeobjekti atraktiivsus väliskasutajate jaoks“ kriteeriumit vastavusega 3 ja mõjuga 2, siis konkreetse andmeobjekti kriteeriumi skoor on $3 * 2 = 6$.

Olles ära hinnanud kõikide kriteeriumite vastavused ja mõjud ning välja arvutanud nende skoori, arvutab hindamisvahend välja nimetatud nelja kategooria (vt pkt 3.2) üldise skoori, mis arvutatakse järgmise valemi järgi:

$$\text{Kategooria skoor} = \frac{\sum \text{Kõikide kategooria kriteeriumite skoor}}{\sum \text{Kõikide kategooria kriteeriumite mõju}}$$

Joonis 2. Kategooria skoori arvutamise valem

Iga kategooria skoor välja arvutatuna võimaldab välja arvutada **andmeobjekti väärtuse**, mis leitakse kategooriate keskmise jagamisel kategooriate arvuga. Alljärgnevas tabelis on toodud illustreeriv näide kolme andmeobjekti väärtuse leidmise kohta.

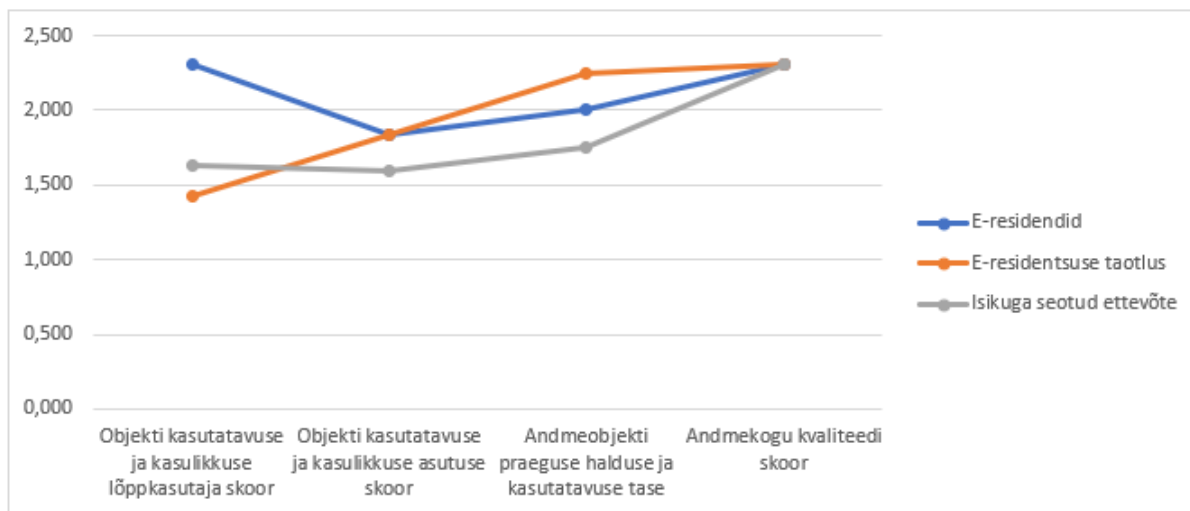
Tabel 3. Andmeobjekti väärtuse arvutamise näide

Andmekogu	E-residentide register		
	Andmeobjektid	E-residendid	E-residentsuse taotlus
Objekti kasutatavuse ja kasulikkuse lõppkasutaja skoor	2,308	1,429	1,625
Objekti kasutatavuse ja kasulikkuse asutuse skoor	1,833	1,833	1,600
Andmeobjekti praeguse halduse ja kasutatavuse tase	2,000	2,250	1,750
Andmekogu kvaliteedi skoor	2,313	2,313	2,313
Andmeobjekti väärtus	52,835 %	48,903 %	45,547 %

3.4 Tulemuste analüüs ja visualiseerimine

Olles ära hinnanud kõik andmeobjektid kõikide kriteeriumite alusel, võimaldab töövahend tulemusi visualiseerida ja analüüsida. Tulemuste analüüs annab asutusele ülevaate tema andmekogudest ja andmeobjektidest ning aitab aru saada, milliste andmeobjektidega on olulisem esmajärjekorras tegeleda.

Üks lahendus tulemuste visualiseerimiseks on vaadata ühe andmekogu lõikes kõiki hinnatud andmeobjekte eelnevalt nimetatud nelja kriteeriumite kategooria kaupa. Kategooriate skooore on võimalik kõrvutada omavahel visuaalselt, mis annab tulemuseks andmeobjekti väärtuse. Iga kriteeriumi kategooria alusel on võimalik hinnata andmeobjekti kasutatavust, hallatavust ning kvaliteeti. Joonisel 4 on näitena kujutatud Tabelis 2 toodud e-residentide registri andmeobjekte.



Joonis 3. E-residentide registri andmeobjektide skooride visualiseerimine

Teine lahendus tulemuste analüüsimiseks on võrrelda teatud kriteeriumeid omavahel ühe andmeobjekti lõikes. Tabelis 3 on välja toodud mõned kriteeriumid, mille omavaheline võrdlemine võib anda asutusele vajaliku sisendi andmeobjekti kasutatavuse või kvaliteedi kohta. Iga võrreldava omaduse kohta on välja toodud valem selle arvutamiseks ning kirjeldus, kuidas tulemust tõlgendada. Saadav tulem võimaldab asutusel saada täpsema ülevaate oma andmeobjektidest ning nende võimalikest ebakõladest, millele tuleb tähelepanu osutada. Kuigi arvutuse tulemusena võib välja tulla, et andmeobjekti võrdluses võib olla sees ebakõla, siis sellele võib olla loogiline seletus. Näiteks kui andmeobjektil on kõrge väärtus välise kasutaja jaoks, kuid välised kasutajad ei kasuta seda andmeobjekti üldse, võib põhjenduseks olla, et tegemist on asutuse jaoks konfidentsiaalse andmeobjektiga ning seda ei tohigi jagada väliseks kasutamiseks.

Tulemuse analüüsi idee on välja filtreerida andmeobjektis esinevad ebakõlad, mille järel asutus saab otsustada, kas need ebakõlad vajavad tähelepanu.

Tabel 4. Andmeobjektide vahelised võimalikud võrdlused koos järeldustega

Valem	Kirjeldus	Valem	Järeldus
As / Sk	Andmeobjekti atraktiivsus sisekasutaja jaoks suhtes andmeobjekti sisemise kasutusega	$\frac{\text{AVG (atraktiivsus sisekasutaja jaoks + mõju asutuse strateegilistele eesmärkidele)}}{\text{Sisemine kasutuse maht erinevates teenustes, raportites ja muudes väljundites}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, on andmeobjekti atraktiivsus sisekasutaja jaoks väike, kuid selle sisemine kasutus on suur. Tulemuses mingi ebakõla, mis tuleb üle vaadata. Juhul kui väärtus on 1, on andmeobjekti atraktiivsus ja kasutatavuses asutuse sees kooskõlas Juhul kui väärtus on üle 1, on andmeobjekti atraktiivsus suur, kuid seda ei kasutata sisemiselt piisavalt ära. Võimalik, et andmeobjekti tuleks asutuses teiste osakondade või süsteemide seas rohkem tutvustada.
Av / Vk	Andmeobjekti atraktiivsus väliskasutaja jaoks suhtes välimise kasutusega	$\frac{\text{AVG (atraktiivsus väliskasutaja jaoks + mõju keskkonnale + huvi avaandmete vastu + kasutatavus riikliku statistika koostamisel)}}{\text{Kasutatavus poliitikakujundamises, andmeteaduses või riigi tasandil masinõppe arendamiseks}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, on andmeobjekti atraktiivsus väliskasutaja jaoks väike, kuid kasutus on suur. Tulemuses on ebakõla. Andmeobjekti tegelikkuses tundub, et andmeobjekt on atraktiivsem kui seda on hinnatud. Juhul kui väärtus on 1, on andmeobjekti atraktiivsus ja kasutatavus väliskasutajate jaoks kooskõlas. Juhul kui väärtus on üle 1, on andmeobjekti atraktiivsus suur, kuid kasutus on väike. Ilmselt peaks andmeobjekti rohkem tutvustama välistele kasutajale, et nad selle kasutusele võtaksid.
Sk / Tõarh	Andmeobjekti sisemine kasutus suhtes tõendus- ja arhiiviväärtuse olemasoluga	$\frac{\text{Sisemine kasutuse maht erinevates teenustes, raportites ja muudes väljundites}}{\text{AVG (tõendusväärtuse olemasolu + arhiiviväärtuse olemasolu)}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, on andmeobjektidel tõendus- või arhiiviväärtus, kuid seda ei kasutata piisavalt hästi asutuse sees ära. Ilmselt võiks tutvustada andmeobjekti kasutusvõimalusi teistele asutuse andmekogudele või osakondadele. Juhul kui väärtus on 1 või rohkem, kasutatakse andmeobjekti piisavalt hästi sisemiselt ära võrdluses tema arhiivi- või tõendusväärtusega.
Vk / Tõarh	Andmeobjekti välimine kasutus suhtes tõendus- ja arhiiviväärtuse olemasoluga	$\frac{\text{Kasutatavus poliitikakujundamises, andmeteaduses või riigi tasandil masinõppe arendamiseks}}{\text{AVG (tõendusväärtuse olemasolu + arhiiviväärtuse olemasolu)}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, on andmeobjektidel tõendus- või arhiiviväärtus, kuid seda ei kasutata piisavalt hästi asutuseväliselt ära. Ilmselt võiks tutvustada andmeobjekti kasutusvõimalusi teistele välistele asutustele ja osapooltele.

Valem	Kirjeldus	Valem	Järeldus
			<ul style="list-style-type: none"> Juhul kui väärtus on 1 või rohkem, kasutatakse andmeobjekti piisavalt hästi väliselt ära võrdluses tema arhiivi- või tõendusväärtusega.
Usaldus / Tõarh	Andmeobjekti võrreldavus väliste allikatega suhtes tõendus- ja arhiiviväärtuse olemasoluga	$\frac{\text{Võrdlemine välise allikaga erinevuste tuvastamiseks}}{\text{AVG (tõendusväärtuse olemasolu + arhiiviväärtuse olemasolu)}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, ei ole andmeobjekt piisavalt usaldusväärne, sest seda ei võrrelda piisavalt väliste allikatega. Juhul kui väärtus on 1 või rohkem, võib andmeobjekti usaldada, sest teda võrreldakse piisavalt regulaarselt väliste allikatega erinevuste tuvastamiseks.
Tõhusus / Tõarh	Andmeobjekti dubleerituse tase ja riskasutatavus suhtes tõendus- ja arhiiviväärtuse olemasoluga	$\frac{\text{AVG (andmeobjekti riskasutuse tõhusus + andmeobjekti dubleerituse tase)}}{\text{AVG (tõendusväärtuse olemasolu + arhiiviväärtuse olemasolu)}}$	<ul style="list-style-type: none"> Juhul kui väärtus on alla 1, ei riskasutata andmeobjekti piisavalt hästi või andmeobjekt on dubleeritud. Juhul kui väärtus on 1 või rohkem, riskasutatakse andmeobjekti piisavalt hästi teistes andmekogudes ning andmeobjekti mittedubleeritus on kooskõlas tema väärtusega.

4. Soovitused asutuse põhiandmete haldamise korraldamiseks

Põhiandmete haldamise eesmärgiks on ühtlustada organisatsiooni põhiliste, riskasutamist vajavate äriolemite (nt. organisatsioon, isik, teenus, asukoht, vara, leping) andmed ning tagada sisemistele- ja välistele andmetarbijatele usaldusväärne allikas. Põhiandmeid peab olema võimalik organisatsiooni infosüsteemide vahel efektiivselt jagada, soovituste järgimine aitab tuvastada andmetest korrektseima ja ajakohaseima versiooni ning tekitada ühtlustatud tõeinstants. Asutuse siseselt korrektselt hallatud põhiandmed võimaldavad hiljem ka riigil paremini mõista ametite/asutuste teenuseid ja -tarbijaid ning teha teadlikumaid otsuseid, mis viivad kõigi jaoks parema tulemuseni.

4.1 Põhiandmete haldus

- Põhiandmete haldamisel tuleks esmalt järgida seadusandlusest tulenevad nõudeid ja riigis kokkulepitud standardeid, meetodikaid ning protsesse, mis võimaldavad tõhusamat andmete kogumist, säilitamist, analüüsimist ja levitamist (vt. ka ptk. 2.3 ja 2.4).
- Andmete ühildamise järel tuleb asutuse rakendustele ja infosüsteemidele tagada ka ligipääs loodud põhiandmeobjektidele, kas siis otsepöördumise, andmeteenuste või analüütilise andmeaia kaudu. Samuti tuleb pakkuda ligipääsu või andmeteenust kasutamiseks alust omavatele välistele kasutajatele.
- Ühiselt kasutatavad andmed, laiem ligipääsetavus nõuab ka täiendavat kaitset. Järgida tuleb andmete konfidentsiaalsuse, terviklikkuse ja kättesaadavuse tagamise regulatsioone (vt. ka ptk. 2.2).
- Andmete regulaarne profileerimine ja auditeerimine kindlustavad haldusprotsesside vastavuse standarditele ja parimatele praktikatele, avastatud probleemid tuleb lahendada esimesel võimalusel.
- Asutus peab kehtestama oma teabehalduse korras selged poliitikad ja protseduurid ka põhiandmete elutsüklile: säilitamine, kasutamine, jagamine ja kustutamine.

4.2 Põhiandmete kvaliteedi haldus

- Andmeomanikud teevad, vajadustest tulenevalt, ettepanekuid asutuse infosüsteemide vahel andmete ühildamiseks ja koondamiseks, et tekitada asutuse põhiandmetest „tõeseim“ versioon.
- Andmehaldurid tuvastavad ühildamise võimalikkuse ja piirangud, sobitavad hiljem valesti liidetud andmed ning mõisted ja loovad täpsemate tunnuste abil sobivamad seosed.
- Põhiandmete kvaliteet peab olema tagatud läbi viie dimensiooni: täpsus, täielikkus, ajakohasus, unikaalsus ja reeglipärasus. Kvaliteedi operatiivne tagamine toimub koostöös IT-ekspertidega.
- Põhiandmete kvaliteedi haldamine peab olema süsteemne ja proaktiivne.

4.3 Põhiandmete kaardistamine, kirjeldamine ja kirjelduste avalikustamine

- Andmekogud tuleb kirjeldada andmehalduse töövahendis, sealhulgas määrata andmete päritolu, tähendus, struktuur, omanikud ja kasutusjuhised.
- Andmeomanikud kirjeldavad põhiandmetega seonduvate tunnuste atribuudid ja kirjeldavad kasutustingimused.
- Andmete valideerimisel, standardiseerimisel ja riskasutamisel tuleb pidevalt jälgida:
 - Toimuks vigaste kirjete tuvastamine ja valideerimine.
 - Andmete struktuuri ja formaadi ühtsust. Tagada andmete sisu ja viiteandmete vaheline vastavus.
 - Põhiandmete rikastamise võimalust erinevate tunnuste abil.
- Põhiandmete kirjeldused tuleb avalikustada andmete teabeväravas, RIHA's või asutuse avalikus andmekataloogis. Sealhulgas tuleb avalikustada:
 - Andmekvaliteedi metodoloogia või hinnang, et kasutajad saaksid kaaluda andmete sobivust

- Andmete tehniline vorming/mudel
- Andmete kasutustingimused või olemasolev teenuste informatsioon

4.4 Põhiandmete riskasutust võimaldavad andmeteenused

- Põhi- ja viiteandmed peavad olema efektiivselt jagatud üle kogu organisatsiooni.
- Andmete ühtlustamisel tuvastatud erinevad andmekirjed, mis viitavad ühele ja samale või kahele erinevale päris maailma objektile. Ühtlustamine toimub läbi otsustuspuu- või tõenäosuspõhiste algoritmide ja reeglite abil.
 - Valepositiivne tulemus – ekslikult seotakse kaks kirjet, mis tegelikult ei viita samale objektile
 - Valenegatiivne tulemus – kaks tegelikult samale objektile viitavat kirjet jäävad seostamata
- Valest ühtlustamisest tulenevad vigu ja ühtlustamisreegleid võib parandada käsitsi.
- Info andmeteenuste käideldavuse ja turvalisuse kohta peab olema kergesti kättesaadav.
- Andmeteenuste käideldavusenõuded peavad lähtuma muuhulgas ka kasutajate vajadustest.
- Andmeteenused peavad olema dokumenteeritud ning seotud andmekirjeldustega.

4.5 Põhiandmete omaniku poolt kasutajatele pakutavad teenustasemed

- Põhiandmete ühtlustatud objektid peavad kajastama asutuse antud hetke parimat teadmist. Lisanduvad andmed võivad senised andmete sobitamise eelduseid ja reeglid muuta.
- Andmete sobitamise reegleid ning kõik tehtud muudatused peavad olema tagasipööratavad.
- Põhiandmete omanik peab kehtestama selged teenusetaseme ja toe põhimõtted ning need avalikustama.
- Põhiandmete omanik peab pakkuma kasutajatele tuge järgmiste tegevuste läbiviimisel:
 - Liidestussoovide vastuvõtt ja neile vastamine
 - Põhiandmeobjektide andmekvaliteedi hindamine (profileerimine)
 - Integratsiooni keerukuse hindamine, kommunikatsioon, kulu analüüs
 - Allika andmehõive test ja mõju hindamine seniste vastavus- ja kvaliteedireeglitele
 - Allika andmekvaliteedireeglite ja -mõõdikute loomine.
 - Integratsiooni lõplik teostus

4.6 Põhiandmete omaniku kommunikatsioon kasutajatega

- Asutustel peab olema kehtestatud andmehalduse kord, mis määrab juhtimisstruktuuri ja vastutused ka põhiandmete haldamise eest, lisaks vastused andmekvaliteedi tagamise, poliitikate väljatöötamise, jõustamise, kommunikatsiooni ning koolituste pakkumise eest.
- Viiteandmete väärtuste muudatused tuleb enne rakendandamist kooskõlastada.
- Põhiandmete omanik peab kaardistama talle teadaolevad kasutajad.
- Põhiandmete omanik peab andma aegsasti teada teenuste katkestustest.
- Põhiandmete omanik peab kasutajatega läbi arutama andmemudeli/atribuutide muudatused ning nendest ka piisavalt varakult ette teatama.
- Põhiandmete omanik peab andma kasutajatele selge võimaluse soovitada muudatusi andmete haldamises.
- Põhiandmete omanik peab pakkuma infokanali andmekvaliteedi vigadest teavitamiseks.

5. Kokkuvõte

Andmete olulisuse hindamine on suur ja oluline samm andmete efektiivsema haldamise, suurema usaldusväärsuse ja parema andmekvaliteedi saavutamiseks nii asutuse kui ka riigi tasandil. Hinnatud andmeobjektid ja andmekogud võimaldavad asutusel saada ülevaate oma põhiandmetest. Kvaliteetsed ja kättesaadavad andmed toetavad asutuse äriprotsesse ning võimaldavad teha andmepõhiseid juhtimisotsuseid.

Andmete olulisuse hindamise meetoodika võimaldab saada ka riigiülese vaate kõikidest riigis kasutusel olevates andmekogudest ja põhiandmetest. Selle jaoks on oluline, et asutused töötaksid läbi oma andmestikud ning täidaksid ja hoiaksid seejärel hindamisvahendit ka ajakohasena. Ainult juhul, kui asutused on täitnud oma osa, tekib ka riigiülene vaade ning seeläbi potentsiaalne kasu riigile tervikuna. Riigi tasandil aitab põhiandmete olulisuse hindamine vähendada dubleerivat andmete kogumist ning seeläbi hoida kokku ressursse. Samuti võimaldab see paremini rakendada andmete *once-only* printsiipi, mis tähendab, et avalike ülesannete täitmisel riigi andmekogusse kogutud põhiandmeid peab olema võimalik ka teistel andmekogudel kasutada, selle asemel, et neid ise uuesti kodaniku või ettevõtte käest koguda.

Põhiandmete haldamisel on oluline, et igal andmestikul oleks määratud andmeomanik, kes tagab andmete kvaliteedi ja ajakohasuse. Kindlasti on oluline, et andmeomanikud omavahel teeksid tihedat koostööd ja jagaksid informatsiooni ja teadmiseid. Samuti on oluline, et riik koordineeriks andmeomanike vahelist koostööd.

Soovitused kogu siinse paketi juurutamiseks lähema paari aasta jooksul on välja toodud eraldi järgmistes dokumentides:

- Õigusanalüüs;
- Soovitused Majandus- ja Kommunikatsiooniministriumile.

6. LISAD

6.1 Lisa 1 – Hindamisvahendi piloteerimise tulemused

Eelpool kirjeldatud (ptk 3) andmete olulisuse hindamise meetodikat piloteeriti kokku nelja erineva avaliku sektori asutuse poolt. Piloteerimiseks valiti välja asutused võttes arvesse, et valimis oleksid erineva suurusega andmestikud ning erineva andmehaldustasemega asutused. Samuti arvestati asutuste võimalusi ja soove osaleda piloteerimises.

Hindamisvahendi piloteerimine algas ühise töötoaga, kus projektimeeskond tutvustas valminud hindamisvahendit asutustele ning arutleti selle arusaadavuse üle. Seejärel oli asutustel umbes kolm nädalat aega täita hindamisvahendit oma asutuse andmetega ning anda hindamisvahendile tagasisidet. Samuti oli piloteerimise perioodi jooksul võimalik jooksvalt projektimeeskonna tuge saada. Seejärel koguneti taas ühiselt kokku töötuppa, mille eesmärgiks oli arutada täidetud hindamisvahendi ja kogu piloteerimise protsessi üle. Töötoas vaadati üle kogu hindamisvahend ning arutleti selle puuduste ja positiivsete külge üle. Samuti arutleti hindamisvahendi kasust asutusele ja riigile tervikuna.

Asutuste tagasiside hindamisvahendile

Piloteerimise käigus hindasid asutused kahte erinevat andmekogu ning mõlemas andmekogus kolme erinevat andmeobjekti. Asutuste hinnangul kulus ühe andmekogu kolme andmeobjekti hindamise peale kokku keskmiselt 3-4 tundi. Sinna alla kuulub andmeobjektide defineerimine ja olulisemate valimine, kriteeriumite alusel andmeobjektide hindamine ning hindamisvahendile tagasiside andmine. Siinkohal toodi olulise aspektina välja, et sellise ajakuluga hindamisvahendi täitmine on võimalik vaid juhul, kui seda teeb asutuse töötaja (andmehaldur või andmeomanik), kes on valitud andmestikuga väga pikalt töötanud ning tunneb seal asuvaid andmeid väga hästi.

Andmeobjektide defineerimine piloteerinud asutustele probleeme ei valmistanud. Keerulisem andmeobjektide juures oli prioriseerimine, et milline neist defineeritud andmeobjektidest on olulisem ja millisega minna hindamisvahendi täitmisega edasi.

Hindamisvahendis olnud kriteeriumid olid üldiselt arusaadavad ning nende mõistmiseks oli väga palju kasu ka täpsustavast selgitusest ning iga kriteeriumi kohta välja toodud skaala selgitusest. Kriteeriumite puhul ei tuntud puudust mõnes täiendavast kriteeriumist, mis võiks olla oluline andmeobjekti hindamisel. Mõned asutused tõid välja, et kohati oli kriteeriumite tõlgendamine keeruline ning oleks soovinud täpsemaid selgitusi. Näiteks oli keeruline hinnata mõnda andmeobjekti, sest skaalade selgitused olid liiga üldsõnalised. Sooviti täpsemaid selgitusi ja konkreetsemaid näiteid skaalade kohta, mis aitaks valida sobiva hinnangu.

Lisaks toodi välja, et mõne kriteeriumi puhul on olulisus tihedalt seotud mõjuga. Näiteks mõne andmeobjekti puhul, kui kriteeriumi olulisus on väike, siis on ka selle mõju väike. Sellistel juhtudel arvati, et võiks kriteeriumi hindamine toimuda ühel skaalal, kuid ei osatud kohe välja pakkuda, milline see skaala võiks siis välja näha.

Kokkuvõttes arvasid asutused, et selline andmeobjekti põhine lähenemine hindamisvahendile on väga hea, sest see hakkab selliselt kokku jooksma ka asutuse kontseptuaalse andmemudeliga ning andmeobjekti tasand võimaldab kasu saada ka protsesside kirjeldamisel. Hindamise koha pealt tuleb parandada kriteeriumite arusaadavust ning täitmisel intuiitiivsust. Ideaalis peaks hindamisvahendis kirjeldatud kriteeriumid olema väga selgesti mõistetavad ja arusaadavad, et selle täitmine oleks võimalikult lihtne ja kiire. Hetkel sellisel kujul hindamisvahendi täitmine on väga ajamahukas ning andmeomanikel üldiselt ei ole piisavalt ressursi, et selline ülesanne ette võtta.

Asutuste tagasiside hindamisvahendi tulemuste analüüsile ja visualiseerimisele

Asutused ei näinud piloteerimise lõpus sellest hindamisvahendist suurt kasu, sest hinnatud andmeobjektide ja andmekogude maht oli liiga väike. Hindamisvahendi tulemustest oodati täpsemaid ja selgemaid soovitusi, kuidas tuleb andmeobjektidega edasi tegeleda, et neid parendada. Samuti oodati, et visuaal tooks paremini välja andmeobjektid, millega ei ole hetkel kõik korras ning vajavad tegelemist.

Lisaks arvati asutuste poolt, et saadud tulemusi on oluline võrrelda ajas pidevalt. Sellisel juhul tekib täpsem arusaam, kuidas sisse viidud muudatused on mõjutanud andmeobjekti paremuse poolt. Oluline on, et hindamisvahendit saaks kasutada regulaarselt. Selle jaoks on oluline, et see oleks vähemalt osaliselt automatiseeritud.

Kasu asutusele ja riigile tervikuna

Asutustel jäi hindamisvahendi täitmisel arusaamatuks, kust tekib riigiülene vaade. Seda on vaja asutustele selgemalt kommunikeerida. Juhul kui kõik asutused teevad andmeobjektide hindamise harjutuse täielikult läbi, siis tekib asutustel endil täielik ülevaade oma andmeobjektidest ning nende väärtustest ning siis on võimalik saada ka riigiülene vaade. Selleks oleks asutuste ja projektimeeskonna hinnangul vajalik riigiülese andmearhitekti roll, kes saaks kõikide asutuste andmeobjektide ülevaate ning koondada need kokku riigiüleseks vaateks. Andmearhitekti rolli olemasolu võimaldab riigi üleselt paremini rakendada andmete *once only* printsiipi.