Andmeladude seire kokkuvõte

Andmekaitse Inspektsioon viis perioodil suvi 2021 kuni kevad 2022 riigi valitsussektoris läbi andmeladude seire.

Seire oli ajendatud asjaolust, et andmeladudele pole täna eraldi õiguslikku regulatsiooni. Andmelao (või andmeaida) näol on seega tegu määratlemata õigusmõistega. Sellele vaatamata on inspektsioon seisukohal, et andmeladudele rakendub avaliku teabe seaduse 51. peatükk, mis reguleerib avaliku sektori andmekogude asutamist ja pidamist. Seda seisukohta oleme väljendanud juba varasemalt ka andmekogude juhendis.

Praktikas on aga vähesed andmelaod andmekogudega samaväärselt reguleeritud. Tihti väidetakse, et andmelao mõiste ning pidamise reeglid pole selged. Samuti tuuakse tehnilisi põhjusi, miks andmekogude reegleid järgida ei saa.

Seire eesmärk oli saada ülevaade valitsussektori andmeladudest, selgitada välja probleemid ning nende põhjused. Seire viidi läbi küsimustiku vormis. Saamaks aru, kas asutuse register või analüütika keskkond vastab inspektsiooni andmelao tähendusele, selgitasime ka oma eeldust andmelaoks kvalifitseerumisel. Nimelt näeb inspektsioon andmelaona *andmekogu osalist või täielikku koopiat, mida kasutatakse muul kui turvakoopia/varukoopia eesmärgil ning mis pole ajutine. Andmeladu võib sisaldada ühe või mitme andmekogu või muu infosüsteemi andmeid.*

Seire küsimustele vastas 20 asutust: Justiitsministeerium, Majandus- ja Kommunikatsiooniministeerium, Kaitseministeerium, Välisministeerium, Kultuuriministeerium, Haridus- ja Teadusministeerium, Sotsiaalministeerium, Siseministeerium, Keskkonnaministeerium, Maaeluministeerium, Riigikantselei, Haigekassa, Töötukassa, Patendiamet, Rahapesu Andmebüroo, Statistikaamet, Riigi Infosüsteemi Amet, Haridus- ja Noorteamet, Keskkonnaministeeriumi Infotehnoloogiakeskus, Riigi Tugiteenuste Keskus.

Vastustest tuvastas inspektsioon 28 andmelao tunnustega andmetöötluskeskkonda. Ühe andmekoguga seotud andmeladusid oli 11. Samas oli näiteid, kus andmeladu sisaldas 12, teises suisa 19 andmekogu ning lisaks täiendavalt muudest allikastest pärinevaid andmeid. Valdav enamus sisaldas kas osaliselt või täielikult isikustatud andmeid. Vaid 4 andmelaos isikuandmed puudusid.

Andmeladude loomise eesmärkidena toodi välja, et andmekogude tehniline arhitektuur ja teenusdisain ei võimalda mahukatel päringutel põhinevaid analüüse teha, päringud võivad koormata ja häirida lähteinfosüsteemi toimimist. Nimetati vajadust konsolideerida andmeid mitmest andmekogust. Markantsemate näidetena vastati, et andmeladu võimaldaks andmeid säilitada kauem, kui andmekogus endas. Samuti võimaldab andmeladu tagada parema andmekvaliteedi kui algandmekogus endas.

Seire tõi ka välja laiema probleemi andmevahetusel. Valdav enamus vastanuist ei pea otstarbekaks kasutada andmekogudest andmete laadimiseks andmelattu riigi infosüsteemide andmevahetuskihti X-tee. Ometi näeb avaliku teabe seaduse (AvTS § 43^9 lg 5) andmekogude regulatsioon ette, et andmevahetus andmekogudega ja andmekogude vahel peab käima X-tee kaudu.

Samuti ei andnud vastused kindlust, et andmelaod on infoturbeliselt samaväärselt kaitstud kui andmekogud, millest andmed pärinevad. Kehtiv regulatsioon kohustab andmekogudele

rakendama infoturbestandardit ISKE või ISO/IEC 27001 ning läbi viima perioodilisi auditeid. Paraku ei täideta vastustele tuginedes valdavalt seda kohustust andmeladude puhul.

Seire järeldused:

1. **Andmeladu tuleb asutada seadusega ning tal peab olema põhimäärus. Vaid ühe andmekogu andmetest koosnev andmekogu saab olla reguleeritud lähteandmekogu põhimääruses.**

   Kehtiva õiguse kohaselt peab andmekogu olema asutatud seadusega ning kui seadus ise pole piisavalt täpne, tuleb lisaks kehtestada põhimäärus. Andmeladu vastab olemuselt AvTS § 43^1 lõikes 1 sätestatud andmekogu tunnustele („riigi, KOV või avalik-õigusliku isiku infosüsteemis töödeldavate korrastatud andmete kogum, mida kasutatakse avaliku ülesande täitmiseks“). Seega kohaldub AvTS 5^1 ptk ka andmeladudele. Niisiis tuleb olemasolevad andmelaod kas seadustada või likvideerida.

2. **Andmelao asutamisel tuleb selgelt välja tuua andmelao eesmärk. Peab olema selge, mida tehakse lähteandmekogus, mida andmelaos.**

   Andmeladu olemuslikult on statistilise ja analüütilise funktsiooni jaoks. Tõsi, kehtivas õiguses ei ole piiranguid, et andmeladu ei võiks kasutada primaarfunktsioonideks (st igapäevaseks operatiivtööks). Ent lubamatu on olukord, kus seadus kirjutab üht, kuid praktika käib teist jalga. Nii tulebki andmelao seadustamisel seaduses selgelt välja tuua andmelao kasutamise eesmärgid ning üle vaadata lähteandmekogude kohta seadusesse kirja pandud eesmärgid (nt eemaldades viited analüüsile ja statistikale, kui seda tegelikult tehakse andmelaos).

3. **Tuleb veenduda, et andmelaos toimuv andmetöötlus on seaduslik.**

   Isikuandmete kaitse üldmääruse (IKÜM) artiklite 5 ja 6 kohaselt võib isikuandmeid koguda ning töödelda üksnes selgelt kindlaks määratud eesmärgil. Edasine muul eesmärgil andmete töötlemine on lubatud vaid IKÜM art 6 lõikes 4 sätestatud tingimustel. Seirest selgus, et andmeladudes kombineeritakse mitmete andmekogude andmeid isikustatud kujul ning kasutatakse seda ka konkreetsete inimeste kohta detailandmete saamiseks, sh profileerimiseks. Selline tegevus peab olema seaduslik, st õigusaktiga ette nähtud. Tuleb veenduda, et selliseks andmetöötluseks on seadusega antud ülesanne, pädevusnorm (kellele ülesanne on antud) ning menetlusnorm (mis näeb ette, et sel viisil töödeldakse). Seejuures juhime tähelepanu, et ei haldusmenetluse seadus ega korrakaitseseadus ei näe ette andmekogude andmete massilist kombineerimist eesmärgiga teatud isikuid välja sõeluda.

4. **Andmetöötlus andmelaos peab olema läbipaistev.**

   Andmetöötluse läbipaistvus on IKÜMi üks põhinõuetest. Sellele aitab esmajoones kaasa andmelao piisav reguleerimine seaduse ning seejärel põhimääruse tasemel. Lisaks eelpoolnimetatud andmetöötluse eesmärgile tuleb seaduse tasemel määratleda andmelao andmekoosseis üldistatud tasemel ning andmete säilitamise tähtajad. Põhimääruses tuleb üles loetleda andmelao andmeallikad ning andmelaole antud püsivad otsejuurdepääsud. Tundlikke andmeid (eeskätt eriliiki isikuandmed, kuid ka muud tundlikud andmed) koondava andmelao puhul tuleb andmelaost andmete väljastamine reguleerida seaduses. Need nõuded tulenevad põhiseadusest (põhiõiguste riive ulatus tuleb piiritleda seaduse tasemel). Andmelaos andmete töötlemist ei saa ammu enam vaadelda tegevusena, mida iga

inimene võib mõistlikult ette näha temaga seotud menetluses. Seetõttu kujutab andmete andmelaos isikustatud töötlemine iseenesest juba täiendavat põhiõiguste riivet võrreldes üksnes menetluse tarbeks ja vältel lähteandmekogus andmete talletamisega.

5. **Andmevahetus andmelaoga (nii andmelao - andmekogu kui ka andmelao - muu infosüsteemi vahel) peab toimuma ainult üle X-tee.**

   Erandiks võiks ehk olla andmevahetus andmekogu ja vaid selle andmekogu andmetest koosneva andmelao vahel, eeldusel, et andmelao olemasolu on põhimääruses selgelt välja toodud ning andmelao kaudu pole loodud andmete väljastamiseks mingeid X-tee väliseid tagauksi. X-tee kasutamine on AvTS § 43[9] kohaselt andmekogudele (ja seega ka andmeladudele) kohustuslik. Inspektsioon konsulteeris täiendavalt Riigi Infosüsteemi Ametiga (RIA) X-tee kasutamise tehniliste võimaluste asjus. RIA ei pidanud vastanute poolt välja toodud X-tee mittekasutamise põhjendusi piisavaks, kuna:

   - X-teel saab suuri andmemahte vahetada kas kasutades X-Road REST teenust või X-Road SOAP manuseid. Kui andmemahud on nii suured, et isegi ülikiire internetiga pole võimalik edastada andmestikku mõistliku aja jooksul (mõned minutid), siis on soovitatav tükeldada andmed ning edastada osade kaupa, et välistada teenustes *timeout*'i tekkimist.
   - RIA poolt on kirjeldatud arendussoov X-tee tuumtarkvara arendajale, et tulevikus tekiks X-Road tarkvarale päringute automaatse tükeldamise tugi.
   - Kui alliksüsteem pakub HTTP v1 ühilduvat API-d, siis seda on üldjuhul võimalik muutmata kujul kasutada X-Road Rest teenusena.

6. **Andmelaos peaks võimalusel kasutama anonümiseeritud või vähemalt pseudonümiseeritud andmeid.**

   Pseudonüümiseeritud andmete lähteandmekogust uuendamiseks vajaliku sidumise saab teha andmete laadimisprotsessis. Depseudonümiseerimine peab olema rangelt ja selgelt õiguslikult reguleeritud. Kui andmeladu primaarfunktsioonideks (st igapäevaseks operatiivtööks) ei kasutata, ei ole detailandmeid (isiku tuvastamist võimaldavaid andmeid) vaja andmelattu kanda.

7. **Andmelao testimiseks tuleks kasutada sünteetilisi andmeid või hägustatud (obfuskeeritud) andmeid.**

8. **Andmelaol võivad olla asutusevälised otsejuurdepääsud ning X-tee teenused <u>vaid juhul</u>, kui selle pidamisel järgitakse kõiki andmekogudele kehtivaid nõudeid**: on nõuetekohaselt seadusega asutatud, andmeväljastus reguleeritud ning kasutatakse X-teed.

   Asjaolu, et andmeladu asub konkreetse ministeeriumi valitsemisalas, ei anna ministeeriumile õigust saada otsejuurdepääsu andmelaole. Andmelaole juurdepääsu andmisel tuleb lähtuda kõikidest andmekogu kohta kehtivatest juurdepääsureeglitest – juurdepääsu saamiseks peab olema õiguslik alus.

9. **Andmekvaliteeti tuleb parandada lähteandmekogus, mitte üksnes andmelaos.**

10. **Andmelaole kehtib ISKE (tulevikus e-ITSi) rakendamise, sh auditeerimise kohustus (AvTS § 43[9] ja KüTS § 9).**

    Andmelaole peab olema määratud turvaklass. Seejuures ei saa näiteks konfidentsiaalsuse turvaosaklass (S) olla madalam kui lähteandmekogus. Paljudest andmekogudest andmete

kokku koondamisel tekib selgelt suurem risk, mis vajab ka tõhusamaid kaitsemeetmeid. Ühe andmekogu koopiat sisaldava andmelao puhul on võimalik, et andmeladu auditeeritakse koos lähteandmekoguga. Kui aga andmeladu sisaldab mitme erineva lähteandmekogu koopiat või andmekogude väliseid muid andmestikke (nt asutuse sisestest töökorralduslikest rakendustest API-de kaudu), pole võimalik, et „andmeladu on auditeeritud koos lähterandmekoguga“.

11. Andmelao päringud tuleb kasutaja tasandil logida.

Logimine andmelaos ei või olla tehtud nii, et ärianalüütika (BI) kiht on üks üldine kasutaja ja pöördub andmebaasi poole. Sel juhul pole logist näha, kes konkreetselt päringu tegi. Andmelao andmebaasi tehtud päringud peavad olema kasutajapõhiselt logitud.

12. Andmeladu ei või minna vastuollu riigi infosüsteemi hajusa arhitektuuriga ega kaasa tuua julgeolekuriski riigile.

Eesti riigi infosüsteemi arhitektuur on üles ehitatud hajusalt, st on eraldi andmekogud, mis suhtlevad omavahel turvalise kanali, X-tee, kaudu. Paljude seni eraldatud andmekogude andmestike ühte kokku koondamine viib selleni, et riigi infosüsteem pole enam piisavalt hajus. Võivad tekkida superandmebaasid, mis on atraktiivsem sihtmärk küberrünnakutele ning mille lekke tagajärjed on hukatuslikumad. Superandmebaaside loomisega kaasneb suurem julgeolekurisk riigile. Lahendus võiks olla loogiline andmeladu, kus ei koondata andmeid ühte kokku, vaid lahendatakse andmepäringute kihina (virtuaalne andmeladu).