EST-Health-30 andmekaitsealane mõjuhinnang

Käesolev andmekaitseline mõjuhinnang on tehtud põhjusel, et Euroopa isikuandmete kaitse üldmääruse (2016/679, edaspidi GDPR) artikkel 35 lõiked 1 ja 3 nõuab enne ulatuslikku eriliiki isikuandmete (terviseandmete) töötlemist kavandatavate isikuandmete töötlemise toimingute mõju hindamist isikuandmete kaitsele.

Kokkuvõte andmekaitsealase mõjuhinnangu tulemustest

Käesolev andmekaitsealane mõjuhinnang on läbi viidud uurimisprojekti „EST-Health-30 - Eesti terviseandmete väärindamine“ kohta. Teadusuuringu üldeesmärk on Eesti terviseandmete väärindamine läbi andmeteaduse meetodite, et pakkuda kvaliteetset tõendust tõhusate patsiendikesksete tervishoiu- ja ennetusteenuste osutamiseks. Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest (andmesubjektid).

Andmekaitsealane mõjuhinnang on läbi viidud ajavahemikus 01.08.2024 - 17.08.2024 ning kehtib kuni uurimisprojektis kasutatud andmete kustutamiseni hiljemalt 31.12.2028. Juhul, kui muutuvad mõjuhinnangu aluseks olnud asjaolud, siis mõjuhinnangut täiendatakse vastavalt.

Kokkuvõttes leiavad mõjuhinnangu koostajad, et uurimisprojektis on kasutusel tõhusad riskimaandamise mehhanismid ja rakendatud meetmete tulemusena ei teki andmesubjekti õigustele suurt ohtu isikuandmete kaitse üldmääruse (2016/679/EU) artikkel 35 lõike 1 tähenduses.

Mõjuhinnang on kooskõlastatud Tartu Ülikooli andmekaitse peaspetsialistiga.

Sissejuhatus

Tartu Ülikool (TÜ) on avalik-õiguslik ning vanim ja suurim Eesti ülikool. Tartu Ülikooli üheks peamiseks eesmärgiks on edendada teadust, kuid spetsiifilisemalt ka edendada Eestit ja tema rahvast uurivaid teadusi ning eestikeelset haridust. Tartu Ülikooli neljast valdkonnast on konkreetselt meditsiinile keskendunud meditsiiniteaduste valdkond. Samas on tervisevaldkond muutumas järjest interdistsiplinaarsemaks ning meditsiiniga on tihedalt seotud ka loodusteaduste valdkond - näiteks genoomika instituut uurib geneetika ja haiguste vahelisi seoseid, arvutiteaduse instituut aga arendab nii geneetilistel kui terviseandmetel rakendamiseks andmeteaduse meetodeid, s.h masinõpet.

Käesolev mõjuhinnang käsitleb EST-Health-30 alusandmestikku, mis luuakse Tervisekassa andmekogu, retseptikeskuse andmekogu, Tervise Infosüsteemi, surma põhjuste registri ja vähiregistri andmete alusel. Alusandmestiku loomist ja sellel põhinevate teadus- ja rakendusuuringute teostamist uurimistetoodikate arendamiseks Eesti terviseandmetel kavandab arvutiteaduse instituudi terviseinformaatika uurimisgrupp (vt taotlus p3 “Vastutavad uurijad” ja p4 “Uuringu läbiviijad”) (edaspidi Uurimisgrupp). Uurimisgrupp on Eesti terviseandmekogudest pärit pseudonüümitud andmestikel viinud varasemalt läbi mitmeid erinevaid teadusuuringuid. Seekordne uurimisprojekt erineb eelmistest uurimisprojektidest selle poolest, et on eelnevatest laiapõhjalisem, kaasates senisest rohkem patsiente ja kaasaegsemaid andmeid. Projekt koosneb kahest etapist, millel on järgmised eesmärgid:

1. etapp: alusandmestiku loomine ja andmekvaliteedi tõstmise meetodite arendamine

1.1. Hinnata valimi põhjal Eesti terviseandmete kvaliteeti ja hõivet võrreldes erinevatest andmeallikatest pärinevat informatsiooni.

1.2. Andmekvaliteedi ja andmehõive tõstmise meetodite sh tehisintellekti-, tekstikaeve- ja imputatsioonimeetodite arendamine.

1.3. Tervisesündmuste esituse ühtlustamine rahvusvahelistele standarditele üle aja ja andmeallikate, sh kasutades tekstikaeve, klasterdamise ja automaattõlke meetodeid.

1.4. Arendatavad meetodid on hiljem rakendatavad ka uutes uuringutes, suurematel andmehulkadel ning riiklikes infosüsteemides andmekvaliteedi ja -hõive parandamiseks.

1.5. Standardiseeritud ja ühtlustatud andmestiku loomine ja kirjeldamine teadusuuringute läbiviimiseks. Andmestikku kirjeldatakse läbi erinevate numbriliste näitajate ja teostatavusanalüüside senisest suuremal detailsusastmel.

2. etapp: teadus- ja rakendusuuringute teostamine uurimistetoodikate arendamiseks Eesti terviseandmetel.

2.1. Arendada haigustrajektooride ja raviteekondade analüüsi metoodikaid kasutades sh tehisintellekti meetodeid, et kirjeldada praeguseid ravipraktikaid, võrrelda neid raviteekondade ja -juhenditega ning modelleerida tervisetulemeid ja ravi majanduslikku mõju.

2.2. Analüüsida erinevate faktorite (nt sugu, vanus, erinevad haigused) mõju ravijärgimusele, hinnata ravijärgimuse mõju erinevatele tervisetulemitele ja luua personaalseid ennustusmudeleid, mis võimaldavad maandada ravi mittejärgimise riske.

2.3. Arendada personaliseeritud ennetusmudeleid, mis võimaldavad vaatlusandmete põhjal tuvastada praeguste haiguse ennetuste ja ravipraktikate kitsaskohti, planeerida senisest paremini haiguste ennetustegevusi ja hinnata pakutud ennetusteenuste majanduslikku mõju

Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest (andmesubjektidest). Nende andmesubjektide kohta saadakse pseudonüümitud andmed Tervisekassa andmekogust, retseptikeskuse andmekogust, Tervise Infosüsteemist, surma põhjuste registrist ja vähiregistrist. Pseudonüümi alusel viib Uurimisgrupp sama isiku andmed kokku ning töötleb neid.

Uurimisprojekt kestab 01.10.2024-31.12.2027. Peale uurimisprojekti lõppu andmed arhiveeritakse üheks aastaks. Arhiveerimisperioodi lõpus, hiljemalt 31.12.2028 andmed kustutatakse.

Mõjuhinnangu läbiviimisest

Läbiviimise aeg

Käesolev andmekaitsealane mõjuhinnang on läbi viidud ajavahemikul 01.08.2024 - 17.08.2024.

Mõjuhinnangu ulatus

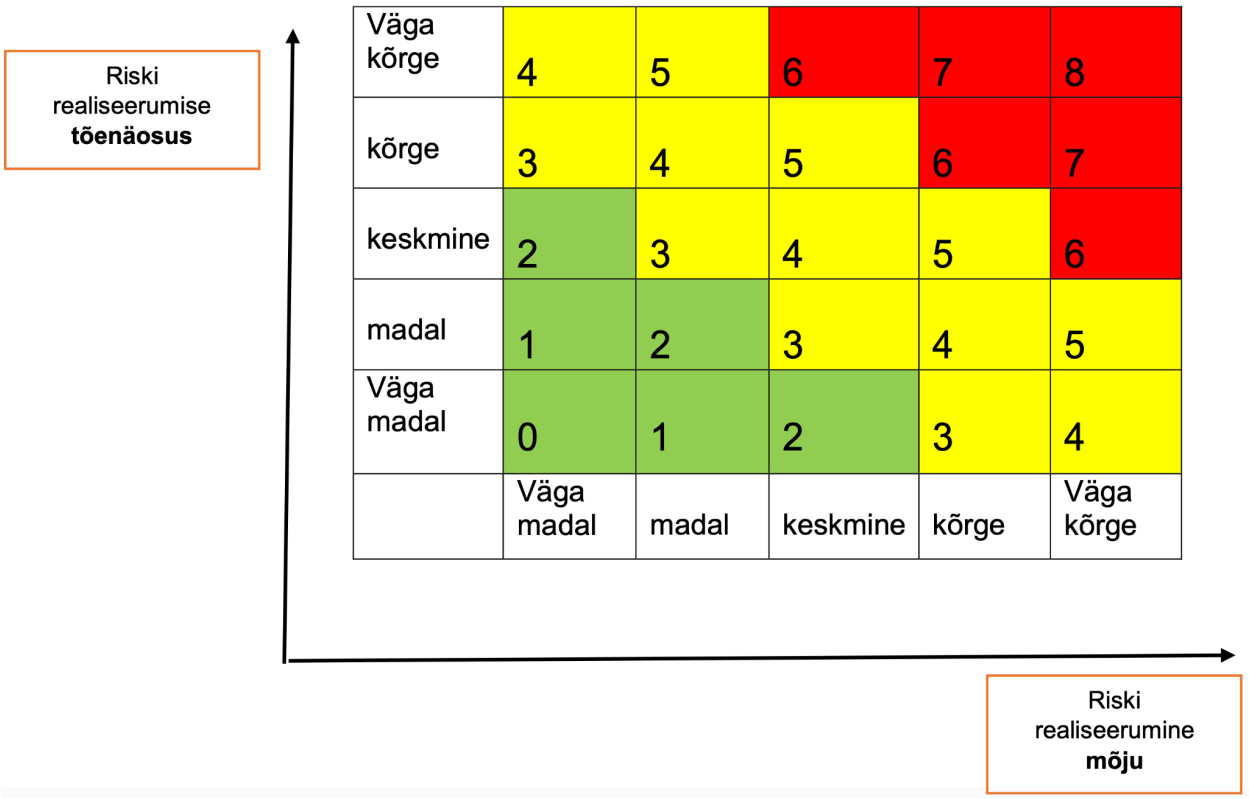
Käesolev andmekaitsealane mõjuhinnang on koostatud andmestiku EST-Health-30 elutsükli kohta Tartu Ülikoolis s.o. Andmete aktiivse kasutamise perioodi kohta oktoober 2024-31.12.2027 ja arhiveerimisperioodi kohta 01.01.2028-31.12.2028.

Metoodika

Käesoleva andmekaitsealase mõjuhinnangu loomisel on kasutatud riskide identifitseerimisel ja hindamisel põhinevat metoodikat. Mõjude hindamisel võetakse arvesse kirjeldatud andmetöötlemise iseloomu, ulatust ja konteksti. Riskid on leitud ja kirjeldatud andmetöötluse ohustsenaariumite hindamisest lähtuvalt. Riskide hindamisel on hinnatud skaalal (0-väga madal; 4-väga kõrge) kahte tegurit:

- riski realiseerumise tõenäosus;
- riski realiseerumise mõju.

Üldine riski tase leitakse kahe skaala ristumispunktis vastavalt alltoodud joonisele ja tabelile. Näiteks kui tõenäosus on 3 ja mõju 2, siis riski tase on 5 ehk keskmine.



Riski tase	
Tulemus	Kirjeldus
6-8	Kõrge
3-5	Keskmine
0-2	Madal

Projektipõhise infosüsteemi kirjeldus

EST-Health-30 andmestikku hoitakse ja kõik teisendused/analüüsid teostatakse Eesti Teadusarvutuste infrastruktuuri poolt pakutavas sensitiivsete andmete privaatses uurimiskeskkonnas (SAPU). Projekti käigus kasutatakse eraldiseisvat SAPU keskkonda, mis ei ole seotud teiste SAPU keskkondadega. Detailne info Eesti Teadisarvutuste infrastruktuuri kohta on leitav aadressil https://etais.ee/ ja lisainfo sensitiivsete andmete privaatse uurimiskeskkonna (SAPU) koht on leitav aadressil https://docs.hpc.ut.ee/public/services/SAPU/ (inglise keeles).

Kasutajate haldus

Projekti käigus kasutatavasse SAPU keskkonda (andmepuur) luuakse spetsiaalsed ja eraldiseisvad kasutajakontod ainult projekti vastutava täitja taotlusel ja heakskiidul. Igale uuringu täitjale antakse ligipääs ainult tööks vajalikele andmetabelitele. SAPU keskkond on

selleks volitatud isikutele kättesaadav ainult aktiivse analüüsi faasis ning muul ajal on keskkond välja lülitatud ning sinna ei ole võimalik siseneda ka kasutajakonto olemasolu korral. Kolmandatel isikutel (kaasa arvatud Tartu Ülikooli teistel töötajatel) puudub juurdepääs kasutatavasse SAPU keskkonda.

SAPU keskkonnas eristatakse nelja erinevat kasutajarolli:

- *Cloud operator* (administraatori õigused) - SAPU tehniline administraator, seda rolli täidab Eesti Teadusarvutuste infrastruktuur - hoolitseb turvalisuse, monitoorimise ülesannete eest, vastutab, et server töötab. Ei kasuta ja ei vaata andmeid.
- *Data owner* (otsene ligipääs masinale ja monitoorimisele) - vastutav uurija, kellele väljastatakse andmed ja kes toob andmed SAPU masinasse. Kõik andmetele ligipääsude andmised ning andmete väljaliigutamise SAPUst toimuvad vaid data owneri ehk vastutava uurija kinnitusel. Data owner ehk vastutav uurija vastutab ka logide jälgimise ja säilitamise eest kogu andmete eluea jooksul.
- *Data custodian* (otsene ligipääs masinale ja monitoorimisele) - uuringu täitja, kes tegeleb peaasjalikult tehniliste küsimustega, sh toetab *data owner*'i andmete SAPUsse viimisega, monitoorib logisid ja vajadusel aitab pilve operaatorit.
- *Data analyst* (ligipääs üksnes läbi virtuaalse töölaua) - uuringu täitja, kes analüüsib ja töötleb andmeid.

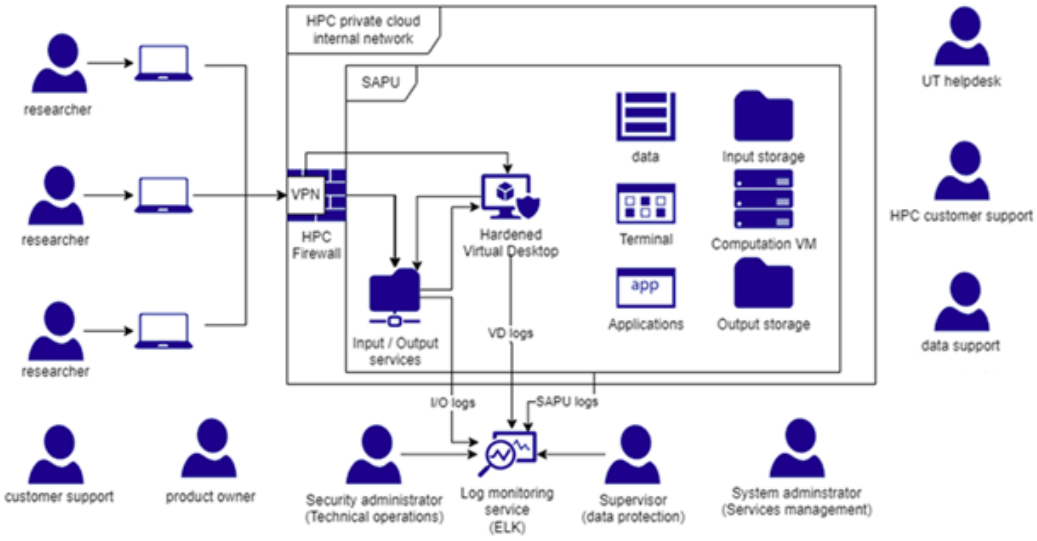
Teadaolevad turvameetmed

- Andmete edastamiseks sõlmitakse leping iga andmeandja ja Tartu Ülikooli vahel. Andmete edastus toimub krüpteeritult läbi turvalise andmevahetusserveri, mis on üles seatud Eesti Teadusarvutuste infrastruktuuri poolt või kasutades andmeallikate tavapäraseid andmete väljastamise viise vastavalt nende sisemistele protseduurireeglitele (enamasti konkreetsele juhtivuurijale parooliga ligipääsetav kataloog andmeallika serveris, andmefail on krüpteeritud).
- Analüüsiks kasutatakse Eesti Teadusarvutuste infrastruktuuri poolt pakutavat sensitiivsete andmete privaatset uurimiskeskkonda (SAPU) millele on piiratud ligipääs vaid eetikaloal loetletud uurijatele.
- Kasutusel on füüsilise ja infotehnoloogilised turvameetmed, mida on täpsemalt kirjeldatud mõjuhinnangu lõpus.

Süsteemi kasutusotstarve

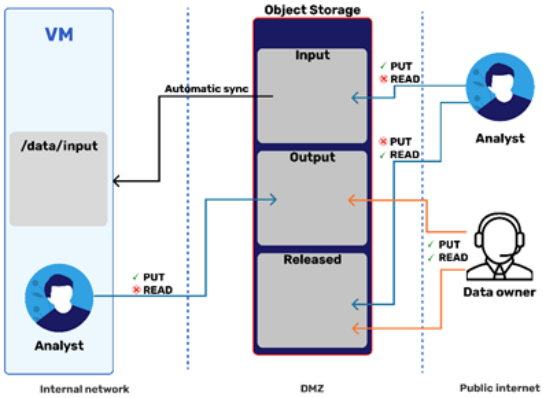
Sensitiivsete andmete privaatne uurimiskeskkond (SAPU) on Eesti Teadusarvutuste infrastruktuuri poolt spetsiaalselt loodud ja pakutav andmetöötluskeskkond, kus analüütikud saavad töötada tundlike andmete kallal, vähendades võimalikku andmete volitamata kopeerimist, ülekandmist või masinatest välja võtmist, pakkudes kõrgemat turvaklassi kui tavaline suure jõudlusega arvutusklaster.

SAPU kõrgetasemeline arhitektuur



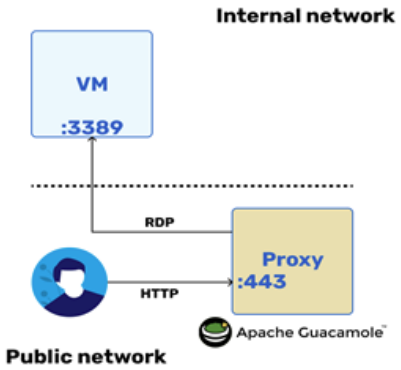
SAPU andmevärav

Kuna analüütikud vajavad võimalust SAPU masinasse viia andmeid, skripte ja muud teavet ning samuti on vajalik SAPU masinast analüüsitulemusi, siis on kasutusele võetud S3 Object Storage põhised eeskirjad kolme kaustaga:



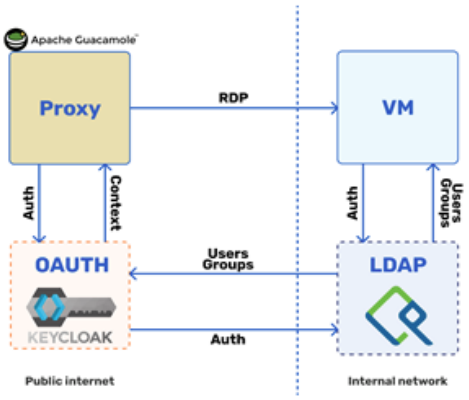
Graafiline kasutajaliides

RDP või masina avalikustamine internetis kätkeb endas mitmeid infoturbe riske ja nende maandamiseks on kasutusele võetud puhverserver. Kasutatakse avatud lähtekoodiga tehnoloogiat Apache Guacamole.



Autentimine ja autoriseerimine

SAPU'l on eraldiseisev LDAP server, millega luuakse ühendus identiteedi ja juurdepääsu haldamiseks.



Andmestiku elutsükkel

Andmestiku elutsükkel on järgmine:

1. Eesti Tervisekassa, Tervise Arengu Instituut (TAI) ja Tervise ja Heaolu Infosüsteemide Keskus (TEHIK) edastavad oma infosüsteemidest lisas "Andmekoosseis" kirjeldatud andmed Tartu Ülikooli vastutavale uurijale.
2. Uurimisgrupi vastutav uurija korraldab andmete tõstmise SAPU-sse.
3. Uurimisgrupi liikmed töötlevad SAPU-s olevaid andmeid ainult teadusuuringu eesmärkide täitmiseks.
4. Alusandmeid uuendatakse regulaarselt, taotluses toodud sagedusega korrates samme 1-3.
5. Vajadusel toimub tulemuste väljastamine SAPU'st. Tulemuste väljastamine toimub ainult vastutava uurija kinnitusel ning pidades silmas anonüümsustaset $k \geq 5$.
6. Uuringu lõppedes SAPU koos andmetega arhiveeritakse üheks aastaks ning vajadusel tagatakse ligipääs andmetele. Peale arhiveerimisperioodi lõppu andmed kustutatakse. Kustutatakse ka kõik SAPU ja andmete varukoopiad.

Isikuandmete töötlemise toimingud

Isikuandmete kogumine

Käesolevas uuringus uusi isikuandmeid ei koguta ja andmesubjektidega ühendust ei võeta. Kasutatakse juba olemasolevaid terviseandmeid ainult taotluses kirjeldatud teadusuuringu eesmärkide täitmiseks. Vastavalt isikuandmete kaitse seadusele § 6 lg 1 võib teadusuuringu vajaduseks isikuandmeid töödelda andmesubjekti nõusolekuta, kui need on pseudonüümitud. Andmed pseudonüümitakse käesolevas uuringus enne Uurimisgrupini jõudmist andmeallikate poolt.

Vaatamata sellele, et andmed on pseudonüümitud ega sisalda isiku otsest tuvastamist võimaldavaid tunnuseid, käsitletakse uuringu andmeid järgmistes peatükkides tulenevalt nende tundlikust iseloomust kui eriliiki isikuandmeid.

Isikuandmete pseudonüümimine ja edastamine

Valimi moodustamine ja pseudonüümide moodustamine

Andmestiku aluseks on isikukoodide põhjal koostatud juhuvalim. Juhuvalimi moodustamise aluseks on isikute isikukoodid, täpsemalt nende põhjal loodud pseudonüümid. Pseudonüümide salajaseks, kuid ühetaoliseks moodustamiseks kasutavad kõik andmeandjad sama räsifunktsiooni ja salajast parooli ehk räsiparooli. Viimast Uurimisgrupile ei väljastata.

Salajase pseudonüümimisparooli moodustamine ja andmeallikatega jagamine

1. Tervisekassa määrab salajase parooli (siin näidetes edaspidi *“mypassword”*)
2. Tervisekassa salvestab parooli tekstifaili, mille sisu on umbes järgmine:
   ```
   Pseudonüümimiseks kasutatav parool on: mypassword
   Kontroll - järgmine käsk:
   echo -n "99999999999SIIAÕIGEPAROOL" | openssl dgst -sha256 | awk '{print $2}'
   peab andma tulemuseks
   a40a173b33e0c3913e3bdb7a7e8878ad9b52925541631f96be5143ac32eb68f7
   ```
3. Tervisekassa krüpteerib tekstifaili digidoc konteineris, määrates adressaatideks konkreetsed andmeväljastuse eest vastutavad isikud nii TEHIK-s kui TAI-s. See tagab, et parooli ei näe Uurimisgrupi liikmed TÜ-s.
4. Tervisekassa edastab krüpteeritud digidoc konteineri TÜ Uurimisgrupile, kes edastab selle omakorda TEHIK-le ja TAI-le. Selline korraldus tagab, et Tervisekassa ei pea ise korraldama ega sekkuma TEHIK ega TAI poolsesse andmeväljastusse.

Pseudonüümimine ja valimi moodustamine

Protsess valimi ja pseudonüümide moodustamiseks iga andmeandja juures on järgmine:

1. Andmeandja veendub, et pseudonüümimiseks kasutama hakatav räsifunktsioon töötab õigesti. Järgmine käsk:
   *echo -n "*99999999999*mypassword" | openssl dgst -sha256 | awk '{print $2}'*
   peab andma tulemuseks
   a40a173b33e0c3913e3bdb7a7e8878ad9b52925541631f96be5143ac32eb68f7
   Lisaks kontrollib andmeandja, et Tervisekassast saadud parool annab räsimisel õige tulemuse.
2. Tervisekassa moodustab kõigile isikukoodidele vastavad räsid, andes räsifunktsioonile ette ühendatud sõnena nii isikukoodi kui salajase parooli. Näiteks isikukoodi 99999999999 korral moodustatakse räsi ülaltoodud käsuga ja selle näite põhjal saadav räsi on
   a40a173b33e0c3913e3bdb7a7e8878ad9b52925541631f96be5143ac32eb68f7
3. Räsi iga sümbol on üks kuueteistkümnesümbolilisest hulgast: a-f või 0-9. Need jaotuvad ühtlaselt. **Valimisse kuuluvad isikud, kelle räsi esimene sümbol kuulub hulka {a,b,c,d,e,f,0} ja teine sümbol hulka {a,b,c,d,e,f,0,1,2,3,4}**. Sellisel juhul satub valimisse (7/16)*(11/16)*100=30.08% isikukoode. *Näide: ka ülaltoodud näidisisikukood 99999999999 satub valimisse, sest tema räsi esimene sümbol “a” kuulub hulka {a,b,c,d,e,f,0} ja teine sümbol “4” kuulub hulka {a,b,c,d,e,f,0,1,2,3,4}.*
4. Nendest isikukoodidest kuuluvad käesoleva andmestiku valimisse isikud, kelle kohta on vastava andmeandja andmebaasis sisestatud kandeid uuringuperioodil. Pseudonüümidena kasutatakse loodud räsisid.
5. Andmeandja säilitab parooli turvaliselt tuleviku andmeuuenduste väljastamiseks.

Antud lahenduse puhul ei ole tarvis moodustada valimit ühe andmeandja juures ning seda siis teiste andmeandjatega jagada. Puudub ka vajadus vahetada isikukood-pseudonüümide tabelit. Andmeandjad vahetavad omavahel üksnes räsifunktsioonis kasutatavat parooli, kasutades selleks turvalist krüpteeritud kanalit, näiteks vastuvõtja isikukoodile kodeeritud *.CDOC* konteinerit. Kuna pseudonüümimine ja valim moodustatakse ainult isikukoodi põhjal, satuvad ka uued isikud pseudonüümi sobivuse korral automaatselt valimisse.

Pseudonüümide moodustamiseks etteantud isikukoodide faili põhjal ja nende seast valimi määramiseks saab kasutada järgmist skripti:

```
#!/bin/bash

PASSWORD='mypassword' #siia panna ainult andmeallikatele teadaolev parool

while read -r line; do
  hash=$(echo -n "$line$PASSWORD" | openssl dgst -sha256 | awk '{print $2}')
  if [[ $hash =~ ^[abcdef0][abcdef01234] ]]; then
    echo "$line" >> valimisse_kuuluvad_isikukoodid.txt
    echo "$line $hash" >> valimisse_kuuluvate_isikukoodide_pseudonyymid.txt
  fi
done < isikukoodid.txt
```

Andmete töötlus enne Uurimisgrupile väljastamist

Andmeandjad väljastavad üksnes valimisse kuuluvate isikute andmeid.

Nende isikute kohta väljastatakse uurimisgrupile andmed järgmistest andmekogudest:

1. Tervisekassa andmekogu (raviarved ja kindlustuskaitse andmed, vastutav töötleja: Tervisekassa)
2. Retseptikeskuse andmekogu (vastutav töötleja: Tervisekassa)
3. Tervise Infosüsteem (vastutav töötleja: Sotsiaalministeerium, volitatud töötleja: TEHIK)
4. Surma põhjuste register (vastutav töötleja: TAI)
5. Eesti vähiregister (vastutav töötleja: TAI)

Täpne andmekoosseis kõigi andmekogude lõikes on kirjeldatud uuringutaotluse juurde kuuluvas lisas. Muuhulgas eemaldatakse otsest isikutuvastamist võimaldavad andmed ning asendatakse pseudonüümiga.

Varasemaid andmeid kui 1. jaanuar 2012. a ei väljastata.

Andmete üleandmine Uurimisgrupile

Andmete edastamiseks sõlmitakse leping iga andmeandja ja Tartu Ülikooli vahel. Andmete edastus toimub krüpteeritult läbi turvalise andmevahetus serveri, mis on üles seatud Eesti Teadusarvutuste infrastruktuuri poolt või kasutades andmeallikate tavapäraseid andmete väljastamise viise vastavalt nende sisemistele protseduurireeglitele (enamasti konkreetsele juhtivuurijale parooliga ligipääsetav kataloog andmeallika serveris koos krüpteeritud andmefailiga).

Andmete uuendamine

Andmestiku on plaanis regulaarselt uuendada. Täienevad olemasolevate andmesubjektide andmed, aga valimisse satub ka uusi andmesubjekte (näiteks isikud, kellel varem puudus isikukood). Protsess uuendamiseks on järgmine:

1. Uurimisgrupp algatab regulaarselt andmete uuendamise protsessi, pöördudes selleks kõigi andmeandjate poole ja täpsustades, millisest ajahetkest alates ja millise ajahetkeni toimunud uuendusi on tarvis.
2. Kõik andmeandjad teostavad vajaliku andmete väljavõtte, pseudonüümimise ja edastavad andmed Uurimisgrupile sarnaselt algsele väljastamisele. Kuna kasutatakse täpselt sama pseudonüümimisalgoritmi ja räsiparooli, saavad valimisse juba varem kuulunud isikud sama pseudonüümi, mis varasemates väljastustes ning uutele luuakse uus pseudonüüm.

Andmete säilitamine

Andmeandjatelt saadud algandmed paigaldatakse SAPUs spetsiaalselt kirjutuskaitstud kausta, et vältida andmete juhuslikku hävimist või kahjustumist. Esimese tegevusena kasutatakse andmetel TÜ poolt arendatud anonüümimisrakendust, mis tuvastab ja asendab algandmete vabatekstilistes dokumendiosades isikunimed, aadressid, telefoninumbrid, isikukoodid, kui neid seal peaks leiduma. Edasine töötlus toimub ainult täiendava anonüümimisprotsessi läbinud andmetel ning algselt saadud andmetele pääsevad ligi üksnes vastutavad uurijad ja *data custodian* rollis (vt allpool) olevad uuringu läbiviijad, et anonümiseerimisrakendust vajadusel uuesti kasutada.

Järgmiseks luuakse automaatskriptidega kvaliteetne terviseandmestik samas serveris paiknevasse PostgreSQL andmebaasi. Selle andmestiku kahjustumine ei ole kriitiline, sest vajadusel on võimalik automaatskriptidega see algandmetest uuesti taastada.

SAPU'st tehakse regulaarselt krüpteeritud varukoopiaid Eesti Teadusarvutuste infrastruktuuri lindirobotile, mis asub füüsiliselt teises asukohas (andmekeskuses).

Säilitamise tähtajad

Andmeid säilitatakse kuni projekti eesmärkide täitmiseni, kuid maksimaalselt kuni uuringuprojekti arhiveerimisperioodi lõpuni: 31.12.2028. Kui uuringu käigus tekib põhjendatud vajadus andmeid säilitada kauem, esitatakse enne uuringu lõppu Eesti bioeetika ja inimõiguste nõukogule (EBIN) vastavasisuline taotlus.

Hoiustamisel kasutatavad turvameetmed

Eesti Teadusarvutuste infrastruktuur järgib infrastruktuuri haldamisel ISKE M taseme nõudeid. Andmeid ja vaheandmeid hävitatakse vajadusel vastavalt ISKE H turbeastmega andmete hävitamise nõuetele. Eesti Teadusarvutuste infrastruktuur käsitleb kõiki teenuse pakkumise käigus teatavaks saavaid/käsitletavaid andmeid konfidentsiaalsetena.

Seoses 2022. aasta lõpus kehtima hakanud „Võrgu- ja infosüsteemide küberturvalisuse nõuded“ määruse ja „Eesti infoturbestandard“ määrusega on Tartu Ülikool ja ka Eesti Teadusarvutuste infrastruktuur kohustatud järgima Eesti Infoturbestandardit ning regulaarselt läbi viima Eesti infoturbestandardi järgimise auditeid. Tartu Ülikool peab esimese auditi läbima vähemalt 3 aasta jooksul (ehk enne 2025. a lõppu) ning käesoleval hetkel selle nimel ka tegutsetakse.

Kuidas välditakse andmete juhuslikku hävimist või kahjustumist?

- Andmetest tehakse regulaarseid krüpteeritud varukoopiaid.
- Regulaarselt teostatakse serverite turvatestimist, uuendamist ja monitoorimist.
- Serverid on varustatud katkematu toiteallikaga (UPS).

- Kasutusel on füüsilised ja organisatoorsed turvameetmed, mis takistavad selleks volitamata isikute füüsilist juurdepääsu serveritele.
- Kasutajal puudub SAPU keskkonnast juurdepääs internetile.
- Kasutajal puudub SAPU keskkonnas õigus installeerida programme.
- SAPU keskkonda saavad siseneda ainult selleks vastavaid õigusi omavad kasutajad.
- Kasutajaid juhendatakse ja koolitatakse SAPU keskkonna kasutamise osas.

Logid, turvakoopiad

Andmete töötlus toimub SAPU keskkonnas, mis rangelt piirab kasutajate võimalust andmeid serverist välja kopeerida ning võimaldab vastutavatel uurijatel jälgida kõikide kasutajate tegevusi. Vastutavad uurijad tagavad, et logitakse kõikide kasutajate kõik tegevused SAPU keskkonnas, sh salvestatakse jooksvalt ka kasutaja ekraanipilti (video). Kasutajad on logidest ja ekraanipildi salvestamisest teadlikud ning see toimib ka heidutava meetmena. Logide monitoorimise korraldavad vastutavad uurijad. Samuti tagavad vastutavad uurijad vajadusel andmeandjatele ligipääsu kasutajate logidele kogu andmete eluea vältel.

Andmetest tehakse üks kord nädalas regulaarseid varukoopiad vastavalt Eesti Teadusarvutuste infrastruktuuri varundamise korrale. Varundamine toimub Eesti Teadusarvutuste infrastruktuuri lindirobotile, mis asub füüsiliselt teises asukohas (andmekeskuses). Varukoopiad on spetsiaalselt krüpteeritud, alles hoitakse viimast kolme varukoopiat. Varukoopiate tegemisel kasutatakse inkrementaalset varukoopiat ja on tagatud, et eelnevaid varukoopiaid ei ole võimalik muuta.

Arhiveerimine

Andmed arhiveeritakse üheks aastaks peale projekti lõppemist Eesti Teadusarvutuste infrastruktuuri poolt krüpteeritud kujul SAPU keskkonnas. Andmetele võimaldatakse ligipääs vastutava uurija kinnitusel vaid põhjendatud juhtudel, näiteks retsenseerimisele saadetud teadusartiklite retsensioonidele vastamiseks. Kõik arhiveeritud andmetega tehtud toimingud logitakse kasutaja tasandil. Andmed hävitatakse hiljemalt 31. detsembriks 2028 kasutades hävitamise ajahetkel parimat kasutuselolevat praktikat vastavalt rakendatavale infoturbestandardile (IS0/IEC 27001). Andmete hävitamine dokumenteerikse hävitamisaktis.

Andmete kasutamine

Andmeid kasutatakse üksnes Uurimisgrupi poolt ainult SAPU serveris ja üksnes andmestiku loomiseks või taotluses kirjeldatud teadusuuringute läbiviimiseks.

Uurimisgrupp

Uurimisgrupi koosseis on käesoleva mõjuhinnangu koostamise hetkel järgmine:

Uuringu läbiviija nimi	Roll projektis	Roll SAPU serveris

Jaak Vilo	Vastutav uurija, professor	Data owner
Raivo Kolde	Vastutav uurija, kaasprofessor	Data owner
Sven Laur	Vastutav uurija, kaasprofessor	Data owner
Sulev Reisberg	Vastutav uurija, teadur	Data owner
Marek Oja	Teadur	Data custodian
Kerli Mooses	Teadur	Data analyst
Taavi Tillmann	Kaasprofessor	Data analyst
Markus Haug	Nooremteadur	Data analyst
Harry-Anton Talvik	Nooremteadur	Data custodian
Hendrik Šuvalov	Nooremteadur	Data analyst
Kunnar Kukk	Nooremteadur	Data analyst
Õie Renata Siimon	Nooremteadur	Data analyst
Maarja Pajusalu	Nooremteadur	Data analyst
Maria Malk	Nooremteadur	Data analyst
Anton Vykhovanets	Nooremteadur	Data analyst
Nikita Umov	Nooremteadur	Data analyst
Laura Lõo	Nooremteadur	Data analyst
Kermo Saarse	Nooremteadur	Data analyst
Sirli Tamm	Andmekvaliteedi spetsialist, programmeerija	Data custodian
Kaire Koljal	Andmekvaliteedi spetsialist, programmeerija	Data analyst
Ami Sild	Terviseandmete insener-analüütik	Data custodian
Helene Loorents	Programmeerija	Data analyst
Neeme Ilves	Spetsialist	Data analyst

Sander Kütisaar (Eesti Teadusarvutuste infrastruktuur)	SAPU pilve operaator	Cloud operator

Uurimisgrupi koosseis ja rollid võivad projekti käigus vastutavate uurijate kinnitusel täieneda. Isikkoosseisu muutmiseks esitatakse EBIN-le vastavasisuline jätkutaotlus. Igale Uurimisgrupi liikmele antakse ligipääs ainult töö eesmärkide täitmiseks vajalikele andmetabelitele. Kui loodud andmestiku kasutamiseks rahuldatakse täiendavadi taotlusi, siis käesoleva taotluse vastutavad uurijad korraldavad ligipääsu ka uute uuringute läbiviijatele lähtudes andmete minimaalsuse printsiibist ja tagavad ka nende logide salvestamise ja monitoorimise.

Andmete väljastamine SAPU serverist saab toimuda üksnes *data owner* rollis uurija kinnitusel ja see juhtub üldjuhul uurimistulemuste avaldamisel nt teaduspublikatsioonis. Väljastuse heakskiitmisel kontrollib vastutav uurija, et väljastatavad andmed on kooskõlas uuringu taotluses kirjeldatud eesmärkide ja sisuga. Uurimisgrupp avaldab üksnes statistilisi tulemusi, kus on tagatud $k \geq 5$ anonüümsus. Uurimisgrupp ei avalda ega jaga kolmandate osapooltega üksikpatsientide andmeid ega pseudonüüme. Kõikide avaldatavate andmestike kohta annab oma kinnituse vastutav uurija.

Uuringute tulemused publitseeritakse nii Eesti kui rahvusvahelistes teadusajakirjades (eelistatult vabalt kättesaadavate artiklitena), konverentsidel ja ettekannetel, üliõpilaste lõputöödes. Andmestiku metakirjeldus (kirjeldus andmeväljade kohta, ei sisalda reaalseid andmeid) OMOP-andmekogude registritesse, (nt http://portal.ehden.eu). Teadusuuringu muid tulemeid (andmestikku kirjeldus, andmehõive statistika ja -hinnangud, andmete standardiseerimise ja tekstikaeve meetodid, üleminekutabelid, teadusuuringute kokkuvõtted jms) tutvustatakse soovi korral andmeandjatele ning otsitakse koostööviise tulemite rakendamiseks andmeandjate andmebaasides.

Andmete kustutamine

Andmed hävitatakse hiljemalt 31. detsembriks 2028 kasutades hävitamise ajahetkel parimat kasutusolevat praktikat vastavalt rakendatavale infoturbestandardile (IS0/IEC 27001). Andmete hävitamine dokumenteerikse hävitamisaktis.. Andmete kustutamise protsessi käigus kustutatakse ka kõik andmete varukoopiad. Kui uuringu käigus tekib põhjendatud vajadus andmeid säilitada kauem, esitatakse enne uuringu lõppu Eesti bioeetika ja inimõiguste nõukogule (EBIN) vastavasisuline taotlus.

Andmete kustutamist teostab Eesti Teadusarvutuste infrastruktuuri meeskond koostöös vastutava uurijaga, et tagada permanentne ja pöördumatu andmete kustutamine. Andmete kustutamine dokumenteeritakse ja andmete kustutamise kohta koostatakse andmete kustutamise akt. Andmete kustutamise protokoll saadetakse kõikidele vastutavatele andmetöötlejatele.

Isikuandmete töötlemise eesmärgid

Töötlemise eesmärgid

Eesmärk on Eesti terviseandmete väärindamine läbi andmeteaduse meetodite, et pakkuda kvaliteetset tõendust tõhusate patsiendikesksete tervishoiu- ja ennetusteenuste osutamiseks.

Uuring viiakse läbi kahes etapis:

1. etapp: alusandmestiku loomine ja andmekvaliteedi tõstmise meetodite arendamine

1. etapi eesmärgid on:

- Hinnata valimi põhjal Eesti terviseandmete kvaliteeti ja hõivet võrreldes erinevatest andmeallikatest pärinevat informatsiooni.
- Andmekvaliteedi ja andmehõive tõstmise meetodite sh tehisintellekti-, tekstikaeve- ja imputatsioonimeetodite arendamine.
- Tervisesündmuste esituse ühtlustamine rahvusvahelistele standarditele üle aja ja andmeallikate, sh kasutades tekstikaeve, klasterdamise ja automaattõlke meetodeid
- Arendatavad meetodid on hiljem rakendatavad ka uutes uuringutes, suurematel andmehulkadel ning riiklikes infosüsteemides andmekvaliteedi ja -hõive parandamiseks.
- Standardiseeritud ja ühtlustatud andmestiku loomine ja kirjeldamine teadusuuringute läbiviimiseks. Andmestikku kirjeldatakse läbi erinevate numbriliste näitajate ja teostatavusanalüüside senisest suuremal detailsusastmel.

2. etapp: teadus- ja rakendusuuringute teostamine uurimismetoodikate arendamiseks Eesti terviseandmetel.

2. etapi eesmärgid on:

- Arendada haigustrajektooride ja raviteekondade analüüsi metoodikaid kasutades sh tehisintellekti meetodeid, et kirjeldada praeguseid ravipraktikaid, võrrelda neid raviteekondade ja -juhenditega ning modelleerida tervisetulemeid ja ravi majanduslikku mõju.
- Analüüsida erinevate faktorite (nt sugu, vanus, erinevad haigused) mõju ravijärgimusele, hinnata ravijärgimuse mõju erinevatele tervisetulemitele ja luua personaalseid ennustusmudeleid, mis võimaldavad maandada ravi mittejärgimise riske.
- Arendada personaliseeritud ennetusmudeleid, mis võimaldavad vaatlusandmete põhjal tuvastada praeguste haiguse ennetuste ja ravipraktikate kitsaskohti, planeerida senisest paremini haiguste ennetustegevusi ja hinnata pakutud ennetusteenuste majanduslikku mõju

Töötlemise õiguslikud alused

Euroopa isikuandmete kaitse üldmääruse (2016/679, edaspidi GDPR) artikkel 9 lg 2 (j) kohaselt on lubatud töödelda eriliiki terviseandmeid s.h terviseandmeid, kui *“töötlemine on vajalik*

avalikes huvides toimuval teaduseesmärgil /…/, ning on proportsionaalne saavutatava eesmärgiga, austab isikuandmete kaitse õiguse olemust ning tagatud on sobivad ja konkreetsed meetmed andmesubjekti põhiõiguste ja huvide kaitseks." Alljärgnevalt on selgitatud, kuidas need nõuded on täidetud.

Kas antud juhul toimub töötlemine teaduse eesmärgil?

Jah, andmeid töödeldakse ainult teaduse eesmärgil, täpsemalt teadusuuringute läbiviimise eesmärgil (vt eespool).

Kas antud juhul on töötlemine vajalik avalikes huvides?

GDPR preambula p 45 kohaselt *"kui /…./ töötlemine on vajalik avalikes huvides oleva ülesande täitmiseks /…/, peaks töötlemise alus olema sätestatud liidu või liikmesriigi õigusaktis."* Antud juhul tuleneb liikmesriigi (Eesti) õiguslik alus järgnevast:

- Eesti isikuandmete kaitse seadus (IKS) § 6 lg 4 lubab töödelda isikuandmeid teadusuuringu vajadusteks, kui IKS nõuete tingimuste täitmist kontrollib asjaomase valdkonna eetikakomitee. Vastav komitee on Eesti bioeetika ja inimõiguste nõukogu (EBIN) näol loodud sotsiaalministri määrusega 24.09.2019 nr 60 *"Uuringueetika komitee moodustamine, selle töökord, liikmete arv ja määramise kord ning uuringu taotluse läbivaatamise tasumäärad"*. EBIN ülesandeks on (§ 3 lk 3) *"isikute põhiõiguste ennetava kaitse tagamine ja uuringutele rakendatavate hindamispõhimõtete ühtlustamine, et kindlustada uuritavate isikute õiguste kaitsemeetmed ning uurijate kohustused neid kaitsemeetmeid järgida."* Uuringumeeskond on esitanud EBIN-le vastavasisulise taotluse.
- Tartu Ülikool on avalik-õiguslik juriidiline isik ning pakub teadustegevusel põhinevaid avalikke teenuseid. Eesti rahvast uurivate teaduste edendamine ning koostöö teiste ülikoolide ja kogu ühiskonnaga on Tartu Ülikooli seadusest tulenev Tartu Ülikooli eesmärk (Tartu Ülikooli seadus § 2 lg 2 ja 3). Taotletava eesti rahvastikul põhineva andmestiku loomise ja sellel läbiviidavate uuringutega edendab Tartu Ülikool nii meditsiini- kui andmeteadust, samuti koostööd teiste ülikoolide ja ühiskonnaga laiemalt. Tartu Ülikool teeb koostööd kogu ühiskonnaga, toetades Eesti ühiskonna arengut ja loob teadustegevusel põhinevaid võimalusi rahvusvaheliseks koostööks (Tartu Ülikooli seadus § 2 lg 5).

Meditsiinis ja rahvatervises on otsuste tegemiseks vaja laiapõhjalistel teadusuuringutel põhinevat tõendust. Uuringute tarvis on reeglina vaja esinduslikke andmeid, mis on võimalikult värsked ja samal ajal piisavalt pika aegreaga. Näiteks on värskeimad andmed (sh imikute ja vastsündinute kohta) vajalikud vastamaks küsimustele, mis puudutavad terviseteenuse osutamise hetkeseisu, uute ravimeetodite kasutust ja efektiivsust ning tervisesüsteemi vastust erinevatele hiljutistele sündmustele. Samas ennustusmudelite loomiseks, haiguse arengu, ravi efektiivsuse, erinevate ravitrajektooride jms hindamiseks on vajalik vaadelda pikemaid ajaperioode.

Ajakohastel andmetel tugineva tõenduse leidmiseks kasutatakse järjest enam päriselu terviseandmete teisesel kasutusel põhinevaid uuringuid. Tänu Eesti unikaalsele

tervishoiuandmete korraldusele on võimalik ühendada patsiendi tervist puudutavad olulised komponendid haiguste (Tervise infosüsteem, vähiregister), ravimite (retseptikeskus), tarbitud teenuste ja kindlustuskaitse (Tervisekassa andmekogu) ning surma põhjuste kohta. Nende andmetel OMOP andmemudelil põhinevate terviseuuringute läbiviimisel on Eesti tervishoiuvaldkonnale ja ühiskonnale laiemalt mitmekordne kasu. Nii aitab igapäevase tervishoiuvaldkonna toimimise raames salvestatud andmete taaskasutamine ühelt poolt hoida kokku andmete kogumisele kuluvat aega ja raha ning teisalt annab kõige täpsema ülevaate tegelikest protsessidest ja trajektooridest erineva terviseseisundi, soo ja vanusega isikutel. Oluline on siinjuures, et terviseandmete teisese kasutamise korral ei suurene patsientide uuringu koormus kuna kasutatakse varasemalt kogutud retrospektiivseid andmeid. Terviseandmete teisene kasutamine võimaldab ilma lisakoormust põhjustamata analüüsida ja osutada efektiivsemalt tervishoiu teenuseid patsiendi gruppidel, kellel traditsiooniliste teadusuuringute tegemine ei ole eetilistel kaalutlustel võimalik (näiteks ravimite mõju lastele või rasedatele[1]). Kaasates analüüsi infot patsiendi kogu raviajaloo kohta pikema aja vältel, mitte ainult uuritava haigusega seotud tegevuste kohta, võib-olla võtmetähtsusega oluliste seoste avastamisel. Lisaks on taotletava teadusuuringu käigus 30% valimi peal välja töötatud meetodid andmekvaliteedi parandamiseks rakendatavad ka teistes teadusuuringutes ning riiklikes andmebaasides kogu rahvastiku andmetel. Taotletav teadusuuring loob eeldused, et päriselu terviseandmeid saaks Eestis laiemalt kasutada ning aitab parandada tervishoiuteenuste kvaliteeti, tervishoiusüsteemi efektiivsust ning toetada tervishoiupoliitiliste tõenduspõhiste otsuste tegemist, toetades seeläbi rahvastiku tervise arengukava 2020-2030 elluviimist. Kvaliteetsem tervishoiusüsteem omakorda toetab tervelt elatud eluaastate suurenemist.

Seega on eriliiki isikuandmete töötlemine vajalik avalikes huvides ning on sätestatud Eesti õiguses.

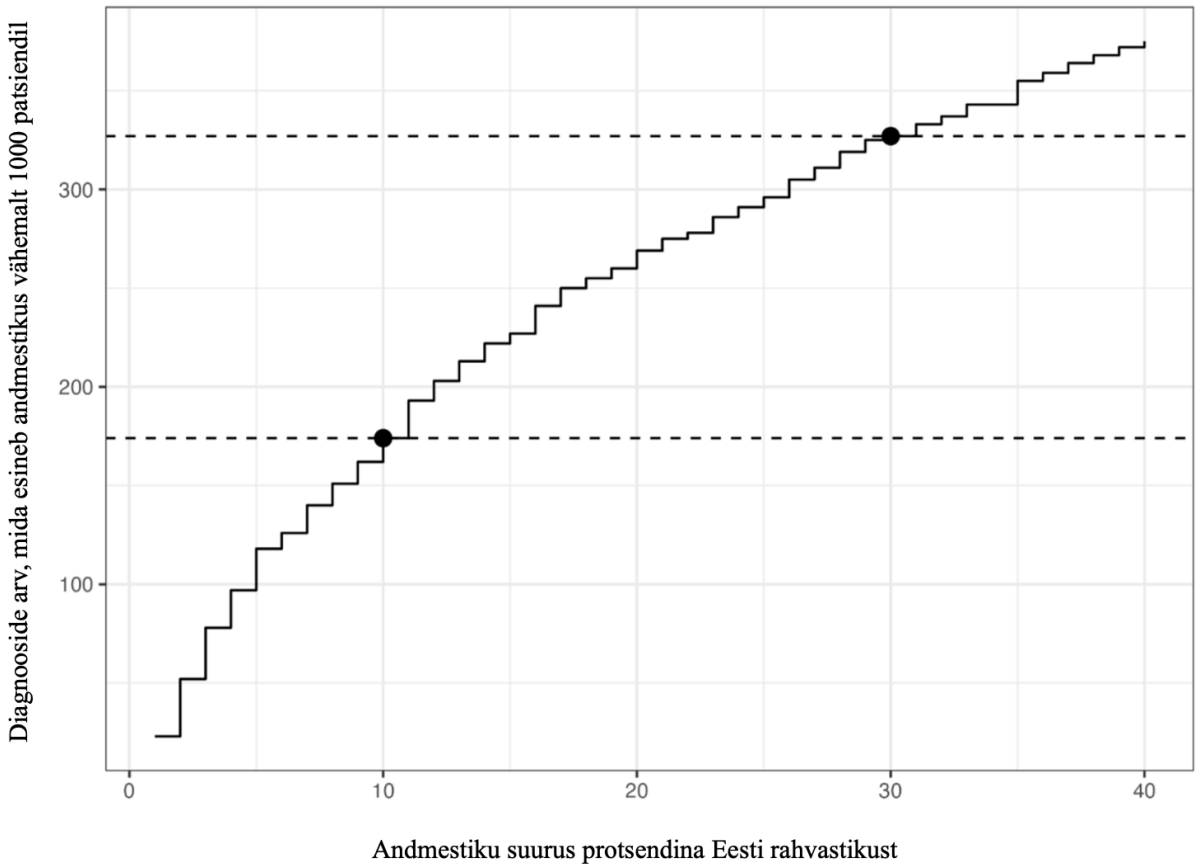
Kas isikuandmete töötlemise ulatus on proportsionaalne saavutatava eesmärgiga?

Eesti eri terviseandmekogudes sisaldub patsientide tervise kohta erinevat liiki ja eri kvaliteediga infot. Selleks, et saada tervisevaldkonna küsimuste uurimiseks patsiendi terviseseisundist kvaliteetne pilt, on vajalik vaadelda korraga andmeid mitmest andmekogust (vt ka taotlus punkt “11. Uurimismetoodika”).

Tervise infosüsteemi epikriisides on kirjas diagnoosid, raviskeemid, saatekirjades ja saatekirja vastustes ka laborianalüüside tulemused, kuid info võib olla puudulik. Epikriisides sisaldub vaid haigusloo kokkuvõte, mitte aga kõik teostatud uuringud või väljakirjutatud ravimid, mis mitmetes uuringutes on oluline info. Teiselt poolt on epikriisid ainsad tervisedokumendid, kus sisaldub vabas vormis kirja pandud info patsiendi kaebuste, üldseisundi, allergiate, ravi kõrvalmõjude jms kohta. Tekstilistest osadest suudame eraldada tehisintellekti meetodite abil ka muud olulist infot, mis kodeeritud väljades puudub, näiteks patsiendi kaebusi, ravimite nõrgemaid kõrvalmõjusid. Tervisekassa andmekogu info tervishoiuteenuste kohta on täielikum ja sisaldab ka teenuste hindu, kuid selle detailsus on samas madalam (mitmed tervishoiuteenused märgitud sama koodiga) ja laborianalüüside kohta puuduvad analüüsitulemused. Kõige parema pildi patsiendile välja kirjutatud ravimite osas annab retseptikeskus, lisaks on seal ka info ravimi väljaostmise kohta, mis on väga oluline indikaator ravijärgimuse hindamiseks. Samas puudub retseptikeskuses info käsimüügiravimite kohta (seda infot võib potentsiaalselt leida

epikriisidest). Paljude tervisevaldkonna uuringute puhul (nt pahaloomulised kasvajad) on oluline uurida suremust, selleks on kõige kvaliteetsem info kirjas surma põhjuste registris, mis sisaldab nii surma kuupäeva kui ka spetsialisti poolt kinnitatud surma põhjuseid.
Andmeväljade valikul oleme lähtunud uurimistöö eesmärkidest tuginedes uurimismeeskonna senistele kogemustele ja minimaalsuse printsiibist. Laiapõhjalisus võimaldab hinnata andmekvaliteeti üle paljude haiguste, sh arvestades kaasuvaid haigusi, ja tagab, et loodud andmetöötlusmeetodid üldistuvad uutele uuringutele sõltumata vaatluse all olevast haigusest ja patsiendi gruppidest. Lähtuvalt sellest, et üheks taotletava teadusuuringu eesmärgiks on, et arendatavad analüüsimeetodid ja ennetusmeetodid üldistuksid ka uutele andmestikele ja haigustele, ei saa me piirata andmestikku ei diagnooside, vanuse ega muude parameetrite järgi. Iga piirang vähendaks andmestiku esinduslikkust üldpopulatsiooni suhtes ja muudaks arendatavad meetodid vähem üldistuvaks ning piiraks loodud meetodite kasutamist teistes uuringutes kui ka rakendatavust Terviseinfosüsteemi andmekvaliteedi tõstmisel.
Andmete vaatlemine üle pika ajaperioodi võimaldab analüüsida tervisesündmuste esituse ja kvaliteedi arengut läbi aja ning uurida terviklikke haigustrajektoore alates ennetustegevustest, esmasdiagnoosist, raviteenuste osutamisest lõpptulemini. Tuginedes uurimisrühma varasemale kogemusele teame, et planeeritavateks uuringuteks vajaliku detailsusastmega terviseandmed on Eestis olemas alates 2012. aastast. Värskeimad andmed (sh imikute ja vastsündinute kohta) on vajalikud vastamaks küsimustele, mis puudutavad terviseteenuse osutamise hetkeseisu, uute ravimeetodite kasutust ja efektiivsust ning tervisesüsteemi vastust erinevatele hiljutistele sündmustele. Samuti on laiem andmestiku ajaraam vajalik, et hinnata riskimudelite 10-aastast ennustusvõimet vastavalt rahvusvaheliselt levinud praktikale.
30% suuruse juhuvalimi vajadus tuleneb uurimisrühma varasemast kogemusest ning rahvusvaheliselt levinud praktikast. Projekti “Tehisintellekti kasutamise võimalused meditsiinis” raames kasutasime 10% juhuvalimit 2012.-2019. aasta andmetest. Projekti käigus saime kinnitust, et erinevate riiklike terviseandmebaaside ühendamine ja laiapõhjalise terviseajaloo kasutamine annab olulist lisandväärtust andmete analüüsis. Näiteks koostöös Naistearstide seltsiga hindasime emakakaela vähi ennetuspraktika vastavust ennetusjuhendile ja erinevate papilloomviirustüvede levikut Eestis. Siiski ilmnesid 10% juhuvalimil koostatud andmestiku kasutamisel ka olulised kitsaskohad. Esiteks, oli andmestiku ajaraam liiga lühike, et vastavalt rahvusvahelistele prakikatele hinnata riskimudelite 10-aastast ennustusvõimet. Kuna andmestikku ei uuendatud, puudus info uute haiguste (nt COVID-19), ravipraktikate ning nende esituse ja andmekvaliteedi kohta. Näiteks seetõttu ei olnud võimalik ravijuhendite arendamisel teha koostööd ravijuhendite püsisekretariaadiga (https://tervis.ut.ee/et/ravijuhendid) ega panustada rahvusvahelistesse Euroopa Ravimiameti uuringutesse. Kõige olulisemana aga oleme korduvalt näinud, et 10% juhuvalim on liiga väike mitmete haigusseisundite või detailsemalt defineeritud kohortide uurimisel. Näiteks eesnäärmevähi rahvusvahelises uuringus osaledes saime vastused vaid osadele uurimisküsimustele, sest täpsemate tervisetulemite analüüsimiseks jäi valim liiga väikeseks.
Vajalik patsientide arv sõltub väga uurimisküsimusest ning ka ühe uuringu sees võib see erineda, kuid arvestades patsientide kliinilise pildi varieeruvust ja retrospektiivsete terviseandmete kvaliteeti, oleks hinnanguliselt minimaalne vajalik patsientide arv uuringute jaoks sadades kuid veel parem üle tuhande. Oleme välja arvutanud, et taotletava ajaperioodi ja andmete ulatuse korral on 10%-lise juhuvalimiga andmestikus diagnoose, mida esineb

vähemalt tuhandel patsiendil 175, samas kui 30%-lises juhuvalimi korral on selliseid diagnoose ligi kaks korda rohkem - 330 diagnoosi (joonis 1). Suurema valimi korral suureneb oluliselt saadavate tulemuste statistiline usaldusväärsus ja samas ka haiguste hulk, millel saame välja töötatud meetodeid rakendada.
Selle analüüsi põhjal näeme, et teadusuuringule seatud eesmärkide saavutamiseks on minimaalne valimi suurus 30%.



Joonis 1. Vähemalt 1000 patsiendil esinevate diagnooside arvu sõltuvus juhuvalimi suurusest, mis on võetud taotletava ajaperioodi ja andmete ulatusega Eesti terviseandmetest.

Kas isikuandmete töötlemine austab isikuandmete töötlemine isikuandmete kaitse õiguse olemust?

Meie hinnangul austab selles mõjuhinnangus kirjeldatud isikuandmete töötlemine isikuandmete kaitse õiguse olemust.

Võrdse kohtlemise printsiip – kirjeldatud isikuandmete töötlemine ei diskrimineeri kedagi rahvuse, soo, vanuse ega muu kuuluvuse alusel. Kõikide ühiskonna- ja vanusegruppide puhul on eesmärk toetada elukvaliteedi parandamist ja tervena elatud aastate arvu suurendamist. Juhuvalimi tõttu on kõikidel inimestel võrdne tõenäosus valimisse sattuda.

Heategemise printsiip – antud andmestikust ja sellel läbiviidavatest terviseuuringutest saadav kasu on uuritavatele pigem kaudne, kuna alustatakse baasteadmiste loomisega ja otsene kliinilises praktikas kasutatav kasu haigusega tegelemiseks võib tekkida alles mitme aasta pärast. Seega saavad antud tööst kasu pigem tulevased patsiendid.

Kahju vältimise printsiip – kirjeldatud isikuandmete töötlemine on kooskõlas mittekahjustamise printsiibiga, kuna ei koorma patsiente ega põhjusta riske nende tervisele.

Taotluses kirjeldatud Uurimisgrupil on pikaaegne kogemus sarnaste andmestike loomisel, terviseuuringute läbiviimisel ja tulemuste publitseerimisel, samuti selle valdkonna üliõpilaste õpetamisel. Uurimisgrupp on andmestiku loomiseks kirjutanud põhjaliku taotluse ja käesoleva andmekaitsealase mõjuhinnangu koos riskide maandusmeetmetega ning palunud neid hinnata Eesti bioeetika ja inimuuringute nõukogul.

Kas tagatud on sobivad ja konkreetsed meetmed andmesubjekti põhiõiguste ja huvide kaitseks?

Meie hinnangul on tagatud sobivad ja konkreetsed meetmed andmesubjekti põhiõiguste ja huvide kaitseks.

GDPR artikkel 6 lg 4 kohaselt võtab vastutav töötleja, juhul kui isikuandmete töötlemine toimub muul eesmärgil kui andmesubjekti nõusolekul ega põhine andmesubjekti nõusolekul, arvesse m.h *“asjakohaste kaitsemeetmete olemasolu, milleks võivad olla näiteks /…/ pseudonümiseerimine”*. Artikkel 89 lg 1 kohaselt kohaldatakse avalikes huvides toimuva teaduse eesmärgil isikuandmete töötlemise suhtes andmesubjekti õiguste ja vabaduste kaitseks asjakohaseid kaitsemeetmeid. *“Need meetmed võivad hõlmata pseudonümiseerimist, kui kõnealuseid eesmärke on võimalik saavutada sellisel viisil,”* kuid nõuab vajadusel täiendavat töötlemist, kui andmesubjektid on jätkuvalt tuvastatavad (*“kui kõnealuseid eesmärke saab täita täiendava töötlemisega, mis ei võimalda või ei võimalda enam andmesubjektide tuvastamist, täidetakse need eesmärgid sel viisil.”*). Vastavad tingimused ja kaitsemeetmed tuleb preambula p 157 kohaselt sätestada liikmesriigi õiguses (*“Teadusuuringute hõlbustamiseks võib isikuandmeid töödelda teadusuuringute eesmärgil, mille suhtes kohaldatakse asjakohaseid tingimusi ja kaitsemeetmeid, mis on sätestatud liidu või liikmesriigi õiguses.”).* Eestis reguleerib isikuandmete kaitset isikuandmete kaitse seadus (IKS), mis lubab isikuandmeid andmesubjekti nõusolekuta teadusuuringu vajadusteks töödelda pseudonüümitult (IKS § 6 lg 1 *“Isikuandmeid võib andmesubjekti nõusolekuta teadus- või ajaloouuringu või riikliku statistika vajadusteks töödelda eelkõige pseudonüümitud või samaväärset andmekaitse taset võimaldaval kujul. Enne isikuandmete üleandmist teadus- või ajaloouuringu või riikliku statistika vajadustel töötlemiseks asendatakse isikuandmed pseudonüümitud või samaväärset andmekaitse taset võimaldaval kujul andmetega.”*). IKS § 6 lg 3 kohaselt on teadusuuringu vajadusteks lubatud kasutada andmeid ka andmesubjekti tuvastamist võimaldaval kujul, kui täidetud on kolm tingimust: (1) pärast tuvastamist

võimaldavate andmete eemaldamist ei ole andmetöötluse eesmärgid enam saavutatavad või neid oleks ebamõistlikult raske saavutada; (2) selleks on ülekaalukas avalik huvi; (3) töödeldavate isikuandmete põhjal ei muudeta andmesubjekti kohustuste mahtu ega kahjustata muul viisil ülemäära andmesubjekti õigusi. Rahvatervise seadus lubab kasutada vähiregistri andmeid teadustööks isikustamata kujul (§ 14^1 lg 2). Tervise infosüsteemi andmete kasutamist teaduse vajaduseks lubab Tervishoiuteenuste korraldamise seadus § 59^3 lg 7, kui vajalikkust ja põhjendust ning isikute põhiõiguste kaitsemeetmeid on hinnanud uuringueetika komitee (§ 59^4 lg 1 ja lg 2). Vastav komitee on EBIN näol loodud sotsiaalministri määrusega 24.09.2019 nr 60 *"Uuringueetika komitee moodustamine, selle töökord, liikmete arv ja määramise kord ning uuringu taotluse läbivaatamise tasumäärad"*. EBIN ülesandeks on (§ 3 lk 3) *"isikute põhiõiguste ennetava kaitse tagamine ja uuringutele rakendatavate hindamispõhimõtete ühtlustamine, et kindlustada uuritavate isikute õiguste kaitsemeetmed ning uurijate kohustused neid kaitsemeetmeid järgida."*

Käesolevas projektis töödeldakse andmeid pseudonüümitud kujul, kuid lisaks rakendatakse andmesubjektide tuvastamise riski maandamiseks ka täiendavat töötlust:

1. Uurimisgrupp ei tea, millised isikud kuuluvad valimisse.
2. Taotletav andmestik ei sisalda isikute nimesid, isikukoode, aadresse jms. Taotletavad andmed pseudonüümitakse andmeandmeandja poolt enne Uurimisgrupile väljastamist. Pseudonüümimisvõti on tagasipööramatu ning seda taotluses kirjeldatud meeskonnale ei avaldata.
3. Andmestiku loomisel ega hiljem ei toimu depseudonüümimist ja uuritavatega ühendust ei võeta.
4. Andmete töötlus toimub spetsiaalsel tundlike andmete platvormil SAPU, mis piirab kasutajate võimalust andmeid serverist välja kopeerida ega võimalda andmeid muul moel linkida teiste andmestikega (maandab linkimisründe riski andmesubjektide tuvastamiseks).
5. Esimeseks tegevuseks SAPU serveris on andmete täiendav töötlemine automaatse anonüümimisrakendusega, et eemaldada andmete vabatekstilistest osadest võimalikud nimed, telefoninumbrid, isikukoodid, aadressid.

Muud kaitsemeetmed on kirjeldatud käesoleva mõjuhinnangu lõpus.

Kuigi käesolevas uuringus ei kasutata andmeid andmesubjekti tuvastamist võimaldaval kujul, vaid pseudonüümitult ja rakendatakse ka muid kaitsemeetmeid andmesubjekti tuvastamise riski maandamiseks, on siiski täidetud ka IKS § 6 lg 3 kolm tingimust: (1) terviseuuringute, s.h ravijärgimuse ja ravitrajektooride uurimiseesmärke oleks ebamõistlikult raske saavutada ilma üksikpatsientide tasemel andmeid analüüsimata; (2) uuringu vastu on ülekaalukas avalik huvi (vt põhjendust eespool); (3) töödeldavate isikuandmete põhjal ei muudeta andmesubjekti kohustuste mahtu ega kahjustata muul viisil ülemäära andmesubjekti õigusi (samuti põhjendatud eespool).

Riskid ja nende maandamine

Riskide kaardistus koos maandamise meetmega on toodud järgmises tabelis. Riski tõenäosust on hinnatud pärast maandusmeetmete rakendamist (nt ilma SAPU kasutamiseta oleks tõenäosused märgatavalt suuremad):

Riski nr	Riski nimetus	Jääkriski tõenäosus (0-väga madal; 4-väga kõrge)	Riski mõju (0-väga madal; 4-väga kõrge)	Riski tase	Tegevused / ettepanekud riski maandamiseks
1	Inimeste terviseandmed saavad avalikuks	1 (madal)	4 (väga kõrge)	Keskmine	SAPU kasutamine (ei saa kopeerida), mis on kättesaadav vaid Tartu Ülikooli sisevõrgust ainult Uurimisgrupi liikmetele, serveri monitooring ja tegevuste logimine, kindel protsess andmete SAPU-st väljatoomiseks, pika kogemusega Uurimisgrupi liikmed, konfidentsiaalsusklausel töölepingus, füüsilised turvameetmed serveriruumil, andmed pseudonüümitud, juhuvalimi kasutamine (mitte kõik Eesti inimesed). Uurimisgrupi liikmete juhendamine ja koolitamine.
2	Andmestikus olevate andmesubjektide suurel hulgal (süstemaatiline) tuvastamine	1 (madal)	4 (väga kõrge)	Keskmine	Otseste isikuandmete eemaldamine andmeandjate poolel enne Uurimisgrupile edastamist, unikaalsete pseudonüümide kasutamine, juhuvalimi kasutamine (mitte kõik Eesti inimesed), SAPU kasutamine (ei saa andmeid kopeerida ega teiste andmestikega linkida)
3	Andmestikus konkreetse andmesubjekti juhuslik tuvastamine	2 (keskmine)	1 (madal)	Keskmine	Otseste isikuandmete eemaldamine andmeandjate poolel enne Uurimisgrupile edastamist, unikaalsete pseudonüümide kasutamine, juhuvalimi kasutamine (mitte kõik Eesti inimesed), väga väike tõenäosus, et juhuslikult tuvastatud isik on uurijale tuttav
4	Andmestikust iseenda tuvastamine	3 (kõrge)	0 (väga madal)	Keskmine	Otseste isikuandmete eemaldamine enne Uurimisgrupile edastamist, unikaalsete pseudonüümide kasutamine, valimisse kuulumine juhuslik, iseenda tuvastamisel ei saa uurija teada uut informatsiooni

5	Andmeallikad avaldavad Uurimisgrupile pseudonüümimis-pa rooli, mis muudab andmesubjektid Uurimisgrupi poolt tuvastatavaks	1 (madal)	2 (keskmine)	Keskmine	Selgelt kindlaksmääratud protsess parooli vahetamiseks andmeallikate vahel ilma seda Uurimisgrupile avaldamata, juhuvalimi kasutamine (mitte kõik Eesti inimesed), SAPU kasutamine (ei saa andmeid kopeerida ega teiste andmestikega linkida)
6	Andmeandja kaotab pseudonüümimiseks kasutatava parooli (pole võimalik enam andmeid uuendada)	1 (madal)	2 (keskmine)	Keskmine	Vajadusel saab parooli uuesti küsida teiselt andmeallikalt. Kui parool on lõplikult kadunud, kustutab Uurimisgrupp kõik seni antud andmed ja küsitakse kõigilt andmeandjatelt kõik vajalikud andmed uuesti (kuigi sel juhul muutub ka valim)
7	Vabatekstiline info võib sisaldada sensitiivseid andmeid	2 (keskmine)	1 (madal)	Keskmine	Andmete saamisel on SAPU serveris esimeseks sammuks anonüümimisrakenduse kasutamine, mis tuvastab ja asendab vabatekstilistes dokumendiosades isikunimed, aadressid, telefoninumbrid, isikukoodid. Juhuleidude korral on Uurimisgrupil kindel protsess nende käsitlemiseks, täiendatakse vastavalt anonüümimisrakendust ja teostatakse anonüümimine uuesti.
8	Andmetele saavad ligi Uurimisgrupi välised isikud	2 (keskmine)	1 (madal)	Keskmine	SAPU serverisse kasutajakontode lisamine ja eemaldamine käib kindlaksmääratud protsessi alusel, regulaarselt vaadatakse üle kõigi kasutajate õigused, lähtutakse TÜ IT-turbe kordadest ja kõik Uurimisgrupi liikmed peavad läbima küberhügieeni ja kodust töötamise infoturbe alased ning andmekaitse alse koolitused ja eksamid (https://cyberhygiene.ut.ee/)
9	Andmeid kasutatakse uurimismeeskonna poolt muuks otstarbeks kui lubatud	1 (madal)	1 (madal)	Madal	SAPU kasutamine (ei saa andmeid kopeerida, automaatne tegevuste logimine), Uurimisgrupi kõrge kvalifikatsioon ja pikaaegne kogemus terviseandmetega töötamisel
10	Avaldatavad tulemused on liiga detailsed (andmesubjektide tuvastamise risk)	1 (madal)	1 (madal)	Madal	Kindel protsess andmete SAPU-st väljatoomiseks, tulemuste avaldamisel kontrollitakse, et tagatud on k≥5 anonüümsus, pika kogemusega Uurimisgrupi liikmed

Kasutusel olevad riskide vältimise meetmed

Tartu Ülikoolis on riskide vältimise aluseks riskianalüüs, mida tehakse igas vajalikus valdkonnas / teemas / projektis ning selle eest on vastutav vastava valdkonna / teema / projekti esindaja.

Riskianalüüsi käigus:

- kirjeldatakse võimalikud riskid,
- hinnatakse iga riski tõenäosust ja võimalikku mõju,
- vastavalt riski tõenäosusele ja võimalikult mõjule määratakse riski tase,
- vajadusel kirjeldatakse riskide kontrollimise ja maandamise tegevused.

Regulaarseid riskianalüüse viiakse läbi vastavalt vajadusele.

Antud projektis vastutab riskide vältimise ja vajalike meetmete rakendamise eest vastutav uurija, kes saab vajadusel abi Tartu Ülikooli siseauditi büroolt.

Füüsilised turvameetmed

Andmetöötlus toimub Tartu Ülikooli teadusarvutuste keskuse infrastruktuuril:

- Jälgitakse Eesti infoturbestandardiga kehtestatud nõudeid.
- Erinevad ressursid on eraldatud võrgu tasandil.
- Töötajaid koolitatakse järjepidevalt.
- Kõik võrguseadmed ja serverid asuvad Tartu Ülikooli majutatud suletud andmekeskustes.
- Andmekeskustes kasutatavad tulekustutussüsteemid toimivad automaatselt, on gaasipõhise lahendusena ning on ette nähtud andmekeskustes kasutamiseks.
- Andmekeskustes ei hoita kergestisüttivaid või tuleohtlikke esemeid.
- Andmekeskuste konstruktsioonides ja sisustuses on viidud miinimumini süttivate materjalide, nagu puu, tekstiil ja sünteetilised materjalid kasutamine.
- Andmekeskused on kaitstud uputuste ja veekahjustuste eest.
- Andmekeskustes on tagatud optimaalne temperatuur ja õhuniiskus.
- Andmekeskused on kaitstud sissemurdmise ja volitamata sisenemise eest.
- Füüsiliselt pääsevad andmekeskusesse nimelist (personaalset) juurdepääsuõigust omavad isikud.
- Isikliku juurdepääsuõigusega isikud pääsevad andmekeskusesse kas võtme või töötõendi ja valvekoodi abil.
- Ilma isikliku juurdepääsuõiguseta isikutel on võimalik andmekeskusesse siseneda üksnes andmekeskusesse juurdepääsu omava isiku juuresolekul.
- Andmekeskuste turvalisuse tagamiseks kasutatakse tehnilist valve- ja läbipääsusüsteemi ning videovalvet.
- Valve- ja läbipääsusüsteem salvestab andmed juurdepääsukaartide kasutamise ja valvestamise kohta.

- Andmekeskused asuva kahe tulekindla ukse taga, mida saab avada vaid isikliku kiipkaardiga või spetsiaalse võtmega.
- Andmekeskused on elektroonilise valve all ning andmekeskusesse sisenemisel tuleb elektrooniline valve isikliku koodi abil deaktiveerida.
- Andmekeskuse elektrooniline valve on deaktiveeritud ainult siis, kui keegi asub füüsiliselt andmekeskuses, on sinna sisenemas või sealt lahkumas.
- Kõik andmekeskusse sisenemised ja elektroonilise valve deaktiveerimised/aktiveerimised logitakse.
- Tuleohutuse tagamisel järgitakse Ülikooli tuleohutuseeskirju.
- Tartu Ülikooli teadusarvutuste keskuses kehtivad tehnilised ja organisatoorsed meetmed infoturbe tagamiseks ning andmete kaitsmiseks. Valik tehnilisi ja organisatoorseid meetmeid (turvakaalutlustel ei ole avalikustatud kõik tehnilised ja organisatoorsed meetmed) on toodud Tartu Ülikooli teadusarvutuste keskuse koduleheküljel https://hpc.ut.ee/terms/information-security (inglise keeles).

Infotehnoloogilised turvameetmed

Andmetöötlus toimub Eesti Teadusarvutusteinfrastruktuuril SAPU keskkonnas, kus:

- Jälgitakse Eesti infoturbestandardiga kehtestatud nõudeid.
- Teostatakse regulaarselt serverite testimist, uuendamist ja monitoorimist.
- Haavatavuste tuvastamiseks kasutatakse monitoorimist, masinõpet ning ka erinevaid läbistusteste. Muuhulgas kasutatakse haavatavuste tuvastamiseks ka juba olemasolevaid haavatavuste tuvastamise tarkvarasid (näiteks Nessus) ning jälgitakse järjepidevalt erinevaid haavatavuste nimekirju. Lisaks Eesti Teadusarvutuste infrastruktuurile skaneerib taristu avalikult kättesaadavaid ressursse ka CERT-EE.
- Erinevad ressursid on eraldatud kasutajaõiguste tasandil.
- Õiguste määramisel lähtutakse minimaalsuse põhimõttest ja vaikimisi administraatori juurdepääsu ei võimaldata.
- Servereid skaneeritakse regulaarselt ja jooksvalt jälgitakse ka võrguliiklust.
- Vaikimisi on keelatud kõik tegevused, mis ei ole otseselt vajalikud töö tegemiseks.
- Kasutatakse andmete varundamist Eesti Teadusarvutuste infrastruktuuri lindirobotile, mis asub füüsiliselt teises asukohas (Eesti Teadusarvutuste infrastruktuuri andmekeskuses).
- Kõik kasutajate tegevused SAPU keskkonnas logitakse ja logisid monitooritakse.
- Kõik potentsiaalsed turvaintsidendid ja turvanõrkuste leidmise katsed logitakse (näiteks sisse logimise katsed, pöördumised erinevate portide poole, kasutajaõiguste muutused jne).
- SAPU keskkonnas olevate kasutajate ekraanipilt salvestatakse.
- SAPU keskkonnast info/andmete välja liigutamine on võimalik ainult, kui vastutav uurija on vastavad andmed üle vaadanud ja selleks nõusoleku andnud.
- SAPU keskkonnast info/andmete välja kopeerimine ei ole võimalik („copy“ käsk).
- SAPU keskkond asub eraldi tulemüüri taga.
- Interneti juurdepääs SAPU masinast on täielikult suletud ja ei ole võimalik teha päringuid internetti.

- SAPU keskkonnas on eelinstalleeritud tarkvara ja kasutajal ei ole võimalik keskkonda tarkvara ise installeerida.
- Andmete liigutamine (kaasa arvatud analüüsi tulemuste) SAPU keskkonnast välja vajab vastutava uurija nõusolekut.
- SAPU keskkondi varundatakse regulaarselt.
- Perioodidel, kui SAPU keskkonda ei kasutata, on keskkond välja lülitatud ja keskkonda ei ole võimalik siseneda.
- Eesti Teadusarvutuste infrastruktuuris kehtivad tehnilised ja organisatoorsed meetmed infoturbe tagamiseks ning andmete kaitsmiseks. Valik tehnilisi ja organisatoorseid meetmeid (turvakaalutlustel ei ole avalikustatud kõik tehnilised ja organisatoorsed meetmed) on toodud Tartu Ülikooli teadusarvutuste keskuse koduleheküljel https://hpc.ut.ee/terms/information-security (inglise keeles).