Isikuandmete kaitse mõjuhinnang projektile: *„E-sotsiaalandmete kasulikus südame-veresoonkonna haigustekke ennustamisel“*

Sissejuhatus: miks on seda dokumenti vaja

Käesolevas teadusuuringus analüüsime Eesti elanike eriliiki isikuandmeid ja loome riskiennustusmudeleid, mille põhjal avaldame teaduspublikatsioone. ~~Uuringu plaanipärase lõpuni jõudmisel ei ole uuring uuritavaid otseselt mõjutanud. On suur tõenäosus, et selle projekti jooksul loodavad rakendused ja tehtud avastused on aluseks tuleviku uuringutele terviseteenuste optimeerimisel. Selliseid tuleviku rakendused võivad automaatselt profileerida suur hulka inimesi. Selle asemel, et tuleviku projektide riske ja seoseid siit dokumendist kustutada, jätame need halli fondiga nähtavaks. See aitab maandada projektide ülest riski (kus esmase projekti risk avaldub teises).~~ Käesoleva teadusuuringu käigus analüüsime juriidilises mõistes **"ulatuslikult" palju uuritavaid** (st >5000 inimese eriliiki andmeid). Projekti andmetöötlus on planeeritud ja metoodiline, seega "süstemaatiline". Piiriülest andmetöötlust ei toimu. Oleme teinud kavandatud andmetöötlusele andmekaitsealase mõjuhinnangu ja hinnanud ohte. Selle põhjal leiame, et abinõude rakendamisel on ohtude ilmemise risk niivõrd väike, et uus andmetöötlus tõenäoliselt ei too eendaga kaasa suure ohu teokssaamist (nt. võimaldaks identiteedivargust, rahalist kahju, mainekahju, diskrimineerimist). Lähtuvalt ei plaani uurimismeeskond käesoleva uuringu raames eelkonsulteerida Andmekaitse Inspektsiooniga. (Küll plaanime seda teha jätku-uuringutes mis võivad suure ohu tõenäosust suurendada). Lähtuvalt peab ülikool arvestama andmetöötluse ohte ja dokumenteerima vastavust isikuandmete kaitse *(ing. k.* GDPR) üldmäärusele.

Selle dokumendi kirjutamisel osalesid:

- Taavi Tillmann, TÜ rahvatervishoiu kaasprofessor
- Nikita Umov, TÜ rahvatervishoiu nooremteadur
- Anneli Uusküla, TÜ epidemioloogia professor
- Marek Oja, TÜ terviseinformaatika teadur
- Tommy Tomson, TÜ infoturbe nõunik
- Terje Mäesalu, TÜ andmekaitse peaspetsialist

Kokkuvõte andmekaitsealase mõjuhinnangu tulemustest

Andmesubjektid on ligikaudselt 660 000 Eesti elanikku, kes 01.01.2012, 01.01.2014 või 01.01.2016 seisuga olid vähemalt 40 aastat vanad.

Antud mõjuhinnang **kehtib** senikaua kuni uuringuks kogutud andmed hävitatakse. Kõige hiljem toimub see vahetult enne projekti lõppemist (31.08.2028) või varem, kui uuringu projekti eesmärgid saavutatakse enne projekti tähtaja lõppu.

Toome välja, et analüüsis kasutatakse pseudonüümsel kujul e-terviseandmeid ja sotsiaalandmed. Iga terviseandmebaasiga kaasneb teoreetiline ja maandamatu isikutuvastuse risk. Nimelt, äärmiselt kogenud ja lisaandmetega varustatud väärkasutaja, teatutud pingutuste tulemusel võib terviseandmetest isikuid tuvastada. Isikutuvastuse risk tõuseb iga lisatud andmebaasiga. Käesolev projekt lisab e-terviseandmetele veel 13 andmebaasi. Vaatamata eelpoolmainitule, on praktiliselt võimatu isikutuvastust teostada tulenevalt kasutusolevatest riskimaandus meetmetest.

Järeldused: Selles projektis on kasutusel tõhusad riskimaandamise mehhanismid (sh andmevõtit ei saa avada ükski Ülikooli töötaja; analüüs toimub eriti sensitiivsete andmete analüüsimise keskkonnas, millele on raskendatud füüsiline ja virtuaalne ligipääs ja kus iga analüüs salvestatakse. Analüüs tuvastas, et peamine jääkrisk, mida annab veel maandada pole mitte tehniline vaid andmekildudele ligipääsu omavate inimeste **käitumine**. Lähtuvalt, plaanime rohkem suhelda ja koolitada projektiga seotuid inimesi, et nad mõistaks ja langetaks projektiga seotud andmekaitse riske.

Täpsemad riskid, mis tuvastati andmekaitse mõjuhinnangu läbiviimise tulemusena on:

Riski nr	Riski nimetus	Riski tase	Ettepanek riski maandamiseks
1	**Andmete kasutamine muuks otstarbeks kui esialgselt kogutud**. See muudab siseringi kultuuri ja/või andmeid avalikustatakse. Ei tuvasta uuritavaid.	1: Madal	Maandada, kasutusele võtta näost-näkku ennetavad vestlused alluvatega; tagada õige eeskuju; keelata, taunida ja proportsionaalselt karistada andmete kasutamine muuks otstarbeks, kui esialgselt kogutud.
2	Tundmatu isik saab **ligipääsu** kõigile andmetele. Ei tuvasta uuritavaid	1: Madal	Aktsepteerida
3	Uurija või tundmatu isik suudab **tuvastada** mõned **üksikud** uuritavad (näiteks kõrvutades lisa andmebaasidega mis pole kooseisus). Võimalik kahju uuritavale.	1: Madal	Aktsepteerida
4	Üks üheksast andmevõtme võimalikest avajatest suudab endale saada piisavalt andmebaase teistelt omanikest, misläbi	1: Madal	Maandada: Hoiatada kõiki andmevõtme avajaid, et nad ei tohi andmeid üksteisele saata ja selle tagajärgedest. Nad peavad saatma vaid Tartu Ülikoolile, kes ei saa võtit ise.

	suudab ta kiiresti **tuvastada kõik** uuritavad. Võimalik kahju uuritavatele.		

1. Sissejuhatus

Tartu Ülikool on Eesti suurim ülikool, avaldades üle poole Eesti teaduspublikatsioonidest. Antud projekti käigus saavad magistritööd teha kaks magistranti ja doktoritööd teha üks doktorant. Seotud ja laiem teadustöö toimub pseudonümiseeritud kujul kolmes instituudis: arvutiteaduse instituut keskendub andmete puhastamisele ja eeltöötlemisele; peremeditsiini ja rahvatervishoiu instituut keskendub andmete lõpp-töötlemisele rahvatervise aspektist.

1.2 Mõjuhinnangu läbiviimisest

Mõjuhinnang viidi läbi ajavahemikul 01.01.2022-01.11.2023, hõlmates vaid neid andmeid, mida analüüsitakse kuni 31.08.2028.a. projektis *„E-sotsiaalandmete kasulikus südame-veresoonkonna haigustekke ennustamisel“* Kasutasime riskipõhist meetodit, tuginedes AKIe näidisele.

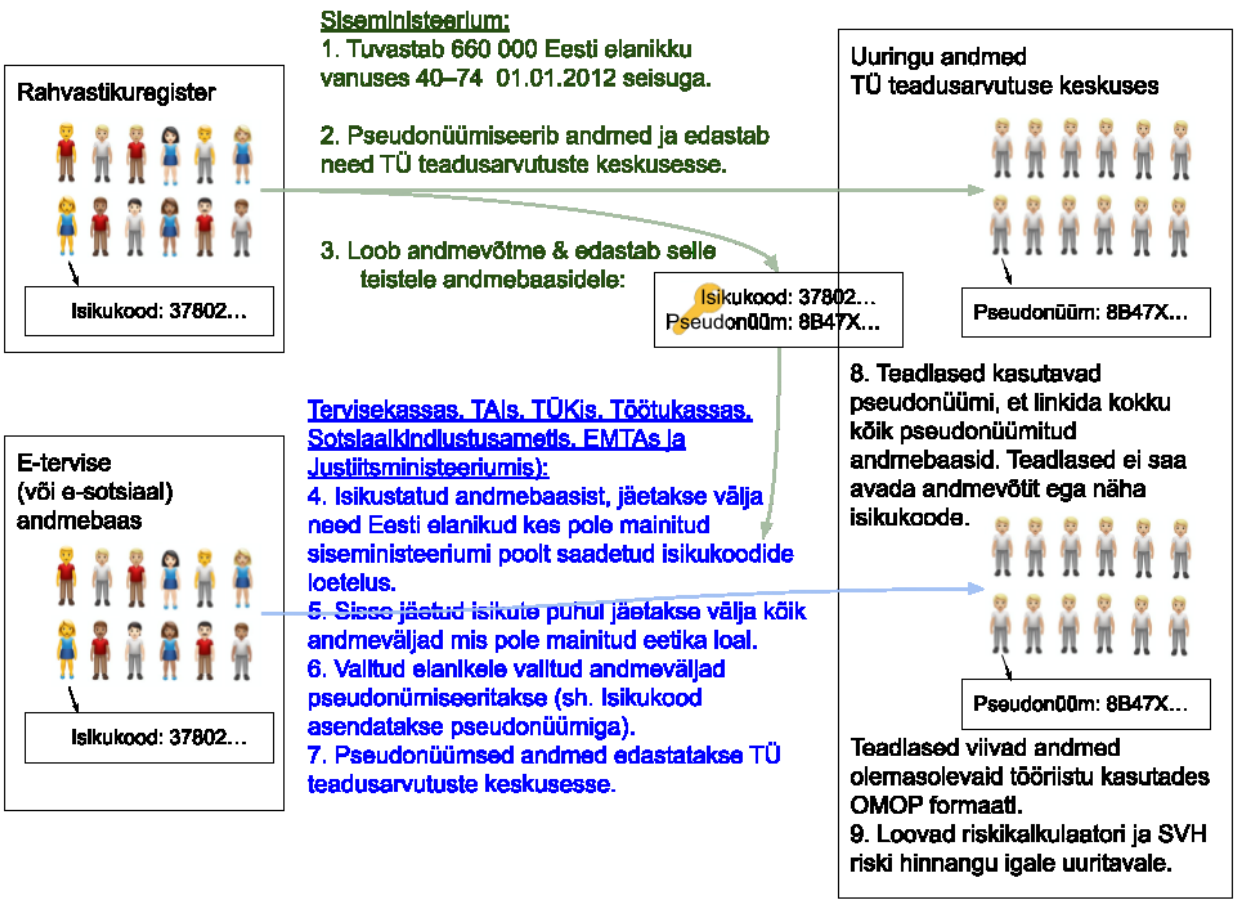
2. Projektipõhise Infosüsteemi kirjeldus

Kõik analüüsid teostatakse Tartu Ülikooli teadusarvutuste keskuse poolt pakutavas sensitiivsete andmete privaatses uurimiskeskkonnas (SAPU). Projekti käigus kasutatakse eraldiseisvat SAPU keskkonda, mis ei ole seotud teiste SAPU keskkondadega. Detailne info Tartu Ülikooli teadusarvutuste keskuse kohta on leitav aadressil https://hpc.ut.ee ja lisainfo sensitiivsete andmete privaatse uurimiskeskkonna (SAPU) koht on leitav aadressil https://docs.hpc.ut.ee/public/services/SAPU/ (inglise keeles).

2.1 Kasutajaks registreerimine: Analüüsi käigus kasutatavasse SAPU keskkonda luuakse spetsiaalsed ja eraldiseisvad kasutajakontod ainult projektijuhi taotlusel ja heakskiidul vastavalt vajadusele. 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.

2.2. Isikuandmete allikad: Projekt lisab teadusarvutuste keskusesse uusi eel-pseudonüümitud andmebaasi ja seob need unikaalse/individuaalse pseudonüümiga.

2.3. Andmejoad: Projekti andmete liikumise kokkuvõtte on näidatud allolevas joonises:



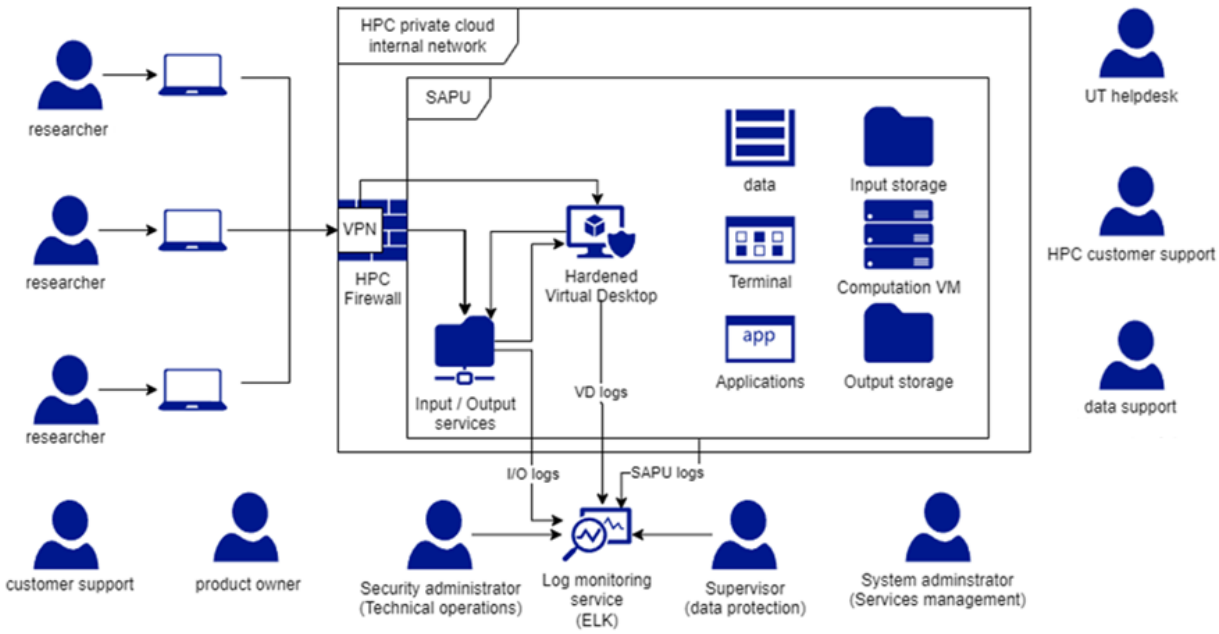
Teadaolevad turvameetmed:

- Andmete edastamiseks sõlmitakse leping iga andmekogu omaniku ja Tartu Ülikooli vahel. Andmete edastus toimub krüpteeritult läbi turvalise SFTP serveri, mis on üles seatud Tartu Ülikooli teadusarvutuste keskuse poolt.
- Analüüsiks kasutatakse Tartu Ülikooli teadusarvutuste keskuse 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 punktis 6.1

Süsteemi kasutusotstarve:

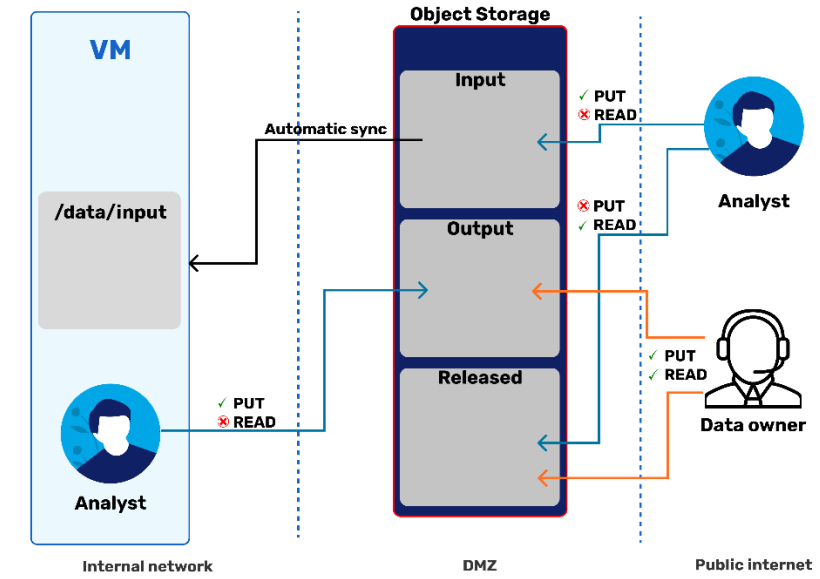
Sensitiivsete andmete privaatne uurimiskeskkond (SAPU) on Tartu Ülikooli teadusarvutuste keskuse poolt spetsiaalselt loodud ja pakutav andmetöötluskeskkond, kus analüütikud saavad töötada tundlike andmete kallal, vähendades võimalikku andmete volitamata kopeerimist, ülekandmist või masinatest välja võtmist,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

SAPU kõrgetasemeline arhitektuur



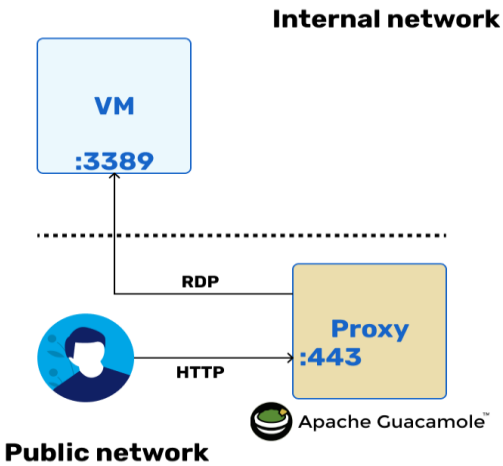
SAPU andmevärav

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



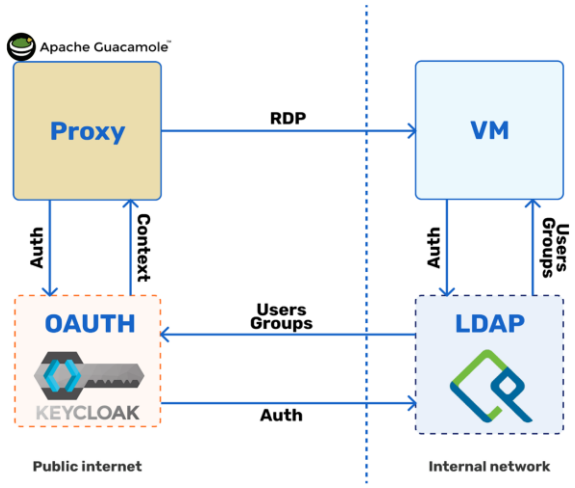
Graafiline kasutajaliides

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.



3. Isikuandmete töötlemise toimingud

Sissejuhatus (kes mis andmed töötleb): Laiemas vaates (enne ja pärast käesolevat projekti) Eesti riik ei analüüsi elanike e-tervise- ja e-sotsiaalandmeid, et hinnata haigustekke riski ega paku neile sekkuvaid teenuseid. Käesolev projekt on osa laiemast ambitsioonist seda saavutada.

3.1 Isikuandmete kogumine

3.1.1. Andmesubjektid: Uuringusse kaasatakse kõik, kes olid kas 01.01.2012, 01.01.2014 või 01.01.2016 seisuga Eesti elanikud ja vähemalt 40 aastat vanad.

3.1.2. Andmebaasid:

Omanik / Vastutav töötleja	Volitatud töötleja	Andmebaasi (lühendatud) nimi	Andmete liigitus: Pseudonümiseeritud ...	Kes kontrollib, kas isikuandmete töötlemise seaduslikud nõuded ja tingimused on täidetud
Tervisekassa		Raviarved	...eriliiki isikuandmed	**TÜ IEK,** tulenevalt isikuandmekaitse seadusest §6(4)
Tervisekassa		Digiretseptid	...eriliiki isikuandmed	
Tervisekassa		Kindlustatud isikud	...eriliiki isikuandmed	
Sots. ministeerium	TAI	Surma põhjuste register	...eriliiki isikuandmed	
Sots. ministeerium	TÜK	Müokardiinfarktiregister	...eriliiki isikuandmed	
Siseministeerium		Rahvastikuregister	...eriliiku isikuandmed	
Eesti Töötukassa		Töötute isikute register	...tundlikud isikuandmed	**Registripidaja** ise hindab, otsustab ja selgitab Uurijatele, millise seadusel alusel andmeid tohib väljastada ja kas selle seaduse nõuded ja tingimused on täidetud. Eetikakomitee või AKI kooskõlastust pole ilmtingimata vaja.
Sots. kindlustusamet		Sotsiaalkaitse infosüsteem	...tundlikud isikuandmed	
Maksu- ja Tolliamet		Maksukohustuslaste register	...tundlikud isikuandmed	
Maksu- ja Tolliamet		Töötamise register	...isikuandmed	
Justiitsministeerium		Kinnistusraamat	...isikuandmed	
Justiitsministeerium		Äriregister	...isikuandmed	

3.1.3. Andmekooseisud:

Vt. Lisatud avaldus kooskõlastuse saamiseks Tartu Ülikooli inimuuringute eetika komiteelt (TÜIEK BigHeart 6.doc, lk 21-25)

3.1.4. Andmekogumise protsess

Uurijad kasutavad turvalisuse huvides vaid pseudonüümitud andmeid. Isikustatuid andmeid (sh isikukoode) töötlevad vaid andmete omanikud (st kas vastutavad töötlejad nagu ministeeriumid ja riigiametid või nende poolt volitatud töötlejad nagu TAI). Seda tehakse seepärast, et ülikoolile edastatavatest pseudonüümsetest andmetest eemaldada isikud, kes ei kuulu analüüsitavasse valimisse (ehk n-ö mittekattuvad isikud). See samm lähtub andmete minimiseerimise printsiibist

Teostamiseks on vaja, et üks valimi koostaja looks andmevõtme tabeli valimisse kuuluvate isikukoodide ja vastavate pseudonüümidega. See esmane asutus edastab andmevõtme kõigile teiste andmebaasi omanikele. Praktilistel kaalutlustel on lihtsam, kui andmevõtme edastamisega tegeleb uurimismeeskond. Selleks, et uurimismeeskond ei pääseks andmevõtmele ligi, krüpteerib valimi koostaja andmevõtme nii, et seda saab avada vaid nimetatud andmebaasi nimetatud töötaja. See samm tõstab turvalisust.

Rahvastikuregistri omanik (siseministeerium) pseudonüümib andmed (sh eemaldades nimed ja isikukoodid, asendades sünnikuupäeva vanusega), loob igale inimesele projektipõhise pseudonüümi ja väljastab pseudonümiseeritud andmed Tartu Ülikooli teadusarvutuste keskusele. Siseministeerium loob eraldi faili ehk andmevõtme, mis sisaldab loetelu valimis olevate inimeste isikukoodidest ja pseudonüümidest. Andmevõti edastatakse krüpteerituna Tartu Ülikooli uurimismeeskonnale. Krüpteering on teostatud nii, et antud tabelit saavad dekrüpteerida ainult seitse nimetatud isikut üheksast riigiasutusest (st Eesti Töötukassa, Maksu- ja Tolliamet, Sotsiaalkindlustusamet, Justiitsministeerium, ~~Tartu Ülikooli genoomika instituut,~~ Tervisekassa, TAI, Tartu Ülikooli kliinikum). Peale seitse isiku nime kooskõlastamist edastavad uurijad need üheksa nime siseministeeriumile. Keegi teine (sh antud taotluses loetletud uurijad ja muu personal TÜ teadusarvutuste keskuses) andmevõtit dekrüpteerida ja avada ei saa. Seega ainsad inimesed, kes näevad ja töötlevad isikukoode, on seitse andmebaasi omaniku esindajat.

Uurimismeeskond täidab seejärel sekretäri funktsiooni ja edastab andmevõtme teiste andmekogude omanikele ja volitatud andmetöötlejatele. Seeläbi koordineerib Tartu Ülikooli uurimismeeskond andmete väljastamist erinevatest allikatest samade pseudonüümidega ilma ise andmevõtit avamata.

Andmevõtme alusel tuvastavad teised andmekogude omanikud, kes on uuringu valimis ja kes mitte. Teistest andmekogudest edastatakse vajalikud andmeväljad pseudonüümsed kujul (sh koos sotsiaalministeeriumi poolt edastatud pseudonüümiga, aga ilma isikukoodideta) Tartu Ülikooli teadusarvutuste keskusesse.

3.1.5. Andmete ühildamine ja ühendamine.

Tartu Ülikooli uurijad kasutavad pseudonüümi, et linkida kokku pseudonümiseeritud andmebaasid. TÜ uurijad ei saa avada andmevõtit ega näha isikukoode.

3.1.7. Andmete muutmine, lisamine

Projekti käigus me ei kontakteeru uuritavatega ega paku neile aktiivselt võimalust oma andmeid muuta, täiendada või lisada. (Küll on see võimalus uuritavatel siiski olemas, vt 3.3.5)

3.1.8. Kuidas tagatakse andmete ajakohasus?

Andmeid ei värskendata. Selle asemel keskendub projekt küsimusele: „kui kasulikud on ühe ajahetke andmed“ ja vastab sellele küsimusele eri ajahetkete lõikes.

3.1.9. Andmesubjekti nõusolek

Andmesubjektilt ei küsita nõusolekut. Töötlemine toimub Isikuandmete kaitse seaduse §6 alusel (Isikuandmete töötlemine teadusuuringu vajadusteks).

3.1.10. Nõusoleku tagasivõtmine

Uuritavate aktiivne teavitamine ei ole praktiliselt teostatav, mistõttu on uuringalane teave leitav aadressil https://tervis.ut.ee/et/teadus. Seal on teavitatud uuritavatel ka võimalik uuringumeeskonnale kirjutada, et nende andmeid uuringus ei kasutataks.

3.2. Isikuandmete säilitamine

Kõik andmeid säilitatakse Tartu Ülikooli teadusarvutuste keskuses vastavas projekti põhises kaustas, millele on juurdepääs vaid eetikaloal mainitud isikutel. Vajadusel säilitatakse andmed krüpteeritud kujul.

3.2.1. Säilitamise tähtajad

Kõik algandmed ja tuletatud andmed kustutatakse kas 31.08.2028 või enne seda juhul kui projekti eesmärgid on varasemalt täidetud. Seda otsust seirab Vastutav täitja kaasprofessor Taavi Tillmann, kes igakuiselt hindab kas projekti eesmärgid on täidetud või mitte. Andmete kustutamise käigus kustutatakse ka kõik andmete varukoopiad.

3.2.2. Hoiustamisel kasutatavad turvameetmed (ISKE)

Tartu Ülikooli teadusarvutuste keskus järgib infrastruktuuri haldamisel ISKE M taseme nõudeid. Andmeid ja vaheandmeid hävitatakse vajadusel vastavalt ISKE H turbeastmega andmete hävitamise nõuetele. Tartu Ülikooli teadusarvutuste keskus 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 Tartu Ülikooli teadusarvutuste keskus 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.

3.2.3. Kuidas välditakse andmete juhuslikku hävimist või kahjustumist?

- Andmetest tehakse regulaarseid varukoopiaid.
- Regulaarselt teostatakse serverite turvatestimist, uuendamist ja seiramist.
- Serverid on varustatud katkematu toiteallikaga (UPS).
- Kasutusel on füüsilised ja organisatoorsed turvameetmed, mis takistavad selleks volitamata isikute füüsilist juurdepääsu serveritel.
- 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.

3.2.4. Logid, turvakoopiad.

Andmete töötlus toimub SAPU keskkonnas, mis rangelt piirab kasutajate võimalust andmeid serverist välja kopeerida ning võimaldab vastutavatel uurijatel jälgida kõiki kasutajate tegevusi. Logitakse kõik kasutaja tegevused SAPU keskkonnas ning lisaks salvestatakse jooksval ka kasutaja ekraanipilti (video). Kasutajad on logidest ja ekraanipildi salvestamisest teadlikud ning see toimib ka heidutava meetmena.

Andmetest tehakse üks kord nädalas regulaarseid varukoopiad vastavalt Tartu Ülikooli teadusarvutuste keskuse varundamise korrale. Varundamine toimub Tartu Ülikooli teadusarvutuste keskuse 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.

3.2.5. Arhiveerimine

Projekti raames toimub andmete varundamine, kuid ei toimu andmete arhiveerimist.

3.2.6. Arhiiviväärtusega andmed: Antud projektis puuduvad. Projekt arhiveerib kõik riskiennustusmudelid, R-koodi skriptid ja muud tööriistad, mis võivad osutuda kasulikuks sarnaste projektide teostamisel tulevikus.

3.3. Isikuandmete kasutamine

3.3.1. Millal ja kelle poolt andmeid kasutatakse?

Andmetele on juurdepääs ja andmeid kasutavad vaid eetikaloal loetletud uurijad. Nad kasutavad neid alates sellest hetkest millal esimesed andmed laekuvad Tartu Ülikooli teadusarvutuste keskusesse (nt 01.10.2023) kuni projekti lõppemiseni (hiljemalt 31.08.2028 või varem kui kõik projekti eesmärgid on varem saavutatud).

3.3.2. Kuidas andmeid kasutatakse

Andmed laekuvad Ülikooli vaid pseudonüümitud kujul ja neid kasutatakse vaid teaduslike analüüside tegemiseks ja riskikalkulaatori loomiseks, alati pseudonüümitud kujul. Seda tehakse R-i keskkonnas Tartu Ülikooli teadusarvutuste keskuses. Täpsemad sammud on siduda kõik andmebaasid omavahel ja analüüsida millised e-sotsiaaltegurid ennustavad südame- ja veresoonkonnahaiguste teket.

3.3.3. Päringute tegemine ja andmete lugemine

Uuringu jooksul ei tehta päringuid üksikindiviidide andmete kohta ja ei loeta selliseid andmeid.

3.3.4. Kättesaadavaks tegemine. Avalikustamine

Andmeid ei tehta kättesaadavaks ja ei avalikustata.

3.3.5. Juurdepääs andmetele ja kättesaadavuse piiramine

Andmetele saavad ligi vaid eetikaloal loetletud uurijad.

3.3.6. Kuidas andmesubjektil on võimalik saada väljavõte andmetest, mida tema kohta on kogutud? Mis kujul see väljavõte antakse? Kas ta saab ebatäpseid andmeid parandada?

Andmesubjekt saadab vastava avalduse vastutava uurija e-postile. Uurita täpsustab temaga, kas ta annab nõusoleku, et seni pseudonüümsetes andmetes (kus me ei saa teda tuvastada) soovib ta vaid ennast üksinda tuvastada. Kui vastavalt allkirjastatud avaldus on laekunud, esitab vastutava uurija antud kodaniku isikukoodi rahvastikuregistri omanikule (Siseministeerium) kes on ainus kes saab avada andmevõtit. Siseministeerium leiab vastava kodaniku pseudonüümi ja edastab selle Tartu Ülikooli vastutava uurija (Taavi Tillmannile). Vastutav uurija siseneb Teadusarvutuse keskusesse, tuvastab kõik andmed mis on seotud antud andmevõtmega ja kopeerib need sobivasse formaati (nt XLS ja/või PDF). Väljavõtte antakse uuritavale eelistatud PDF formaadis, eeldusel, et formaadi vahetus ei vähenda oluliselt andmete arusaadavust. Vastavalt uuritava eelistusele võib andmevahetus toimuda kas e-kirja või posti teel.

Juhul kui uuritav soovib ebatäpsete või eksitavaid andmeid parandada, siis eelisjärjekorras proovime seda üheskoos parandada algandmetes (nt soovitame uuritaval siseneda rahvastikuregistrile, kirjutame tema perearstile vms.), mida peegeldame ka meie andmetes. Kui see osutub võimatuks, parandame andmeid vaid meie andmetes ja selgitame uuritavale miks algandmed jäid muutmata.

3.3.7. Kuidas välditakse loata või ebaseadusliku juurdepääsu eest?

Andmete töötlus toimub spetsiaalsel tundlike andmete analüüsiplatvormil (SAPU, https://docs.hpc.ut.ee/public/services/SAPU/) vaid selleks uuringuks ülesseatud virtuaalserveril, mis piirab kasutajate võimalust andmeid serverist välja kopeerida ning võimaldab vastutavatel uurijatel jälgida kõigi kasutajate tegevust.

Andmetele ligipääs on piiratud üksnes käesolevas taotluses nimetatud uurimisgrupi liikmetele. Vältimaks juhuslikke rünnakuid internetist, on kõik kasutatavad serverid kättesaadavad ainult Tartu Ülikooli sisevõrgust. Servereid haldab Tartu Ülikooli teadusarvutuskeskus, mille turvameetmeid ja protsess on hinnatud ISKE-M tasemele.

Loata juurdepääs on teoreetiliselt võimalik vaid siis, kui kõrvaline isik saab administraatori õigustes juurdepääsu Tartu Ülikooli teadusarvutuste keskuse infrastruktuurile, suudab luua fiktiivse konto, teab täpselt millise alamkaustaga seda siduda, saab juurdepääsu ja oskab seejärel pseudonüümsetes andmetes orienteeruda. Juhul kui selline isik peaks üritama tuvastada ühte isikut isikutasandil peaks ta tegema suure hulk analüüse, raporteid ja võrdlusi (eelistatud teiste väliste andmebaasidega mida tal oleks sisuliselt võimatu isikutasandil siduda), mis võtaks palju aega.

Uuringu meeskond teostab regulaarseid auditeid, et millised kasutajatunnused on antud kataloogile sisse pääsenud. Kui uued või tundmatud kasutajatunnused seal ilmnevad, see tõstatab häiret uurimaks kas ebaseaduslik juurdepääs võis toimuda.

Juhul, kui parasjagu ei toimu aktiivset analüüsi faasi, siis on SAPU keskkond välja lülitatud ja keskkonnale ei pääse keegi juurde.

3.4. Isikuandmete edastamine

3.4.1. Kas andmeid edastatakse?

Vastutav töötleja (Tartu Ülikool) ei edasta andmeid kolmandatele isikutele, asutustele või riikidesse.

3.4.2. Kas andmesubjekt saab taotleda oma andmete ülekandmist teisele töötlejale.

Ei, kuna andmesubjekt ei esitanud isikuandmed omaenda nõusoleku alusel ja töötlemine pole vajalik lepingu täitmiseks. Lisaks ei ole me teadlikud teistest töötlejatest kes suudaksid teha vastavat töötlemist.

3.5. Isikuandmete kustutamine

3.5.1. Millal ja kuidas andmeid kustutatakse?

Kõik algandmed ja tuletatud andmed kustutatakse kas 31.08.2028 või enne seda juhul kui projekti eesmärgid on varasemalt täidetud. Kustutamist teostab Teadusarvutuse keskuse meeskond koostöös vastutava uurijaga (Taavi Tillmann), et tagada permanentne ja pöördumatu kustutamine.

3.5.1. Hävitamine, blokeerimine

Kõik algandmed ja tuletatud andmed hävitatakse kas 31.08.2028 või enne seda juhul kui projekti eesmärgid on varasemalt täidetud. Projekti käigus ei teki paberdokumente ning andmeid ei trükita välja.

3.5.2. Kas andmesubjekt ise saab kasutada õigust taotleda oma andmete kustutamist (õigus „olla unustatud“)?

Jah. Vastav protsess on just nagu kirjeldatud ülaltoodud sammus 3.3.6

3.5.3. Ebaõigete andmete muutmine või kustutamine?

Antud projekti käigus me ei kontakteeru uuritavatega ega paku võimalust neil andmeid kontrollida muuta, täiendada, lisada või kustutada.

3.5.4. Kustutamise dokumenteerimine (logi selle kohta).

Andmete kustutamise dokumenteeritakse ja andmete kustutamise kohta koostatakse andmete kustutamise akt.

4. Isikuandmete töötlemise eesmärgid

4.1.1. Töötlemise eesmärgid

Antud projekti eesmärgid on:

1) Tuvastada milline kogum e-sotsiaalandmeid kõige paremini ennustab südame veresoonkonna haiguste (SVH) 10.a. teket tervete inimeste seas?
2) Analüüsida, kas ülaltoodud ennustust on sama täpne teha kasutades selleks kas 01.01.2012, 01.01.2014 või 01.01.2016. hetkel kogutud andmeid?

4.1.2. Kas ja kuidas eesmärkidele vastavus tuvastatakse organisatsiooni poolt enne andmete kogumist või andmete kogumise ajal?

Esimeses faasis, enne andmete kogumist enamik andmebaaside omanike käest tuvastatakse vastavus andmete ja eesmärkide vahel seeläbi, et taotlus peab saama **Tartu Ülikooli inimuuringute eetika komitee** poolt kooskõlastuse. Nemad hindavad nad isikuandmete kasutamist kõigist mainitud andmebaasidest.

4.1.3. Kuidas tagatakse, et andmeid töödeldakse üksnes neil eesmärkidel, milleks neid kogutakse?

Kõik uurimisgrupi liikmed on kas lepingulised Tartu Ülikooli töötajad, kelle töölepingus sisaldub muuhulgas konfidentsiaalsusklausel ja kel on sarnaste uuringutega varasem kogemus või üliõpilased, kellega sõlmitakse konfidentsiaalsusleping. Kõigile uurimisgrupi liikmetele on selgitatud uurimistöö sisu, andmekaitse- ja eetikanõudeid. Sealhulgas selgitatakse liikmetele, et teistel eesmärkidel andmeid töödelda ei tohi (ja kuidas vastava soovi korral saame esitada lisa taotlusi eetika komiteedele). Uurimistöö läbiviimisel lähtutakse muuhulgas Tartu Ülikooli juhendist “Andmekaitse teadustöös" (https://wiki.ut.ee/pages/viewpage.action?pageId=196183311). Valdav osa uurimismeeskonnast omab varasemat sarnase sisuga terviseuuringute läbiviimise kogemust kus oleme andmeid töödelnud vaid kokku lepitud eesmärkide raames.

4.1.4. Kas andmeid töödeldakse ka muul viisil, mis on esialgsete töötlemise eesmärkidega vastuolus?

Ei

4.2. Töötlemise vajalikkus ja proportsionaalsus

4.2.1. Mis põhjustel töödeldakse andmeid töötlemise eesmärkide saavutamiseks just sel viisil nagu seda praegu tehakse? Kas eesmärkide saavutamiseks on võimalik andmeid töödelda ka mõnel muul viisil? Kas on olemas lihtsam viis (tavaliselt see tähendab ka vähem riski põhjustav) eesmärkide saavutamiseks? Oluline on siin täpselt ja argumenteeritult selgitada, miks andmeid töödeldakse just sel viisil ja sellises ulatuses!

Kokkuvõtteks, andmete laiuse ja sügavuse valik tuleneb uurigu eesmärgist ja vajalikus valimimahust. Südame- ja veresoonkonna haigustel on mitmeid riskitegureid. Sellest tulenevat võib mõne olulise riskiteguri arvestamata jätmine viia ekslike uurimistulemusteni

Detailsemalt, kaalusime mitmeid erinevaid lahendusi enne käesoleva töötlemise plaani valimist. Muu hulgas:

Kaalusime varianti analüüsida igat andmebaasi **ükshaaval** ning omavahel sidumata. See võimaldaks hinnata teaduskirjanduses kirjeldatud riskitegurite olemasolu andmebaasides. Selline lähenemine ei võimalda uurijatel südame- ja veresoonkonnahaiguste riskitegureid tuvastada. Kaalusime varianti, et iga **üks sotsiaalandmebaas on seotud vaid ühe tervise andmebaasiga** aga mitte üksteisega. See võimaldaks tuvastada üksikute riskitegurite riskisuhteid haigustekkele. Kuna riskitegurid on omavahel korreleerunud, siis oleks need riskisuhteid kallutatud ülespoole (st tegelik riskisuhe oleks madalam). Kallutatud riskitegurite lisamine ühte riskiennustusmudelisse on praktiliselt võimatu kuna see ülehindaks riski mitmekordselt (kui mitte enam) ja suurendaks ebatäpsust eri gruppide vahel viisil, mida ei ole võimalik ennustada ega korrigeerida.

A) Kaalusime varianti, kus osa andmete hägustamisest ja sidumisest teostab Tartu Ülikooli asemel **Statistikaameti** teadlase keskkond. Paraku vastasid nemad, et nende keskkond on liiga väike (kuni 20GB) meie projekti otstarbeks (vajame ca 900 GB). Lisaks on OMOPi tarkvara teisaldamine Tartu Ülikoolist mujale keerukas protsess, mis ei pruugi olla praktiliselt teostatav.
B) Kaalusime varianti, kus andmed kunagi ei tooda kokku aga **analüüs toimub nn födereeritult** (ingl k *federated learning*) (st saadetakse igale andmebaasi omanikule analüüsimiseks). Paraku ei ole selline tarkvara meie teada kasutuses, mis võimaldaks meil jooksutada analüütilist koodi, mis ühe tehinguga optimeeriks parameetreid üle 13. andmebaasi. Isegi kui see oleks tehtav, kõige mahukam samm on e-tervise andmete teisaldamine OMOP formaati mis käesoleval hetkel on teostatav ainult Tartu Ülikoolis. Ilma OMOPi formaadita on e-tervise andmed selle projekti eesmärkide vaatenurgast igati kasutuskõlbmatud.

4.2.2. Minimaalsuse põhimõte (andmeid kogutakse üksnes ulatuses, mis on vajalik andmetöötluse eesmärkide täitmiseks)

Kuigi andmekooseis paistab pikk, tegelikult on seal igal andmeväljal otsene vajadus, et täita ühte kolmest funktsioonist:

1) Tuvastada võimalikult täpselt, kes on uuringu algul **haige**, kas neil on südamehaigust-ennustavaid kaasnevaid haigusi, ja kas neil tekib südamehaigus uuringu jooksul. Kui siinkohal lubada ebatäpsust (mis küll langetaks isiku tuvastamise ohtu), suurendab see tõenäosust, et loodud riskikalkulaatori abil sekkuvaid ravimeid soovitatakse välja valedele inimestele.
2) Tuvastada võimalikult täpselt südamehaiguse **riskitegureid,** mis kirjanduses alusel on võivad ka Eestis teoreetiliselt ennustada südamehaiguse teket. Kui siinkohal valida vähem riskitegureid (mis küll langetaks isiku tuvastamise ohtu) siis paralleelselt suurendaks see loodud riskikalkulaatori ebatäpsust ja võimalust, et jätku uuringud soovitavad sekkumist ravimitega valedele inimestele.

4.2.3. Täpsus, täielikkus ja ajakohasus

Vt. "3.5.3. Ebaõigete andmete muutmine või kustutamine".

Kasutame puuduvate andmete puhul asjakohased meetodeid (sh puuduvus, kui haiguse ennustaja, ja/või puuduvate andmete imputeerimine, vastavalt puudumise mustrile). Teades, et andmete kvaliteet paraneb ajas uurime, kas riskikalkulaatorite täpsus tõuseb, kui neid luua hilisemal aastal (2014) vs. varasemal aastal (2012).

Me teame, et algandmed ei pruugi olla täpsed. Samas kui kõrvutada selliste andmete alusel teenuse pakkumist olemasolevaga, kus me sisuliselt eeldame, et me ei tea mitte midagi ühegi elaniku kardiovaskulaar riski kohta (st. AUROC=0.50) ja ei paku kellelegi mitte midagi, siis sellise madala referentspunktiga võrreldes usume, et ka osaliselt ebatäpsete andmete pealt haiguste ennustamine võib pakkuda märkimisväärset täpsuse tõusu. Teisisõnu, isegi osaliselt täpsete andmete puhul näeme, et klaas on ikkagi pooltäis võrreldes tühja klaasiga.

5. Riskid

Riski number	Riski nimetus	Riski tõenäosus	Riski mõju	Riski tase	Ettepanek riski maandamiseks
1	**Andmete kasutamine muuks otstarbeks kui esialgselt kogutud**. See muudab siseringi kultuuri ja/või andmeid avalikustatakse. Ei tuvasta uuritavaid.	3: Võimalik	1:Vähe-tähtis	1: Madal	Maandada, kasutusele võtta näost-näkku ennetavad vestlused alluvatega; tagada õige eeskuju; keelata, taunida ja proportsionaalselt karistada andmete kasutamine muuks otstarbeks, kui esialgselt kogutud
2	Tundmatu isik saab **ligipääsu** kõigile andmetele. Ei tuvasta uuritavaid	1: Harva-esinev	3: Mõõd-ukas	1: Madal	Aktsepteerida
3	Uurija või tundmatu isik suudab **tuvastada** mõned **üksikud** uuritavad (näiteks kõrvutades lisa andmebaasidega mis pole kooseisus). Võimalik kahju uuritavale.	1: Harva-esinev	4: Suur	1: Madal	Aktsepteerida
4	Üks üheksast andmevõtme võimalikest avajatest suudab endale saada piisavalt andmebaase teistelt omanikest, misläbi suudab ta kiiresti **tuvastada kõik** uuritavad. Võimalik kahju uuritavatele.	1: Harva-esinev (praktiliselt võimatu)	4: Suur	1: Madal	Maandada: Hoiatada kõiki andmevõtme avajaid, et nad ei tohi andmeid üksteisele saata ja selle tagajärgedest. Nad peavad saatma vaid Tartu Ülikoolile, kes ei saa võtit ise.
	Riskiennustuskalkulaator pole päriselus nii täpne nagu arvame, selle omadused või protsess ei meeldi uuritavatele jne.	Väljaspool skoopi.	Seda hilisemat riski maandab teaduslik protsess (st. *peer review*) pluss hiljem loodava tehnoloogia seejärgne hindamine regulaatorite, rahastajate ja tervishoiuteenuste osutajate poolt. Käsitleme neid riske jätku-projektides.		

6. Kasutusel olevad riskide vältimise meetmed

Kirjelda üldiselt, kuidas organisatsioonis tegeletakse riskide vältimisega. Kas sellega tegeleb konkreetne ametikoht jms.
Organisatsioonis vastutab riskide vältimise eest Vastutav uurija (Taavi Tillmann). Ta saab abi ja sisendit kolleegidelt, ülemuselt ja vajadusel ülikooli siseauditi büroo andmekaitse spetsialistidelt.

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 (Taavi Tillmann), 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 süsteetilised 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 elektroonine valve on deaktiveeritud ainult siis, kui keegi asub füüsiliselt andmekeskuses, on sinna sisenemas või sealt lahkumas.
- Kõik andmekeskusse sisenemised ja elektroonilise valve desaktiveerimised/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 Tartu Ülikooli teadusarvutuste keskuse infrastruktuuril SAPU keskkonnas, kus:

- Jälgitakse Eesti infoturbestandardiga kehtestatud nõudeid.
- Teostatakse regulaarselt serverite testimist, uuendamist ja seiramist.
- Haavatavuste tuvastamiseks kasutatakse seiramist, masinõpet ning ka erinevaid läbistusteste. Muuhulgas kasutatakse haavatavuste tuvastamiseks ka juba olemasolevaid haavatavuste tuvastamise tarkvarasid (N: Nessus) ning jälgitakse järjepidevalt erinevaid haavatavuste nimekirju. Lisaks Tartu Ülikooli teadusarvutuste keskusele skaneerib Tartu Ülikooli teadusarvutuste keskuse 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 Tartu Ülikooli teadusarvutuste keskuse lindirobotile, mis asub füüsiliselt teises asukohas (Tartu Ülikooli teadusarvutuste keskuse andmekeskuses).
- Kõik kasutajate tegevused SAPU keskkonnas logitakse.
- Kõik potentsiaalsed turvaintsidendid ja turvanõrkuste leidmise katsed logitakse (N: 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 kolmanda isiku (andmete omaniku) nõusolekut.
- SAPU keskkondi varundatakse regulaarselt.
- Perioodidel, kui SAPU keskkonda ei kasutata, on keskkond välja lülitatud ja keskkonda ei ole võimalik siseneda.
- 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).