



## Innovatsiooniprojekti ideekavand<sup>1</sup>

### AVALIKU SEKTORI INNOVATSIOONIVÕIMEKUSE TÕSTMINE

Ideekavandit täites palume tutvuda riigikantselei veebilehel toodud [soovituste ja juhistega projekti esitajale](#).

Innovatsiooniprojekti nimi	TI-valmis andmestikud. Teadmuskihit tehisintellekti jaoks.
Innovatsiooniprojekti fookusvaldkond	<input type="checkbox"/> Droonitehnoloogiate valdkond <input checked="" type="checkbox"/> Tehisintellekti lahenduste valdkond
Innovatsiooniprojekti panus valitsuse tegevuskava prioriteetidesse	<input type="checkbox"/> Riigi kriisikindluse suurendamine <input checked="" type="checkbox"/> Majanduse kasvule kaasa aitamine <input checked="" type="checkbox"/> Riigi tõhus juhtimine
Innovatsiooniprojekti esitajad (tulevased RK partnerid) (asutus/asutused) <sup>2</sup>	Statistikaamet
Projektijuht või ideekavandi esitaja kontaktisik (nimi, asutus, e-posti aadress ja telefon)	Ideekavandi esitajad: Terje Trasberg (andmehõive ja tööluse osakonna juhtaja, terje.trasberg@stat.ee) ja Veiko Berendsen (andmehalduse juhtivekspert, veiko.berendsen@stat.ee)
Innovatsiooniprojekti kestus (kuudes)	18 kuud <i>Ajaarvestust alustame üldjuhul partnerluslepingu sõlmimisest.</i>
Innovatsiooniprojekti kogumaksumus (sh käibemaks, kui on abikõlblik)	Kogumaksumus: 1 076 400 eurot Partner: Statistikaamet
Käibemaks	<input checked="" type="checkbox"/> jääb kulu tegija kanda (käibemaks abikõlblik) <input type="checkbox"/> saab küsida riigilt tagasi (käibemaks ei ole abikõlblik)  <i>Vastav info täita iga partneri kohta (kopeeri ridu ning kirjuta partneri nimi juurde)</i>

<sup>1</sup> Juhul kui ideekavand on mõeldud **asutusesiseseks kasutamiseks**, siis lisage vastav alus ideekavandi päisesse.

<sup>2</sup> **Partner EL struktuurivahendite mõttes**, kes viib ise läbi innovatsiooniprojekti elluviimisega seotud hanked, sõlmib lepingud ning vastutab aruandluse eest.

## 1. Probleemikirjeldus (max 2 lk)

**Kirjeldage lahendamist vajavat probleemi, selle olulisust ning keda see probleem puudutab.**

- Selgitage, miks on probleem aktuaalne.
- Hinnake probleemi mõju (nt rahaline kokkuhoid, keskkonna- või sotsiaalne kasu). Kirjeldage probleemi tausta. Mida on probleemi lahendamiseks Eestis juba tehtud või mis on tegemisel? Tooge välja relevantsed teiste riikide kogemused probleemi lahendamisel.

Eesti kui digiriigi üheks suurimaks rikkuseks on avaliku sektori andmerohkus. Riigi infosüsteemides, registrites, avaandmete portaalides, statistika andmebaasides ja valdkondlikes rakendustes tekib igapäevaselt suur hulk andmeid, mida kasutatakse avalike teenuste arendamiseks, poliitikakujundamiseks, teadustöök ja ühiskonna arengute jälgimiseks. Andmepõhise riigi järgmine arenguhüpe ei seisne aga enam ainult selles, kas andmed on olemas, vaid selles, kas neid on võimalik kiiresti leida, õigesti mõista, usaldusväärselt tõlgendada ja taaskasutada – nii inimeste kui ka tehisintellekti lahenduste abil.

Suuri keelemudeleid ehk LLM-e kasutatakse üha enam info leidmisel, kokkuvõtmisel ja otsuste ettevalmistamisel. Need sobivad hästi tekstilise info töötlemiseks, kuid arvandmete ja struktureeritud andmestike puhul on nende kasutamine endiselt piiratud. Tabelid, andmebaasid, klassifikaatorid, statistilised näitajad ja registriandmed ei ole keelemudelitele iseenesest arusaadavad. Ilma täiendava kontekstita ei pruugi mudel leida õiget andmestikku, mõista näitaja tähendust, eristada sarnaseid mõisteid, arvestada meetodikat ega kontrollida andmete ajakohasust. Selle tulemusena võivad tehisintellekti lahendused anda arvandmete kohta ebatäpseid, aegunud või eksitavaid vastuseid.

Probleemi tuum ei ole ainult andmete otsimises, vaid selles, et andmete, nende tähenduse ja konteksti vahel puudub sageli masinloetav seos. **Selleks, et tehisintellekt saaks avaliku sektori andmeid usaldusväärselt kasutada, on vaja semantilist ehk teadmuspõhist kihti, mis seob andmestikud, metaandmed, mõisted, klassifikaatorid, meetodika, kvaliteediinfo ja selgitava dokumentatsiooni.** Nii saab tehisintellekt aru, mida konkreetne andmestik või näitaja tähendab, kuidas see on tekkinud, milliste teiste andmetega see seostub ning millised piirangud selle kasutamisel kehtivad.

Eestis on selleks mitmed olulised eeldused juba olemas. Statistikaameti koordineerimisel on avaliku sektori asutused koostanud andme- ja ärisõnastikke, mis kirjeldavad andmete sisu ühtse standardi alusel ning on graafina esitatavad tesaurustena. Loodud on Eesti andmekirjelduse standard ning Eesti andmete teabevärv, mis koondab infot andmekogude ja avaandmete kohta. Statistikaameti on olemas kogemus metaandmete, meetodilise dokumentatsiooni, andmekirjelduste ja sõnastike haldamisel ning tehtud on esimesi katsetusi tehisintellekti kasutamiseks andmekirjelduste koostamisel. Seega on olemas alus – andmekirjeldused, sõnastikud ja metaandmed –, kuid järgmine samm on siduda need ühtseks teadmuskihiks, mida saavad kasutada nii inimesed kui ka tehisintellekti lahendused.

Samas on lahendamata mitu olulist küsimust. Eesti andmete teabevärv kasutab tavapäraselt otsingut ega toeta veel tehisintellektil põhinevat andmete leidmist. Kuigi teabevärv sisaldab suurt hulka andmestikke, sh avaldatud riiklikku statistikat, ei võimalda see vastusena esitada andmete väärtusi koos nende tähenduse ja kontekstiga. Lisaks paikneb suur osa andmetest, meetodikast ja selgitavast infost endiselt eri portaalides, rakendustes ja dokumentides, mistõttu ei ole need kasutaja ega tehisintellekti vaatest ühtselt leitavad ja tõlgendatavad.

Projekti aktuaalsus tuleneb sellest, et tehisintellekti kasutamine avalikus sektoris kasvab kiiremini kui andmete masinmõistetavaks muutmise võimekus. **Kui seda lünka ei lahendata, jäävad TI-lahendused avaliku sektori andmete kasutamisel piiratud ja ebausaldusväärseks.** Kui aga

andmed, metaandmed ja mõisted on omavahel seostatud ning tehisintellekti jaoks kasutatavad, loob see aluse loomulikus keeles andmepäringutele, automaatsetele kokkuvõtetele, mitme andmeallika ühendamisele, andmekvaliteedi kontrollile, RAG-lahendustele struktureeritud andmetel ning tulevikus ka usaldusväärsetele AI-agentidele ja otsustustoe lahendustele.

Käesoleva projekti eesmärk ei ole luua üksnes uut otsingulahendust. Otsing on üks võimalik kasutusjuht, mille abil saab teadmuskivi toimimist näidata. **Projekti fookus on tehnilise ja organisatsioonilise raamistiku loomisel, mis võimaldab avaliku sektori struktureeritud andmestikke muuta tehisintellekti jaoks kasutatavaks.** Raamistik kirjeldab, milliseid andmekirjeldusi, sõnastikke, metaandmeid, tehnilisi komponente, rolle, vastutusi ja haldusprotsesse on vaja, et sellist teadmuskivi lahendust oleks võimalik riigis järjepidevalt luua ja hallata.

Raamistik luuakse kogu avaliku sektori vajadust silmas pidades, kuid seda katsetatakse Statistikaameti andmestike ja rakenduste peal. Statistikaamet sobib selleks hästi, sest siin on olemas väga erinevat tüüpi andmeid, rakendusi ja neid kirjeldavat teavet: avaldatud riiklik statistika, juhtimislauad, ERIKA andmeteenus, meetodid, klassifikaatorid, andmekirjeldused, sõnastikud, ulatuslik dokumentatsioon ja erinevad rakendused. See võimaldab testida, kuidas teadmuskivi lahendus toimib olukorras, kus andmed ja neid selgitav info paiknevad eri süsteemides, kuid kasutaja või tehisintellekti vaatest peaksid need moodustama ühe arusaadava terviku.

Projekti skooabis ei ole isikuandmete ega muude piiratud juurdepääsuga andmete sisuline kasutamine tehisintellekti lahendustes. Andmeväärtusteni liigutakse ainult andmestike puhul, mille kasutamiseks on olemas õiguslik alus ja tehniline võimalus, näiteks avaldatud statistika. Piiratud juurdepääsuga andmete puhul kasutatakse üksnes metaandmeid: lahendus saab viidata andmestiku olemasolule ja kasutustingimustele, kuid ei ava ega töötle andmete sisu.

Projekt mõjutab laia ringi inimesi, eristame kolme peamist gruppi:

- **Analüütikud, teadlased ja andmepetsialistid** kulutavad täna palju aega andmete, nende kirjelduste, meetodilise tausta ja eri andmestike seoste leidmisele. Näiteks noorte tööturul osalemise analüüsimiseks tuleb teada, millised andmestikud on sobivad, kus need asuvad ning millised metaandmed, meetodika ja definitsioonid neid selgitavad.
- **Otsustajad, poliitikakujundajad ja valdkonnaekspertid** vajavad kiiresti usaldusväärseid vastuseid, mis põhinevad kontrollitud andmetel. Täna sõltuvad nad sageli analüütikute abist ka avaldatud statistika leidmisel, tähenduse mõistmisel ja eri allikate kokkuviiamisel. *Näiteks Eesti ekspordi andmete kasutamisel peab teadma, et eraldi avaldatakse kaupade ja teenuste ekspordi, otsustama, kas arvestada reeksporti, ning mõistma, milline meetodika on konkreetse küsimuse jaoks sobiv. Rahvusvaheliste võrdluste puhul tuleb arvestada ka meetodiliste erinevustega. Selliste valikute tegemine eeldab täna statistikaalaseid teadmisi.*
- **Andmehaldurid, andmeomanikud ja metaandmete koostajad** peavad looma ja haldama andmekirjeldusi, sõnastikke, klassifikaatoreid ja muud metaandmestikku, mis võimaldab andmeid leida, mõista ja taaskasutada. Praegu on suur osa sellest tööst käsitsi tehtav ning metaandmete kvaliteet ja detailsus on asutuste lõikes ebaühtlane. Olemasolevate kirjelduste sidumine kiiresti arenevate tehisintellekti lahendustega nõuab uusi töövõtteid ja tehnoloogilisi lahendusi.

Mõju avaldub eelkõige tööaja kokkuhoius, andmete paremas taaskasutuses ja kiiremas otsustusprotsessis. Kui vajalik teave on paremini leitav, seostatud ja kasutatav ka TI-põhiste tööriistade kaudu, väheneb käsitöö maht ning suureneb avaliku sektori andmete sisuline kasutus. Tulevikus loob see eelduse ka AI-agentidele ja otsustustoe lahendustele, mis tuginevad usaldusväärsetl Eesti andmestikele ning nende tähendust kirjeldavale teadmuskihile.

### **Rahvusvaheline kogemus**

Rahvusvaheline kogemus kinnitab, et struktureeritud andmete usaldusväärne kasutamine tehisintellekti lahendustes on üleilmne väljakutse. Sellega tegelevad nii tehnoloogiaettevõtted, näiteks Google ja Anthropic, kui ka avaliku sektori organisatsioonid. Rahvusvahelises statistikakogukonnas arendatakse samal eesmärgil semantilisi mudeleid ja masinloetavaid metaandmeid. Näiteks UNECE HLG-MOS eestvedamisel arendatav Core Ontology for Official Statistics (COOS) loob ametliku statistika ühist semantilist raamistikku. Samuti näitavad valdkondlikud lahendused, näiteks StatGPT-tüüpi katsetused, et üldine keelemudel ei ole ametliku statistika ja struktureeritud andmete usaldusväärseks kasutamiseks piisav.

Kuigi tehnoloogilised standardid ja alused on rahvusvahelised, tuleb Eesti andmestike semantika, kirjeldused, mõisted ja kontekst luua ning siduda kohalikus keeles ja Eesti avaliku sektori andmehalduse põhjal. Seetõttu peab Eesti teadmuskihhi looma ise, tuginedes rahvusvahelistele standarditele ja kogemustele.

### **Statistikaameti kogemus**

Statistikaamet on varem katsetanud tehisintellekti kasutamist andmekirjelduste koostamise toetamiseks. Katsetuse tulemusena selgus, et andmekirjelduste koostamist oli võimalik kiirendada 2–3 korda. Kasutusele võeti RAG-lähenemine ja valdkonnapõhine allikabaas, mis võimaldas näiteks põllumajanduse andmestike kirjeldamisel kasutada ainult selle valdkonna asjakohaseid allikaid. Pilooti katsetati Statistikaameti kasutuses olevatel administratiivandmetel.

Katsetus näitas ka seniseid piiranguid: allikate haldamine oli liiga manuaalne ning piisava terminoloogilise ja mõistelise täpsuse saavutamine keeruline. Tulemused ja rahvusvaheline tagasiside kinnitasid siiski, et valitud suund on strateegiliselt õige. Edasiseks on vaja paremini struktureeritud ja hallatud teadmuskeskonda ehk n-ö treenitud korpust, millele lisanduvad allikabaas, terminihaldus ja paindlikum promptimine.

Käesolev taotlus on jätk sellele piloodile. Projekti eesmärk on lahendada senised puudused ning laiendada lähenemist andmekirjelduste koostamise toetamiselt laiemale kasutusele: andmete, metaandmete, mõistete ja metoodika sidumisele teadmuskihiks, mis toetab nii andmete leidmist, tõlgendamist kui ka kasutamist tehisintellekti lahendustes.

## **2. Projekti eesmärk**

***Sõnastage konkreetne, selge ning mõõdetav eesmärk, mille saavutamist või mitte saavutamist on võimalik hinnata.***

***Tooge välja, kas projekti on eesmärk viia läbi piloot, eksperiment, katsetus või töötada välja prototüüp.<sup>i</sup>***

- Kirjeldage, kuidas plaanite projekti eesmärgi saavutamist mõõta.*

Projekti eesmärk on töötada välja ja katsetada Statistikaameti andmestikel prototüüpi, mis muudab struktureeritud andmestikud tehisintellekti jaoks paremini kasutatavaks. Selleks tuleb:

- Võtta kasutusele ja kohendada Eesti vajadustele rahvusvahelised standardid ning tehnoloogilised lahendused**, mis võimaldavad mõista ja töödelda andmestike struktuuri, sisu ja tähendust.

- **Töötada välja tehniline ja organisatsiooniline toimetumudel** teadmuskihi loomiseks ja haldamiseks Statistikaameti, tuginedes olemasolevatele andmehalduse komponentidele (andmekirjeldused, sõnastikud, metaandmed ning senised AI-katsetused). Toimetumudel kirjeldab, milliseid tehnilisi komponente, rolle, vastutusi, kvaliteedinõudeid ja haldusprotsesse on vaja, et TI-valmis andmestikke oleks võimalik järjepidevalt luua, hallata ja laiemalt kasutusele võtta.
- **Luua ja katsetada lahendusi ning arhitektuurseid komponente, mis võimaldavad siduda struktureeritud andmed, metaandmed, mõistesüsteemid ja selgitava dokumentatsiooni ühtseks teadmuskihiks, mida saab kasutada tehisintellekti lahendustes.**
- Katsetada otsingulahenduse prototüüpi, mis **demonstreerib teadmuskihi kasutamist** ning võimaldab teha avaldatud struktureeritud andmestike sisule loomulikus keeles usaldusväärseid päringuid ja saada vastuseid koos andmete tähendust ning konteksti selgitava teabega.
- **Katsetada võimalusi toetada andmekirjelduste ja metaandmete loomist ning haldamist tehisintellekti abil.**

**Projekti keskmes ei ole isikuandmed ega üksikisikute kohta järelduste tegemine.** Fookus on avaandmestike, metaandmete ja teaduse paremal kirjeldamisel, sidumisel ning kasutamisel.

#### Oodatavad tulemused:

##### Lõppkasutaja vaates

- Statistikaameti veebilehele tulnud kasutaja saab esitada küsimuse loomulikus keeles ning leida ühest kohast nii statistika, selgitava meetoodilise teabe kui ka andmete tähenduse ja konteksti.
- Vastused põhinevad kontrollitud avaandmetel (sh riiklikul statistikal) ning sisaldavad viiteid algallikatele, andmestike kirjeldustele, metaandmetele ja vajaduse korral ka nende aluseks olevate üksikandmete kirjeldustele (mitte üksikandmetele endile). See võimaldab kasutajal kiiresti hinnata, kas küsimusele saab vastata avaldatud statistika põhjal või on vaja liikuda edasi detailsemate ava- või üksikandmete kasutamiseni Statistikaameti teiste teenuste kaudu. *Jätkates varasemat näidet: Kui kasutaja küsib „Kui suur on Eesti eksport?“, selgitab lahendus esmalt, kas küsimus puudutab kaupade või teenuste ekspordi (või mõlemat), kas arvesse võetakse reeksporti ning kas soovitakse kogu Eestist eksporditud kaupu või Eesti päritolu kaupade ekspordi. Rahvusvahelise võrdluse soovi korral aitab lahendus valida sobiva andmeallika ning selgitab, miks võivad näiteks Statistikaameti, Eurostati või maksebilansi andmed mõnevõrra erineda. Vastus sisaldab viiteid algandmetele ja meetoodilistele selgitustele.*

##### Andmehaldurite ja metaandmete koostajate vaates

- Tekib võimalus siduda statistikaandmed, metaandmed, sõnastikud ja selgitav dokumentatsioon ühtseks teadmuskihiks, mida saab kasutada tehisintellekti lahendustes.
- Metaandmete ja andmekirjelduste loomist ning ajakohastamist saab osaliselt automatiseerida tehisintellekti abil, vähendades käsitööd ja parandades kirjelduste kvaliteeti.

#### Väljundid:

1. **Analüüs ja tehniline kontseptsioon**, mis kirjeldab, kuidas rahvusvahelised standardid ja tehnoloogilised lahendused (sh CDIF, RAG, teadmusgraafid jm) on kohandatud Statistikaameti vajadustele ning millised põhimõtted on ülekantavad laiemalt avalikku sektorisse.
2. **Tehniline ja organisatsiooniline toimetumudel**, mis kirjeldab teadmuskihi tehnilist ülesehitust ning selle toimimiseks vajalikku organisatsioonilist korraldust, sh osapoolte rolle, vastutust ja koostöö põhimõtteid. Hinnatakse ka toimetumudeli ülekantavust ning koostatakse ettepanek selle rakendamiseks laiemalt avalikus sektoris.
3. **Toimivad tehnoloogilised komponendid (MVP-d):**
  - metaandmete ja dokumentatsiooni allikabaas RAG-i jaoks;
  - teadmusgraaf (sh andme- ja ärisõnastikud ning muud mõistesüsteemid);

- liidestused suurte keelemudelitega;
  - teadmuskihti haldav rakendus.
4. **TI-toega otsingulahenduse MVP**, mis võimaldab esitada loomulikus keeles küsimusi ühiskonnaähtuste kohta ning saada vastuseid, mis tuginevad Statistikaameti avaldatud riiklikule statistikale ja teistele avaldatud andmetele. Vastused sisaldavad viiteid kasutatud andmestikele, metaandmetele, metoodikale ning selgitavad andmete tähendust ja konteksti.
  5. **TI-toega andmekirjelduste loomise ja rikastamise lahendus**, mis võimaldab olemasolevate andmekirjelduste, sõnastike ja muu dokumentatsiooni põhjal koostada ning rikastada uusi andmekirjeldusi poolautomaatselt.
  6. **Katsetuse tulemuste hinnang ja soovitus**, mis annavad hinnangu lahenduse toimivusele ning kirjeldavad eeldusi ja soovitusi selle rakendamiseks laiemalt avalikus sektoris.

#### Möödikud lahenduse kvaliteedi hindamiseks:

- a) **Lahendus suudab anda usaldusväärseid vastuseid** Statistikaameti andmestike põhjal, kus vastus tugineb kontrollitud andmeallikatele ning sisaldab viiteid kasutatud andmetele, nende tähendusele ja metoodilisele taustale.
- b) **Lahendus suudab kasutada ühes päringus eri teadmusallikaid**, ühendades struktureeritud andmed, metaandmed, sõnastikud ja selgitava dokumentatsiooni.
- c) **Lahendus vähendab kasutaja vajadust teada, millises süsteemis vajalik teave paikneb**. Kasutaja leiab vajaliku info loomulikus keeles esitatud küsimuse abil sõltumata sellest, kas see asub statistikaandmebaasis, veebilehel või muus teadmuse allikas.
- d) **AI abil loodud või rikastatud andmekirjeldused vastavad kvaliteedilt käsitsi koostatud kirjeldustele** ning vajavad ainult piiratud ulatuses inimese täiendusi.
- e) **Ekspertide hinnangul on tehniline ja organisatsiooniline toimemudel rakendatav** ning loob eeldused lahenduse laiendamiseks teistesse avaliku sektori organisatsioonidesse.

### 3. Võimalikud lahendussuunad (max 2 lk)

#### **Kirjeldage võimalikke lahendusi ning tegevusi, millega projekti eesmärk saavutatakse.**

- *Kirjeldage võimalikke lahendussuundi, põhjendage eelistatud lahendussuuna valikut (NB! Valitud lahendussuund ei ole siduv, see võib projekti käigus muutuda).*
- *Kirjeldage probleemi lahendamiseks vajalikke tegevusi, mida antud katseprojekti raames plaanitakse teha.*
- *Selgitage, kuidas lahendust katsetatakse. Selgitage, kuidas läbi viidavat katsetust ja selle edukust hindate.*

#### **Lahenduse kavandamise lähtekohad:**

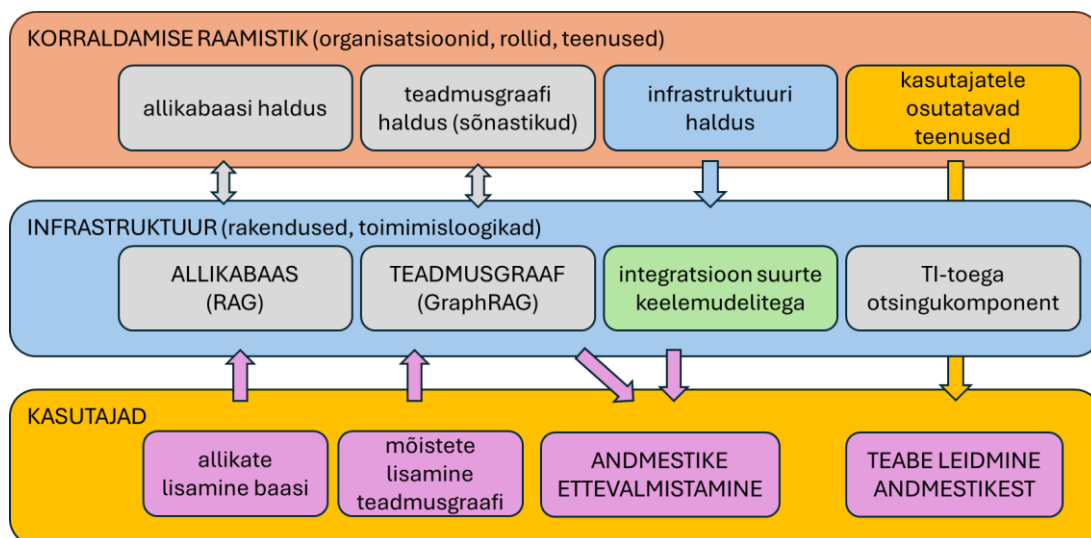
- **Toetume rahvusvahelisele kogemusele ja standarditele.** Eri valdkondi ja lähenemisnurki rahvusvaheliselt ühendab toetumine standarditele. Seetõttu võtame lahenduse kavandamisel aluseks rahvusvahelised standardid ja raamistikud. Toimimisraamistikuna kasutame CDIF-i (*Cross Domain Interoperability Framework*). Tegemist on ulatusliku raamistikuga (nt hõlmab ka juurdepääsuõiguste haldust ODRL-i abil), millest rakendame projekti eesmärkidest ja Statistikaameti vajadustest lähtuvad komponendid.
- **Läheneme valdkondadeülelalt.** Statistikaamet koondab väga erinevate valdkondade statistikaandmeid ning kasutajate küsimused eeldavad sageli nende omavahelist seostamist.

Seetõttu ei kavandata lahendust ühe valdkonna vajadustest lähtudes, vaid selliselt, et see toimiks erinevate statistikavaldkondade puhul. Seda lähenemist toetavad ka rahvusvahelised arengusuunad (nt Microsoft *From Local to Global: A GraphRAG Approach to Query-Focused Summarization*), kus otsitakse lahendusi valdkondliku ja valdkondadeülese teadmuse ühendamiseks (Local → Community → Global). Projekti käigus otsime lahendust küsimusele, kuidas siduda eri valdkondade teadmuskihid tervikuks, säilitades samal ajal nende valdkondliku tähenduse ja eripära.

- **Toetume senitehtule.** Arvestame olemasolevate riiklike andmehalduse komponentidega ning integreerime katsetuses Statistikaameti olemasolevad teadmuse komponendid. Selleks kasutame (1) terminoloogiate ja taksonoomiate haldust (andme- ja ärisõnastikud) ning (2) Statistikaameti olemasolevaid andmeallikaid ja lahendusi, sealhulgas stat.ee veebikeskkonda ja seal paiknevat statistika andmebaasi (<https://andmed.stat.ee/et/stat>) ning erinevaid rakendusi (nt. juhtimislauad, palgarakendus, töötururakendus), ERIKA andmeteenust, ja metoodilist dokumentatsiooni.

### Võimalik lahendustee:

Projekti käigus katsetatakse lahendust Statistikaameti andmestikel. Lahendus kavandatakse modulaarse arhitektuurina, et selle toimivust oleks võimalik hinnata piiratud keskkonnas (Statistikaameti keskkondades) ning edukuse korral laiendada ka teistele avaliku sektori andmestikele. Joonisel on esitatud lahenduse võimalik arhitektuur ja peamised komponendid.



**Lahenduse kavandamisel lähtume modulaarsest lähenemisest.** Projekti eesmärk ei ole ette määrata ühte kindlat tehnoloogilist teostust, vaid töötada välja ja katsetada lahendus, mida saab vajadusel edasi arendada ning mille üksikuid komponente saab aja jooksul välja vahetada või uuendada.

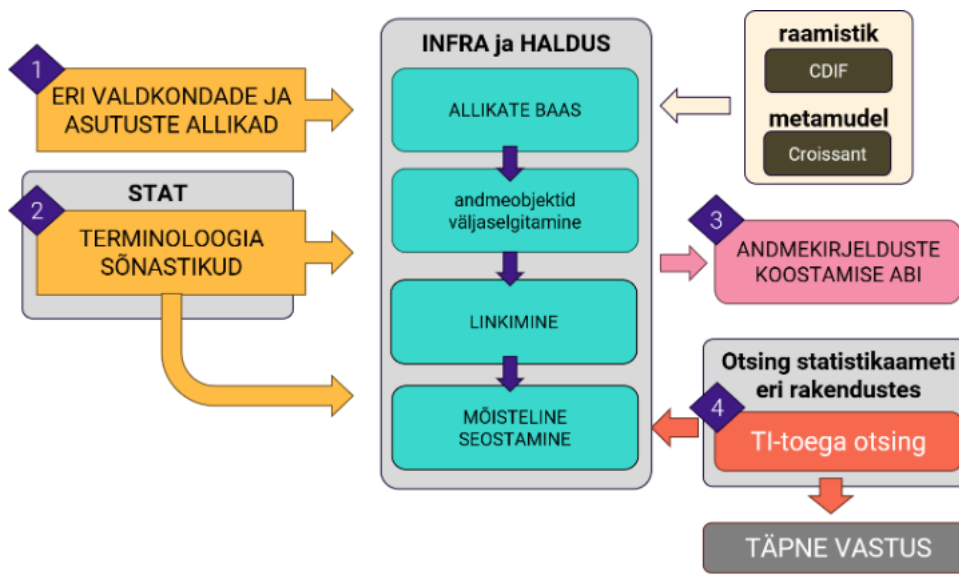
Seejuures käsitleme projekti kahest vaatenurgast. **Esimene puudutab tehnoloogiat ja infrastruktuuri** – eesmärk on luua arhitektuur, mis ei sõltu liigselt ühest konkreetsest tehnoloogiast või teenusepakujast ning suudab kohaneda kiiresti areneva tehisintellekti maastikuga. **Teine puudutab organisatsiooni ja töökorraldust.** Kuigi tehnilist lahendust katsetatakse Statistikaameti andmestikel, töötatakse projekti käigus välja ka tehniline ja organisatsiooniline toimemudel. Selle eesmärk on kirjeldada, milliseid rolle, vastutust ja koostööpõhimõtteid oleks vaja lahenduse rakendamiseks laiemalt avalikus sektoris.

Tehnoloogilises lahenduses on 6 osa:

1. Allikabaasi moodustamine RAG-i jaoks ja selle haldus (võimalikult suur automatiseeritus)

2. Andme- ja ärisõnastike ehk mõistelise teadmusbassi tehnoloogiline lahendus (RDF, SKOS ja teenused)
3. Andmekirjelduste koostamise abi, mis võimaldab andmestike või muutujaid (sama mis tunnus või andmeelement) kirjeldada lõppkasutajale mõistetaval kujul
4. TI-toega otsing, mis kasutab küsimustele vastamiseks arhitektuuril toimivat komponenti ning allikabaasi ja mõisteid
5. Infrastruktuuri ning rakenduste ja teenuste komponent tehnilises mõttes - mis asjadest koosneb ja kuidas töötab
6. Toimimisloogikat määrav ja arendust juhtiv standarditel põhineva lähenemise käsitlus, mis annab korraliku ja jätkusuutliku raami.

Üldine vaade võimalikust lahendusest on järgmine:



#### 4. Projekti uuenduslikkus

**Tuua selgelt välja projekti uuenduslikkus – mida tehakse senisest teisiti, kas see hõlmab uusi tehnoloogiaid, protsesse, toimetamismeetodeid, disaini, turgu vms?**

- Selgitage lahenduse uuenduslikkust nii Eesti kui globaalses kontekstis.
- Mis on projektis sellist, mis vajab katsetamist?

Projekti tehnoloogiline uuenduslikkus ei seisne uute tehisintellekti tehnoloogiate loomises, vaid olemasolevate tehnoloogiate ühendamises viisil, mis võimaldab kasutada struktureeritud statistikaandmeid koos nende metaandmete, mõistesüsteemide ja selgitava dokumentatsiooniga ühe tervikliku teadmuskihina. Selline lahendus võimaldab suurtel keelemudelitel kasutada usaldusväärselt nii andmete väärtusi kui ka nende tähendust, konteksti ja päritolu.

Uuenduslikkus seisneb kahe seni eraldi arenenud suuna ühendamises. Ühelt poolt kasutatakse kaasaegseid tehisintellekti tehnoloogiaid (RAG, GraphRAG ja suured keelemudelid), teiselt poolt rahvusvahelisi andmekirjelduse ja koostoime standardeid (CDIF, Croissant, SKOS, RDF). Projekti käigus ühendatakse need terviklikuks arhitektuuriks, mida katsetatakse Statistikaameti andmestikel.

Eesti avaliku sektori jaoks on uuenduslik see, et sama lahendus kavandatakse toimima erinevatel statistikavaldkondadel. Statistikaamet toodab riikliku statistika programmi alusel eri valdkondade statistikat, kasutades selleks muuhulgas ka erinevate haldusalade andmekogusid, ning projektis katsetatakse nende ühise teadmuskivi loomist.

Kõige olulisem tehnoloogiline katsetus seisneb selles, kuidas siduda tehisintellektiga üheaegselt struktureeritud andmed, metaandmed, mõistesüsteemid ja selgitav dokumentatsioon nii, et need moodustaksid ühe usaldusväärse teadmuskivi. Meie teada ei ole sellist lahendust seni sellisel kujul eri statistikavaldkondi hõlmavalt rakendatud. Projekti käigus katsetatakse eelkõige kahe tehnoloogilise terviku koos toimimist: (1) RAG ja GraphRAG ning (2) CDIF ja Croissant.

## 5. Projekti elluviimisega (katsetusega) seotud riskid ja nende maandamismeetmed

***Kirjelda peamisi riske, mis võivad takistada projekti elluviimist või eesmärkide saavutamist, ning kavanda maandamismeetmed.***

### Tehnoloogilised riskid

**Risk 1:** Projekt ühendab mitut kiiresti arenevat tehnoloogiat (teadmustraafid, RAG/GraphRAG lahendused, semantilised mudelid, keelemudelid ning andmehalduse komponendid), mille koosrakendamiseks ei ole veel välja kujunenud ühtset rahvusvahelist praktikat. Seetõttu võib lahenduse integreerimine osutuda keerukamaks või nõuda projekti käigus esialgsete tehniliste valikute muutmist.

**Maandamismeede:** Lahendus kavandatakse modulaarse arhitektuurina, mis võimaldab üksikuid komponente vajaduse korral välja vahetada või täiendada. Eelistatakse rahvusvaheliselt levinud standardeid ning arendatakse lahendust järk-järgult, valideerides iga komponendi toimivust enne järgmiste etappidega jätkamist.

**Risk 2:** Tehisintellekti tehnoloogiad arenevad väga kiiresti ning suurte tehnoloogiaettevõtete lahendused võivad hakata senisest paremini kasutama struktureeritud andmeid ilma osa projektis kavandatud tehniliste komponentideta. See võib muuta mõne kavandatud tehnilise lahenduse vähem oluliseks.

**Maandamismeede:** Projekti põhiväärtus ei seisne üksikutes tehnoloogiates, vaid Eesti andmestike kirjelduste, metaandmete, sõnastike ja teadmustruktuuride sidumises ühtseks teadmuskihiks ning selle toimimiseks vajaliku organisatsioonilise ja tehnilise raamistikuga. See teadmuskihit jääb vajalikuks sõltumata sellest, milliseid tehnilisi lahendusi tulevikus keelemudelid toetavad. Modulaarne arhitektuur võimaldab vajaduse korral asendada üksikuid komponente ilma kogu lahendust ümber ehitamata.

**Risk 3:** Erinevate teadmuse allikate (struktureeritud andmed, metaandmed, sõnastikud, klassifikaatorid ja dokumentatsioon) sidumine ühtseks teadmuskihiks ei pruugi anda oodatud kvaliteediga vastuseid või võib osutuda keerukamaks, kui projekti alguses eeldati.

**Maandamismeede:** Lahendust arendatakse ja hinnatakse järk-järgult, alustades vajadusel piiratud arvu statistikavaldkondadega ning laiendades teadmuskihiti samm-sammult järgmistele valdkondadele. Selline lähenemine võimaldab hinnata teadmuskivi kvaliteeti eri tüüpi andmestikel, tuvastada kitsaskohad varajases etapis ning täiustada lahendust enne selle laiendamist.

### Organisatsioonilised riskid

**Risk 4:** Projekti elluviimiseks vajalik koostöö Statistikaameti, RIA ja JustDigi vahel ei pruugi projekti jooksul kujuneda piisavalt tihedaks või võivad partnerite prioriteedid muutuda. See võib raskendada tehnilise ja organisatsioonilise toimetumise väljatöötamist.

**Maandamismeede:** RIA kaasatakse projekti juhtühma ning JustDigiga konsulteeritakse projekti vältel jooksvalt. Teemat käsitletakse regulaarselt ka Statistikaameti ja JustDigi andmehalduse koostöökohtumistel, et lahendada tekkivaid küsimusi ning kujundada ühine arusaam toimetumisel.

**Risk 5:** Projekti võtmeisikute töökoormuse muutumine või personali vahetus võib mõjutada projekti elluviimise kiirust ja teadmiste järjepidevust.

**Maandamismeede:** Projekti viiakse ellu meeskonnana ning teadmised dokumenteeritakse jooksvalt. Statistikaametil on ulatuslik kogemus keerukate arendus- ja innovatsiooniprojektide juhtimisel (sh varasemad Innovatsioonifondi projektid, Eurostati rahastatud arendusprojektid ning teaduskoostöö projektid), mis võimaldab vajadusel ülesandeid ümber korraldada ja tagada projekti järjepidev elluviimine.

#### **Andmete ja sisu kvaliteediga seotud riskid**

**Risk 6:** Olemasolevad andmekirjeldused, sõnastikud ja metaandmed on Statistikaameti eri statistikavaldkondades erineva kvaliteedi ja detailsusega. See võib raskendada teadmuskihhi loomist ning suurendada vajadust käsitsi korrastamiseks.

**Maandamismeede:** Piloottlahendust arendatakse etapiviisiliselt, alustades statistikavaldkondadest, kus andmekirjeldused ja metaandmed on paremini korrastatud. Seejärel laiendatakse katsetust järk-järgult järgmistele valdkondadele, et hinnata lahenduse toimivust erineva kvaliteediga lähtematerjali korral. Projekti käigus katsetatakse ka tehisintellekti kasutamist andmekirjelduste ja metaandmete rikastamiseks ning hinnatakse selle mõju käsitöö vähendamisele.

## **6. Projekti ajakava**

**Koostage realistlik ajakava, mis hõlmab kõiki projekti tegevusi ning annab sellega sisendi projekti eelarve koostamisele.**

- Ajakava koostamisel arvestage vajalike eel- ja järel- või vahetegevustega (nt partnerluslepingu sõlmimise ettevalmistus kuni 2 kuud, vajalike lubade saamine projekti jooksul vms).
- Milliste võimalike puhvritega oleks ajakavas mõistlik arvestada?
- Jagage tegevused loogilisteks etappideks, arvestage tegevuste omavahelisi seoseid ning ajalist järgnevust või paralleelsust.
- Hangete läbiviimise ajaraami kavandamiseks kasuta hankekalkulaatorit [Hankekalkulaator - EIS](#)

<b>Tegevused</b>	<b>Tegevuse algus (mitmes kuu)</b>	<b>Tegevuse lõpp (mitmes kuu)</b>	<b>Kestus kokku (mitu kuud)</b>
Eeltöö			
Partnerluslepingu ettevalmistamine ja sõlmimine			
Hanke ettevalmistamine ja läbiviimine	I kuu (November 2026)	VI kuu (aprill 2027)	6 kuud
Lepingu sõlmimine		VI kuu	
I etapp			
Tehnoloogilise lahenduse analüüs	VII	VIII	3 kuud
Toimemudeli lahenduse analüüs	VIII		3 kuud pikema perioodi jooksul
II etapp			

Infrastruktuur ja vajalikud tarkvarad toimimiseks	IX	IX	1 kuu
Allikabaasi (RAG) realiseerimine ja liidesed	IX	XI	3 kuud
Graafi komponendi realiseerimine	X	XII	3 kuud
Otsingu komponendi realiseerimine (kasutab GraphRAG-i)	XII	X	4 kuud
Treenimine (töö allikatega), katsetamine	X	XV	6 kuud
Andmekirjelduste koostamise abimehe komponendi realiseerimine	XIV	XVI	3 kuud
Puhver	XVII	XVII	1 kuu
Lõppraport ja soovitused	XVIII	XVIII	1 kuu
<b>KOKKU</b>			<b>18 kuud (2028.a. aprilli lõpuks valmis)</b>

## 7. Projekti eelarve

**Koostage realistlik eelarve detailsusega, mis hõlmab kõiki projekti tegevusi ning võimaldab seeläbi hinnata planeeritud kulude vajalikkust ja mõistlikkust.**

- Arvutage eelarves summad kogumaksumusena (st sisaldavad kõiki makse), sh projektijuhi kogukulu.
- Lisage eelarvele kirjeldusena selle kujunemise põhjendused, arvutuste ja hinnangute alused.
- Eelarve kogusumma palume esitada 1000 euro täpsusega.

**Kohandage eelarvetabelit oma projekti vajadustele vastavaks.**

Tegevused	Partner 1 kulud
I etapp: ANALÜÜS	
Tehnoloogilise lahenduse analüüs	60 000 + km
Toimemudeli analüüs	30 000 + km
II etapp: ARENDUS JA KATSETAMINE	
Vajalikud tarkvarad toimimiseks (sh katsetamise faasiks)	30 000 + km
Tehnoloogilise lahenduse (sh allikabaasi ja otsingukomponendi) realiseerimine ja liidesed	500 000 +km
Andmekirjelduste koostamise abimehe komponendi realiseerimine	50 000+km
Ülekantavuse analüüs (toimemudel)	40 000+km
MUUD KULUD	
Projektijuhtimine (18 kuud)	100 000
IT ekspert (6 kuud, majasiseste katsetuste ettevalmistamine ja läbiviimine)	35 000
Väline ekspertiis (toetab Statistikaametit otse, nt aitab valideerida tulemusi)	20 000
Tarkvara litsentsi kulud (Statistikaamet)	5000+km
Lähetused	10 000
Arutelu- ja valideerimise üritused, katsetuse kommunikatsioon (eraldi sisseostetav teenus)	20 000+km
<b>KOKKU (koos käibemaksuga)</b>	<b>1 076 400 eurot</b>

Eelarve koostamisel arvestati turukonsultatsioonil osalenud ettevõtete hinnanguid. Eelarve täpsustub hanke-eelsete turukonsultatsioonide käigus.

## 8. Võimalikud lahenduste pakkujad

**Tooge välja võimalikud hankepartnerid, kes soovitud lahendussuunas tooteid/ teenuseid/ pakuvad.**

- Otsige ja nimetage võimalikke probleemile lahenduste pakkujaid (nt erinevate valdkondade eksperdid, teadlased, ettevõtted, kes on probleemi lahendamiseks varasemalt tegelenud). Mõelge nii Eesti kui rahvusvaheliste pakkujate peale.

Eestis on võimalikeks lahenduste pakkujateks ülikoolid ja valdkonnas tegutsevad ettevõtted nagu Cybernetica, STACC, MindTitan, CGI, Texta ja mitmed teised ettevõtted. Kindlasti on hange avatud ka rahvusvahelistele tegijatele.

Statistikaametil on hea võrgustik ja koostöösuhted ka rahvusvaheliste valdkondlike ekspertidega, keda on võimalik kaasata kas ekspertidena Statistikaametile appi tellija poolele.

## 9. Projekti meeskond ja töökorraldus

**Tooge välja projekti edukaks elluviimiseks kaasatavad või vajalikud osapooled (asutused ja/või inimesed) ning täiendav ekspertiis, mida meeskonda juurde vajate.**

- Kirjeldage rollide ja töö jaotust projektimeeskonnas.
- Kirjeldage projekti juhtimise korraldust.
- Märkige ära, kui suure koormusega projektijuht (võimalusel ka teised võtmeisikud) projekti panustavad.
- Kirjeldage, missugust täiendavat ekspertiisi tuleb juurde kaasata (nt tehniline ekspertiis, andmekaitse), mis on meeskonnaliikmete poolt katmata.

**NB! Kui nimetate konkreetseid meeskonnaliikmeid, siis nendega (või nende juhtidega) peab olema projektis osalemine läbi räägitud!**

### Projekti meeskond

**Projektijuht** – värvatakse positiivse rahastusotsuse korral (0,8–0,9 koormusega). Projekti ettevalmistamise ja käivitamise eest vastutab kuni projektijuhi tööle asumiseni Terje Trasberg.

**Veiko Berendsen** – andmehalduse juhtivekspert. Vastutab projekti sisulise kontseptsiooni, teaduskihi, semantiliste mudelite ning tehnilise ja organisatsioonilise toimetamise väljatöötamise eest.

**Roby Palumäe** - semantikavarade ekspert. Vastutab oskussõnavara koondamise, ühtlustamise ja rakendamise ning andmevarade ja metaandmete kirjeldustega integreerimise eest, samuti asutuste vahelise koostöö eest.

**Marvin Üürike** – andmeteadur ja tehisintellekti ekspert. Vastutab AI-lahenduste arhitektuuri ja tehnilise katsetuse eest.

**IT spetsialist** - toetab katsetuse tehnilist läbiviimist Statistikaameti keskkondades. Täpne ametinimi ja funktsioon täpsustub projekti käigus.

**Kliendikanalite spetsialistid** – tunnevad praegust süsteemi, toetavad katsetuste läbiviimisel.

Projekti kaasatakse vastavalt vajadusele Statistikaameti andmehalduse meeskonna liikmeid, leviosakonna eksperte (stat.ee ja kasutajateenused) ning juriste hangete ettevalmistamiseks ja läbiviimiseks.

Projekti edenemist toetab regulaarselt kogunev **juhtrühm**, kuhu kuuluvad:

- Terje Trasberg, andmehõive ja töötamise osakonna juhataja;
- Anne Jürgenson, innovatsioonitoetuse juht;
- Annika Uibopuu, andmehalduse teenuse juht;
- **RIA esindaja** (andmehalduse valdkond), kelle roll on nõustada projekti ning toetada lahenduse koostöövõime Eesti andmehalduse ökosüsteemiga, eelkõige Eesti andmete teabevärava (ATV) vaates.

**Justiits- ja Digiministeeriumiga** tehakse projekti jooksul regulaarset koostööd, et arutada lahenduse toimetamist ja selle võimalikku laiemat rakendamist avalikus sektoris.

Statistikaameti juhtkond jälgib projekti edenemist regulaarsete ülevaadete kaudu.

Projekti käigus tehakse koostööd **Rahandusministeeriumiga**, kus kavandatakse valdkondliku teadmuskäigu loomist valimislubaduste analüüsi jaoks (taotlus esitatud Justiits- ja Digiministeeriumi AI meetmesse). Koostöö eesmärk on vahetada kogemusi, jagada õppetunde ning kasutada projektide käigus saadud teadmisi vastastikku, et vältida dubleerimist ja kiirendada lahenduste küpsemist.

Ilmselt on sarnaseid vajadusi ka teistes avaliku sektori valdkondades. Hea meelega jagame juba projekti käigus saadud kogemusi ja tulemusi teiste huvilistega, et vältida dubleerivaid arendusi ning luua eeldused lahenduste, komponentide ja teadmuse taaskasutamiseks ning vastastikuseks edasiarendamiseks.

#### **Rahvusvaheline koostöö**

Projekti käigus tehakse koostööd teiste riikide organisatsioonidega, kes tegelevad sarnaste probleemide lahendamisega. Vajaduse korral kaasatakse rahvusvaheline ekspert või konsultant, kes aitab hinnata tehnilisi lahendusi ning valideerida projekti tulemusi.

## **10. Projekti tulemuste elluviimine**

***Kirjeldage oma valmisolekut ja võimekust pärast katseprojekti edukat lõppu projekti tulemusi kestlikult ellu viia.***

- *Kas projekti tulemuste edasine arendus ja kasutuselevõtt seostub asutuse prioriteetsete tegevustega, on tööplaanis vms?*
- *Kas tulemuste hilisemaks elluviimiseks vajalik rahastus ja muud ressursid on olemas või tegeletakse selle leidmisega?*
- *Tooge välja olulisemad riskid projekti tulemuste hilisemal kasutuselevõtul. Kuidas plaanite neid riske maandada?*
- *Kirjeldage, kas ja mil määral on tulemused skaleeritavad ning kasutatavad avalikus sektoris laiemalt.*

Projekt on seotud otseselt Statistikaameti arenguplaanidega:

- Riikliku statistika seaduse kohaselt on Statistikaameti ülesanne lisaks riikliku statistika tegemisele **koordineerida riigi andmehaldust ja juhtida klassifikaatorite süsteemi**. Statistikaametil on üle viie aasta kogemust Eesti riigi andmehalduse koordineerimisel koostöös Justiits- ja Digiministeeriumiga ning teiste avaliku sektori asutustega. Andmehaldus on üks neljast Statistikaameti 2023.–2030. aasta arengukava arengusuunast, mille keskne mõte on: „**Andmed on meie vara. Meil on ülevaade riigi andmetest. Andmete kodu on korras.**” Käesolev projekt toetab seda suunda, aidates luua raamistikku, mille abil on avaliku sektori andmed, metaandmed, mõisted ja seosed paremini leitavad, mõistetavad, kvaliteetsed ja taaskasutatavad ka tehisintellekti lahendustes. Statistikaameti andmehalduse roll ei piirdu ainult ameti enda vajadustega, vaid toetab kogu avaliku sektori andmete korrastamist, kirjeldamist ja taaskasutust. Projekti tulemusi saab kasutada andmevarade kaardistamisel, andmetest tervikliku ülevaate loomisel, andmekvaliteedi seires ning metaandmete ajakohasuse tagamisel. Ühtne oskussõnavara ja metaandmete taaskasutus vähendavad dubleerivat tööd andmete kirjeldamisel ja kogumisel ning aitavad seeläbi vähendada halduskoormust ja ressursikulu. Lisaks tehnilisele tulemusele suurendab projekt asutuste valmisolekut arendada AI-põhiseid teenuseid ja teadmuspõhiseid lahendusi, mis tuginevad korrastatud, kirjeldatud ja masinmõistetavatele andmetele.

- Põhifunktsiooni ehk statistika tegemise kõrval on Statistikaameti üks ülesandeid **teha usaldusväärne statistika kõigile kättesaadavaks**. Avaldatud andmetest ei ole ühiskonnal kasu, kui nad ei ole kasutamiseks kiiremail ja kasutajasõbralikumail moel leitavad. Ameti kehtiva strateegia üks eesmärkidest on, et meie teenused on kiired ja paindlikud ning vastavad kliendi vajadustele. Statistika kättesaadavaks tegemise kõrval on oluline strateegiline eesmärk suurendada andmekirjaoskust, et riigis avaldatud andmeid kasutataks oskuslikult. See projekt aitab otseselt parandada andmete leitavust tänapäevasemal moel, mida ühiskond ootab.
- Projekti tulem toetab Statistikaameti **andmete taaskasutuse teenust**, mille eesmärgiks on teadlastele, avalikule sektori jt andmekasutajatele võimaldada turvalist ligipääsu üksikandmetele. Selleks on arendatud ERIKA keskkond, mille osaks on ka mahukas andmekataloog. Projekti tulemused võimaldavad andmevajajatel jõuda mugavalt ja kiiresti teadmiseni, kas ja millised vajalikud andmed on ERIKA keskkonna kaudu võimalik kasutamiseks saada Statistikaametist. ERIKA keskkond on potentsiaalselt kasutusele võetav ka teiste (riigi)asutuste poolt, kes soovivad konfidentsiaalseid andmeid kättesaadavaks teha (arendused on arutelu faasis).

### **Tulemuste elluviimise rahastus**

Projekti käigus loodav lahendus võetakse esmajärjekorras kasutusele Statistikaametis. Projekti lõppemise järel on võimalik selle edasist arendamist ja haldamist esialgu rahastada Statistikaameti andmehalduse arendamiseks eraldatud vahenditest. Need vahendid on aga ajutise iseloomuga ning pikaajaline jätkusuutlikkus eeldab andmehalduse võimekuse kujundamist avaliku sektori püsifunktsioonina koos sellele vastava rahastusega.

Projekti üheks eesmärgiks ongi hinnata lahenduse kasutegurit, ülalpidamiskulusid ja ressursivajadust, et luua alus teadlike otsuste tegemiseks selle edasise rahastamise kohta. Eeldame, et projekti tulemusena väheneb käsitöö andmehalduses, paraneb andmete leitavus ning väheneb ekspertide ajakulu korduvatele andmepäringutele, TI pakutud väärandmete kasutamine ja andmete leidmisega seotud nõustamisele. Need mõjud aitavad osaliselt katta lahenduse kasutuselevõtu ja haldamise kulusid ning suurendavad investeringu tasuvust.

### **Skaleeritavus**

Projekti käigus katsetatakse lahendust Statistikaameti andmestikel, kuid juba arhitektuuri kavandamisel lähtutakse põhimõttest, et selle üksikuid komponente oleks võimalik kasutada ka teistes avaliku sektori organisatsioonides.

Skaleeritavust käsitletakse kahel tasandil. Esmalt hinnatakse, kuidas saaks projekti tulemusena loodud teadmuskihiti ning selle tehnilisi komponente tulevikus rakendada riiklikes andmehalduse teenustes, eelkõige Eesti andmete teabevärava (ATV) edasiarendamisel. Selleks töötatakse välja tehnilise ja organisatsioonilise toimetudeli ettepanek ning kaasatakse projekti vältel RIA ja Justiits- ja Digiministerium.

Teiseks kavandatakse lahendus selliselt, et ka teised avaliku sektori organisatsioonid saaksid vajaduse korral üle võtta kas kogu lahenduse või üksikud komponendid (nt teadmuskihiti loomise põhimõtted, metaandmete rikastamise lahendused või AI-toega otsinguteenused). Samuti hinnatakse võimalusi kasutada Statistikaametis loodavat teadmuskihiti ka teiste avaliku sektori organisatsioonide AI-lahenduste alusena.

Projekti käigus hinnatakse ka lahenduse pikaajalist ressursivajadust, ülalpidamiskulusid ja sobivaid juurutusmudeleid, et tagada selle praktiline ja kulutõhus rakendamine nii Statistikaametis kui ka võimaliku laiendamise korral.

#### **Tulemuste elluviimise riskid**

**Risk:** Projekti tulemusena valmiv lahendus ei pruugi jõuda Statistikaametis igapäevasesse kasutusse, kui selle kasutamine nõuab liiga palju käsitööd, olemasolevate tööprotsesside muutmist või täiendavaid arendusi.

**Maandamismeede:** Lahendust arendatakse tihedas koostöös tulevaste kasutajatega ning seda katsetatakse projekti jooksul reaalse kasutusjuhtude peal. Eelistatakse olemasolevatesse tööprotsessidesse sobituvaid lahendusi ning hinnatakse nende praktilist kasutatavust enne projekti lõppu.

**Risk:** Projekti käigus valmiv lahendus (kui see on töötav) on edukalt katsetatud Statistikaametis, kuid selle laiem rakendamine avalikus sektoris eeldab täiendavat rahastust, mitme asutuse koostööd ning selgelt kokku lepitud rolle ja vastutust. Nende eelduste puudumisel võib lahenduse kasutuselevõtt piirduda Statistikaametiga.

**Maandamismeede:** Projekti käigus töötatakse välja tehnilise ja organisatsioonilise toimetumudeli ettepanek, mis kirjeldab lahenduse edasiseks rakendamiseks vajalikke rolle, vastutust ja koostööpõhimõtteid. RIA kaasatakse projekti juhtühma, et tagada lahenduse kooskõla riikliku andmehalduse ja Eesti andmete teabevärava arendamisega. Justiits- ja Digiministeeriumiga tehakse projekti jooksul regulaarset koostööd, et hinnata lahenduse laiendamise võimalusi ning kavandada selleks vajalikud järgmised sammud, sealhulgas võimalikud rahastusallikad.

**Risk:** Projekti käigus väljatöötatud lahenduse kasutuselevõtu ja ülalpidamise kulud võivad osutuda suuremaks, kui projekti alguses hinnati, mistõttu võib lahenduse laiem kasutuselevõtt osutada majanduslikult ebaotstarbekaks.

#### **Maandamismeede:**

Projekti käigus hinnatakse erinevaid tehnoloogilisi lahendusvariante (sh pilve- ja vajaduse korral kohapealseid lahendusi), nende ülalpidamiskulusid ning skaleeritavust. Eelistatakse modulaarset arhitektuuri, olemasolevate komponentide taaskasutamist ning võimalikult suurt automatiseeritust. Projekti tulemusena antakse hinnang lahenduse pikaajalisele ressursivajadusele ja sobivale juurutusmudelile.

## **11. Mõju ettevõtlusele**

✓ Projekt omab positiivset mõju innovatsioonile ettevõtlussektoris. Kõige otsesemalt väljendub mõju läbi ettevõtete, kes osalevad tegevuste elluviimiseks korraldatavatel hangetel ja/või konkurssidel. Innovatsiooni hankimine avaliku sektori poolt aitab kaasa innovatsioonitegevuste kasvule erasektoris.

## 12. Seos nutika spetsialiseerumise valdkondadega

- Eesti teadus- ja arendustegevuse, innovatsiooni ning ettevõtluse (TAIE) arengukaval 2021-2035 on fookusvaldkonnad, s.o Eesti arenguvajadustele ja -võimalustele vastavad riigi, ettevõtete ja teadusasutuste koostöös eelisarendatavad teadus- ja arendustegevuse, innovatsiooni ja ettevõtluse valdkonnad. Ettevõtluse ja majandusliku arengupotentsiaaliga TAIE fookusvaldkonnad on ühtlasi Eesti nutika spetsialiseerumise valdkonnad (täpsem info: <https://www.hm.ee/korgharidus-ja><https://www.hm.ee/korgharidus-ja-teadus/teadus-ja-arendustegevus/taie-fookusvaldkonnad>[teadus/teadus-ja-arendustegevus/taie-fookusvaldkonnad](https://www.hm.ee/korgharidus-ja-teadus/teadus-ja-arendustegevus/taie-fookusvaldkonnad)).
- Kirjeldage teie projekti võimaliku lahenduse seost vähemalt ühe valdkonnaga (rõhuasetusega teadmus- ja tehnoloogia siirdel).

Digilahendused igas eluvaldkonnas (vt teekaarti)	Projekt panustab teekaardi "Digilahendused igas elu valdkonnas" fookusvaldkondade arendamisse, täpsemalt eelkõige "digilahendused äriprotsesside innovatsiooni toetamiseks" ja "teadus- ja arendustegevus andmevaldkonna arendamiseks" (nt avaandmete kvaliteet, tehisaru teemad).
Tervisetehnoloogiad ja teenused (vt teekaarti)	
Kohalike ressursside (toit, puit, maapõueressursid, teisene toorme ja jäätmed) väärindamine (vt teekaarti)	
Nutikad ja kestlikud energialahendused (vt teekaarti)	

## 13. Seos strateegias Eesti 2035 toodud arenguvajadustega

- Selgitage, kuidas panustavad projekti tegevused ja valitud lahendussuund "Eesti 2035" strateegias kirjeldatud arenguvajadustesse.
- Tooge välja, kui projekti tegevused panustavad muudesse olulistesse valdkondlikesse arengukavadesse või -dokumentidesse.

Projekt toetab eelkõige Eesti 2035 sihte „Eesti majandus on tugev, uuendusmeelne ja vastutustundlik“ ning „Eesti riik on uuendusmeelne, usaldusväärne ja inimesekeskne“, luues avalikule sektorile uue võimekuse kasutada kvaliteetseid ja hästi kirjeldatud andmestikke tehisintellekti lahendustes. Projekti tulemusena muutuvad andmed paremini leitavaks, taaskasutatavaks ja masinmõistetavaks, mis loob aluse uute andmepõhiste teenuste ja otsustustoe lahenduste arendamiseks ning aitab tõsta avaliku sektori tõhusust ja innovatsioonivõimekust.

## 14. Avalike ülesannete täitmine projekti elluviimisel

- Selgitada ning tuua välja seosed ja viited, missuguse seaduse, määruse, haldusakti või lepingu alusel täidab ideekavandi esitaja asutus innovatsiooniprojekti ellu viies avalikke ülesandeid.

- *Kui ideekavandi esitaja on MTÜ, siis selgitada, kuidas ta pakub otsest avalikku teenust (loe [Teenuste korraldamise ja teabehalduse alused–Riigi Teataja](#), §2 lg2).*

Statistikaamet täidab projekti elluviimisel talle seadusega pandud avalikke ülesandeid, eelkõige riikliku statistika seaduse alusel.

Riikliku statistika seaduse kohaselt (§ 9) on Statistikaameti ülesandeks mh riikliku statistika tegemine, andmehalduse koordineerimine ja klassifikaatorite süsteemi juhtimine. § 33 järgi on Statistikaametil statistika levitamise kohustus.

Projekti käigus arendatakse lahendusi, mis parandavad Statistikaameti hallatavate andmete leitavust, mõistetavust ja taaskasutatavust ning toetavad andmehaldust ja statistika kasutamist.

Statistikaseaduse § 1 lg 2 ütleb omakorda, et "riikliku statistika eesmärk on kajastada ühiskonna olukorda ja muutusi ning varustada ühiskonda rahvastiku, sotsiaalvaldkonna, majanduse ja keskkonna arengu seisukohalt olulise teabega, sealhulgas arengukavade ja prognooside koostamiseks, erinevate poliitikate kujundamiseks, teadus- ja rakendusuuringute tegemiseks ning teadmuspõhiste otsuste langetamiseks."

## 15. Rahastus mitmest allikast

- *Kas probleemi lahendamiseks või planeeritud lahenduse katsetamiseks on taotletud või taotletakse toetust teistest rahastamisallikatest?*
- *Kui jah, siis tuua välja rahastusallikas, summa ja tegevused ning kas toetus on taotlemisel või projekt on saanud rahastusotsuse.*

Seosed:

- Probleemi lahendamisega seotud andmehalduse tegevusteks on Statistikaameti kasutusel **Euroopa Ühtekuuluvus- ja siseturvalisuspoliitika Fondide programmi rahastus**, mis on eraldatud ettevõtlus- ja infotehnoloogiainistri 14.03.2023 käskkirja nr 44 alusel. Sellele toel koondatakse, ühtlustatakse ja struktureeritakse andmehalduse valdkonna oskussõnavara ning seotakse seda andmevarade ja metaandmete kirjeldustega. Käesolev projekt ei dubleeri nimetatud tegevusi ega tähenda topeltrahastamist. Varasem rahastus katab andmehalduse aluskomponentide – oskussõnavara, andmekirjelduste, metaandmete ja nende seoste – loomist ja korrastamist. Käesolev projekt kasutab neid eeldusena ning keskendub tehnilise ja organisatsioonilise raamistiku loomisele, mille abil saab olemasolevad komponendid siduda tehisintellekti jaoks kasutatavaks teadmuskihiks ning katsetada selle toimimist avaliku sektori struktureeritud andmestikel.
- Riigikantselei on saanud Innofondist rahastuse projektile "**Andmepõhine poliitikakujundaja.**" Käesolev projekt täiendab Riigikantselei projekti. Kui Riigikantselei projekt keskendub poliitikakujundajale mõeldud rakendusele, mis kasutab erinevate registrite konfidentsiaalseid üksikandmeid poliitikaküsimustele vastamiseks, siis käesolev projekt loob selleks vajalikud eeldused, arendades andmestike semantilist kirjeldamist, metaandmeid ja teadmuskihiti. Käesoleva projekti fookus ei ole registrite mikroandmete töötlemisel ega nende analüüsimisel, vaid andmestike tähenduse, konteksti ja leitavuse parandamisel, et tulevased TI-lahendused saaksid neid usaldusväärset kasutada. Teatud kattuvus võib esineda otsinguprototüübi osas. Selle eesmärk käesolevas projektis ei ole iseseisva otsingulahenduse arendamine, vaid teadmuskihiti ja semantiliste lahenduste toimivuse demonstreerimine ning valideerimine. Seetõttu kasutatakse võimaluse korral olemasolevaid lahendusi ja komponente ning tehakse koostööd Riigikantseleiga, et vältida

dubleerimist. Arendustööd piirduvad ulatuses, mis on vajalik prototüübi sidumiseks Statistikaameti andmestike ja teadmuskihiga.

### Kinnitused

✓ Oleme teadlikud, et Riigikantselei võib saata ideekavandi eksperthinnangu saamiseks valdkonna ekspertidele.

Kinnitan, et esitatud innovatsiooniprojekt on teiste partnerite juhtkondadega kirjalikult kooskõlastatud.

### Allkirjastamine

- Ideekavand tuleb allkirjastada projekti esitava(te) asutus(t)e allkirjaõigusliku juhtkonnaliikme poolt (nt kantsler, asekanstler, KOVi juht, KOVi volikogu esimees, ministeeriumi allasutuse juht/asejuht vms) ja saata [riigikantselei@riigikantselei.ee](mailto:riigikantselei@riigikantselei.ee).

<sup>i</sup> **Katsetus** vastab küsimusele: kas see töötab? Katsetamise puhul ei vaadata alati, kas lahendus praktiliselt toimib.

**Piloteerimine** vastab küsimusele: kas see töötab päriselus ja on mõistlik kasutusele võtta? Hinnata praktilist toimivust.

**Ekspereiment:** Igasuguse eksperimendi eesmärk on kontrollida hüpoteese põhjuslike seoste kohta. Laialdaselt kasutatava määratluse kohaselt on eksperimeint selline katse, mis on kavandatud põhjuslike seletusteni jõudmiseks (Cook ja Campbell 1979, tsiteeritud Elmese, Kantowitzi ja Roedigeri 2013: 146 järgi). Kui teeme x siis juhtub y. **Prototüüp** - masina, seadme või mingi rakenduse esialgne teostus, algne mudel, mida edasi arendatakse.