

TALLINNA TEHNIKAÜLIKOOL  
Infotehnoloogia teaduskond

Rutt Lindström, 175720IDDR

# **SNOMED CT Eesti laiendi failide ülevaatuse tööriist**

Diplomitöö

Juhendaja: Priit Rospel  
MSc

Tallinn 2021

## **Autorideklaratsioon**

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Rutt Lindström

17.05.2021

## Annotatsioon

SNOMED CT on üks suurimaid rahvusvahelisi kodeeritud terminoloogiaid meditsiini valdkonna infosüsteemides kasutamiseks. Tervise ja Heaolu Infosüsteemide Keskus (edaspidi TEHIK) on SNOMED CT Eesti laiendi haldaja.

SNOMED CT-d avaldatakse mitmel eri moel, üks oluline avaldamisviis on SNOMED Internationali poolt kindlaks määratud formaadis laiendifailide komplekt. Enne Eesti laiendi avaldamist saadetakse see masinloetav failikomplekt TEHIKule ülevaatamiseks, kuid selle kontrollimiseks või inimloetavale kujule konverteerimiseks puuduvad täna töövahendid.

Käesoleva töö eesmärk oli SNOMED CT Eesti laiendi masinloetava failipaketi avaldamiseelse kontrolli tarbeks prototüüptööriista loomine. Eesmärgini jõudmiseks oli tarvis lahendada hulk alamülesandeid, mis jaotusid eri peatükkide vahel:

- 1) ülevaate tegemine terminoloogia haldamisest ja SNOMED CT olemusest;
- 2) SNOMED CT tehniliste võimaluste ja piirangute kirjeldamine;
- 3) loodava tööriista taustsüsteemi ja seotud tööprotsesside analüüs;
- 4) tööriista kavandi loomine;
- 5) tööriista tehniline realisatsioon ning valminud prototüübi lühike järelanalüüs.

Valminud tööriist on täisfunktsionaalne prototüüp, kuid see ei ole mõeldud asutuse ametlikuks tarkvaraks ning täidab eelkõige näidise rolli, kusjuures prototüübi lahutamatuks osaks on selle loomise käigus valminud analüüs ja dokumentatsioon.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 34 leheküljel, 7 peatükki, 15 joonist, 1 tabelit.

# **Abstract**

## **Review Tool for SNOMED CT Estonian Extension Release Files**

SNOMED CT is one of the leading international terminologies for use in clinical information systems. Health and Welfare Information Systems Centre (TEHIK) is the national release centre for SNOMED CT and responsible for authoring the Estonian extension. Shortly before SNOMED International publishes a new version of Estonian SNOMED CT extension the release files are sent to TEHIK for a pre-release check. Today, TEHIK does not have a tool for reviewing these machine readable pre-release files and therefore the review is done manually and using insufficient office software.

The author of the paper aims to build a functional prototype of the review tool for SNOMED CT pre-release files. The tool supports importing RF2 release files, calculating the statistics of changes and overall content of the files and show the main content in a human-readable form. The system is designed to use SNOMED International's terminology services to enrich the referenced components with human-readable descriptions from relevant modules of SNOMED CT.

The paper consists of several topics as follows:

- 1) Brief introduction to the basics of terminology and SNOMED CT;
- 2) Tooling, technical possibilities and restrictions of SNOMED CT;
- 3) Analysis of the current workflow and problems related to the topic;
- 4) Requirements for the planned tooling;
- 5) Creation of the functional prototype and a brief analysis of the result.

The prototype tool is a functional software that serves as an example and documentation. It is not designed to become an official organisational software but it could be used as a temporary solution.

The thesis is in Estonian and contains 34 pages of text, 7 chapters, 15 figures, 1 table.

## Lühendite ja mõistete sõnastik

<b>API</b>	<i>Application Programming Interface.</i>
<b>CDA</b>	<i>Clinical Document Architecture</i> , HL7 poolt loodud tervisedokumentide andmevahetusstandard.
<b>FSN</b>	<i>Fully Specified Name</i> ; mõiste definitsioonina käsitletav tekstkirje SNOMED CT-s.
<b>HL7</b>	Health Level Seven International.
<b>HTTP</b>	<i>Hypertext Transfer Protocol.</i>
<b>JSON</b>	<i>JavaScript Object Notation</i> ; andmevahetusvorming.
<b>MLDS</b>	<i>Member Licensing and Distribution Service</i> ; SNOMED CT veebikeskkond litsentside haldamiseks ning laiendifailide avaldamiseks.
<b>MRCM</b>	<i>Machine Readable Concept Model</i> ; SNOMED CT masinloetav mõistemudel.
<b>OWL</b>	<i>Web Ontology Language.</i>
<b>REST</b>	<i>Representational State Transfer.</i>
<b>RF2</b>	<i>Release Format 2</i> . SNOMED CT laiendifailide avaldamisformaad.
<b>RHK</b>	Rahvusvaheline Haiguste Klassifikatsioon (ingl <i>International Statistical Classification of Diseases and Related Health Problems</i> ).
<b>SCTID</b>	SNOMED CT identifikaator
<b>SNOMED CT</b>	<i>Systematized Nomenclature of Medicine: Clinical Terms.</i>
<b>SQL</b>	<i>Structured Query Language.</i>
<b>TEHIK</b>	Tervise ja Heaolu Infosüsteemide Keskus.
<b>TIS</b>	Tervise infosüsteem; riiklik andmekogu.

# Sisukord

1 Sissejuhatus .....	9
2 SNOMED CT: teoreetiline alus ja praktiline kasutamine .....	10
2.1 Terminoloogia haldamise põhimõtted .....	11
2.2 SNOMED CT taust.....	12
2.3 SNOMED CT ülesehitus .....	13
2.3.1 Mõiste ( <i>concept</i> ) .....	14
2.3.2 Kirjeldus ( <i>description</i> ).....	15
2.3.3 Seos ( <i>relationship</i> ).....	16
2.3.4 Viiteloend ( <i>reference set</i> ) .....	17
2.4 SNOMED Internationali tööriistad ja teenused.....	17
2.4.1 SNOMEDi täisteenus ehk <i>Managed Service</i> .....	18
2.4.2 Veebipõhised lehitsejad ja terminoloogiaserver.....	19
2.4.3 SNOMED RF2 laiendifailide pakett .....	20
3 Lähteülesande kirjeldus ja analüüs .....	22
3.1 Hetkeolukorra kirjeldus .....	22
3.2 Lahenduse analüüs.....	24
3.2.1 Terminoloogiateenuse abil RF2 failide loomine .....	24
3.2.2 Avaldamiseelne RF2 failide ülevaatus .....	25
4 Tööriista kavand .....	27
4.1 Funktsionaalsed nõuded .....	27
4.2 Mittefunktsionaalsed nõuded.....	29
4.3 Tööriistade valik .....	30
5 Tehniline realisatsioon.....	32
5.1 Serrakendus ja andmebaas.....	32
5.2 Klientrakendus ja SNOMED CT terminoloogiateenused .....	34
5.3 Rakenduse töövoog ja näited .....	34
5.3.1 Muudatuste statistika töövoog.....	34
5.3.2 Viiteloendite ülevaatus .....	36
5.3.3 Teised vaated ja funktsioonid .....	38
6 Tulemuse analüüs .....	40
7 Kokkuvõte .....	42

Kasutatud kirjandus .....	43
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks .....	45
Lisa 2 – SNOMED CT RF2 failide representatsioon andmebaasi tabelitena.....	46
Lisa 3 – SNOMED CT Eesti laiendis viidatud sisu ning seotud ärireeglid.....	47
Lisa 4 – Vastavustabel JSON ja RF2 representatsioonide vahel.....	50

## Jooniste loetelu

Joonis 1. SNOMED CT loogiline mudel. (Bhattacharyya 2016: 26).....	13
Joonis 2. Mõiste graafiline kirjeldus SNOMED CT lehitsejas. (SNOMED 2021i).....	14
Joonis 3. Mõistega seotud kirjeldused SNOMED CT lehitsejas. (SNOMED 2021i) ....	15
Joonis 4. Veebipõhise SNOMED CT lehitseja lihtotsing. (SNOMED 2021i).....	19
Joonis 5. Väljavõte Snowstorm terminoloogiateenuse vastusest. (SNOMED 2021h)...	20
Joonis 6. SNOMED CT RF2 laiendifail: eesti keele viiteloend.....	21
Joonis 7. Klientrakenduse veebitõmmis: menüü ja statistika kuvamise leht.....	35
Joonis 8. Klientrakenduse koodinäide .....	35
Joonis 9. Serverrakenduse koodinäide: statistika kontroller .....	36
Joonis 10. Serverrakenduse koodinäide: äriloogika kihi statistika arvutamise meetod .	36
Joonis 11. Klientrakenduse veebitõmmis: viiteloendite nimekiri. ....	37
Joonis 12. Klientrakenduse veebitõmmis: viiteloendi sisu.....	38
Joonis 13. Klientrakenduse veebitõmmis: kirjelduste ülevaatus. ....	38
Joonis 14. SNOMED CT ülesehitus SQL ja RF2 jaoks. (SNOMED 2021b: 18) .....	46
Joonis 15. SNOMED CT Eesti laiendi sisu kujunemine ja viitamine. ....	47



# 1 Sissejuhatus

SNOMED CT (ingl *Systematized Nomenclature of Medicine: Clinical Terms*) on üks juhtivaid infosüsteemides ja andmevahetuses kasutatavaid meditsiiniterminoloogiad. Selle rahvusvahelist väljaannet haldab SNOMED International ning Eesti laiendit Tervise ja Heaolu Infosüsteemide Keskus (edaspidi TEHIK).

Käesolev töö on saanud tõe TEHIKu andmekorraldustalitusel SNOMEDI Eesti laiendiga töötamisest ning töö eesmärk on luua tööriist Eesti laiendi sisu avaldamiseeliseks kontrollimiseks. Täna hallatakse SNOMEDI Eesti laiendit valdavalt SNOMED Internationali poolt pakutavate tööriistadega, kuid laiendi masinloetavate failide ülevaatamiseks SNOMED International spetsiaalset tööriista ei paku ning seda ei ole ka TEHIKul.

Töö esimeses osas tutvustatakse terminoloogiahoolduse põhimõtteid ja SNOMED CT-d põhjalikumalt. Seejärel antakse ülevaade SNOMED Internationali pakutavatest valmistarkvaradest ja teenustest, mis on tehniliseks raamistikuks tööriista kavandamisel.

Töö teises osas analüüsitakse SNOMED CT Eesti laiendiga seotud tööprotsesse ja sellega seotud probleeme, mille lahendamiseks tööriista loomist planeeritakse. Seejärel luuakse tööriista kavand ehk funktsionaalsete ja mittefunktsionaalsete nõuete kogum.

Töö viimases osas realiseeritakse tööriista täisfunktsionaalne prototüüp. Lahenduskäik kirjeldatakse nii SNOMEDI-spetsiifilise keerukuse võtmes kui ka teiste tehniliste valikute osas, nagu arendusvahendid, arhitektuur, keeled ja raamistikud. Töö lõpeb valminud prototüübi lühikese analüüsiga, mille eesmärk on hinnata valminud töö tulemust ning jõuda järeldusteni, kuidas prototüübi funktsionaalsust kõige paremini reaalses infosüsteemis realiseerida.

Töö käigus loodud rakendus on näidisrealisatsioon, mis on oma funktsionaalsuselt täielik, kuid ei ole ette nähtud kasutamiseks asutuse ametliku tarkvarana.

## 2 SNOMED CT: teoreetiline alus ja praktiline kasutamine

Esimesed digitaalsed haiguslood võeti kasutusele juba 1960ndatel ja 1970ndatel. Need olid vormilt tekstilised ja tänapäeva mõttes algelised, kuid juba siis näitasid uuringud, et täisdigitaalse haigusloo kasutuselevõtt suutis oluliselt vähendada haigla ravikulusid. (Benson, Grieve 2021: 12-13)

Lihtsaim ja algelisim terminoloogia haldamise ja kodeerimise viis on koodiloend. Koodiloend koosneb kood-tekst paaridest ja selle eesmärk on lihtsustada infosüsteemide andmevahetust, tõlkimist ning eri kodeerimissüsteemide vahelist vastendamist. (Elkin 2012: 54) Koodiloendiks võib pidada ka väljavõtet suuremast terminoloogiast või klassifikaatorist. Sel juhul on loendi eesmärgiks kitsendada konkreetset kontekstis kasutatavate kirjete hulka ning kirjete ontoloogia on vajadusel (nt andmete analüüsil või kontrollil) tuletatavad alusterminoloogiast lähtuvalt. (Elkin 2012: 99–100)

Klassifikaator on hierarhiline koodiloend, mille eesmärk on lõplikult kodeerida ühe kindla domeeni mõistestik. Klassifikaator on reeglina monohierarhilise ülesehitusega ning sisaldab nn jääkkategooriaid, kuhu on võimalik klassifitseerida kõik see, mis mujale ei sobi. Klassifikaatorid võeti esmalt kasutusele statistilise eesmärgiga ja seda juba enne digitaliseerimise võidukäiku. Üks enim kasutatud klassifikaatoreid tervishoius on RHK ehk rahvusvaheline haiguste klassifikaator, mis asutati 1893. aastal surma põhjuste klassifikaatorina. (WHO 2021) Farmaatsia valdkonna levinuim klassifikaator on ATC ehk ravimite anotoomilis-terapeutiline-keemiline klassifikatsioonisüsteem, mis pärineb 1970ndate aastate Norrast. (Ravimiamet 2019: 195)

Klassifikaatorist keerukama ülesehitusega on ontoloogilised terminoloogiad, kus kodeeritud mõistet kirjeldab lisaks tekstile ja asukohale taksonoomias ka talle omistatud atribuutide väärtused. Terminoloogia on olemuslikult avatud süsteem, mis üritab küll kodeerida ja kirjeldada võimalikult palju oma valdkonnast, kuid jääb alati laiendatavaks. (Elkin 2012: 54) Kõige laiahaardelisem terminoloogia meditsiinivaldkonnas on SNOMED CT (Benson, Grieve 2021: 293).

## 2.1 Terminoloogia haldamise põhimõtted

Terminoloogia haldamise kõige fundamentaalsem nõue on mõistekeskus (*concept orientation*), mis sobib hästi kokku tarkvaraarenduse objektorienteeritusega. Mõistekeskuse kandvaks ideeks on, et terminoloogia keskseks elemendiks on reaalmaailma nähtus (mõiste), mitte selle tekstiline kirjeldus ega kood. (Elkin 2012: 55)

Mõistekeskusest lähtuvad ka teised põhiprintsiibid:

- ühele mõistele vastab terminoloogias alati üks kood;
- ühele koodile ei tohi vastata rohkem kui üks mõiste;
- mõistete kirjeldus peab olema kontekstivabalt üheselt arusaadav;
- terminoloogia peab olema sisemiselt seotud ja järjepidev, nt paralleelsetes domeenides peab olema sarnane seosteloogika. (Elkin 2012: 55)

Terminoloogiaalduse tehnilised põhimõtted:

- kontekstivabad koodid (kood ei tohi endas kanda lisatähendust);
- kasutada tuleks alfanumeerilisi koode, mille pikkus ei ole rangelt defineeritud;
- hoiduda koodide taaskasutamisest;
- koodide ja mõistete versioneerimine, vastavus ajalooliste ja uute kirjade vahel;
- metaandmed (kirje loomise/muutmise aeg, allikas, uue ja vana kirje seos jms);
- sõltumatus keelest ning tõlkimisvõimalus;
- ajakohasus ehk pidev parandamis- ja täiendamisevõimalus. (Elkin 2012: 59–60)

Ülalkirjeldatud põhimõtted on terminoloogiasüsteemide väljatöötajate poolt laialt omaks võetud ning levinud ka meditsiinivaldkonna infotehnoloogias. Erinevate standardimisprojektide teoreetiline pagas sõnastati tungivalt soovituslike nõuetena 1998. aastal J.J. Cimino artiklis „Desiderata for Controlled Medical Vocabularies in the Twenty-First Century“. Lisaks eelpool mainitud põhimõtetele rõhutati selles artiklis ka meditsiiniterminoloogia polühierarhilise ehituse vajadust, granulaarsuse reguleerimise võimalust, tautoloogiliste ühendite vältimist (“vaagnaluu murd vaagna piirkonnas”) ning

hoidumist klassifikatsiooni jääkkategooriatest, millel puudub kliiniline sisu (nt “Muu”, “Täpsustamata” jms). (Cimino 1998: 394–403)

Kõiki siinses peatükis toodud põhimõtteid peetakse fundamentaalselt oluliseks ka SNOMED CT haldamisel ja arendamisel (Benson, Grieve 2021: 293–295).

## **2.2 SNOMED CT taust**

SNOMED CT terminoloogia algus ulatub 1965. aastasse, kui Ameerika Patoloogide Kolledž avaldas koodistiku nimega Systematized Nomenclature of Pathology (SNOP). See kirjeldas patoloogilisi leide neljal teljel: paiknemine, morfoloogia, etioloogia ja funktsiooni häirumine. See oli ühtlasi esimene mitmeteljeline tervishoius kasutatav kodeerimissüsteem. Järgmise 10 aasta jooksul arendati seda edasi kogu haiguste spektri jaoks ning sündis SNOMED. Sellele järgnenud edasiarendused kulmineerusid tänaseks SNOMED CT-ks aastal 2002. (Benson, Grieve 2021: 290, 295)

SNOMED CT on äärmiselt mahukas. Rahvusvaheline versioon koosneb umbes 350 000 mõistest, 1,5 miljonist tekstilisest kirjeldusest ning 3 miljonist seosest (SNOMED 2021i). Sellele lisandub riikide kohalikes laiendites loodud sisu. SNOMED CT on kasutatav vaid infotehnoloogiliste vahenditega – seda nii mahukuse kui ka keerulise ülesehituse tõttu, mis ühelt poolt teeb küll kasutamise keerulisemaks, kuid teiselt poolt lisab funktsionaalset võimekust ning laienemisvõimalust. (Benson, Grieve 2021: 295)

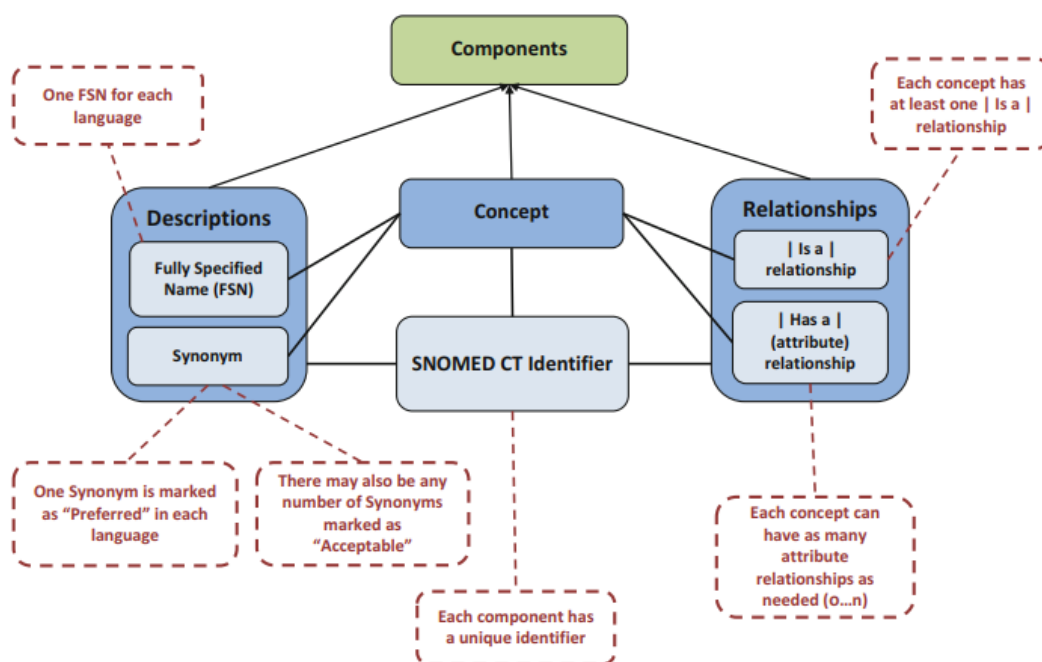
SNOMED CT rahvusvahelise versiooni ja üldise haldamise eest vastutab SNOMED International. Riiklikke laiendeid haldavad liikmesriigid, kuid mitte kõigil liikmetel ei ole oma laiendit. Väljaspool liikmesriike on SNOMEDit võimalik kasutada SNOMED Internationali otselitsentsi alusel. (SNOMED 2021a)

Kuigi Eesti on SNOMEDi liige ning SNOMEDi kasutamine Eestis on kõigile tasuta, pole see väga tuntud ega laialt levinud. Seda illustreerib ka madal kajastatavus tervisetehnoloogia alastes teadustöodes. Reeglina on SNOMED CT toodud välja kui võimalus olemasolevat süsteemi edasi arendada (nt Jürimaa 2016: 61), mõnikord ka mööndustega mainides, et kuigi SNOMED CT on kahtlemata võimekas, võib see siiski tuua kaasa liigse granulaarsuse andmete sisestamisel (Ross 2011: 32–33, 52). Pikemalt on SNOMED CT käsitletud Kerli Tarkini magistritöös “Possibilities to Improve Semantic Interoperability in Medical Data Exchange in Estonia with SNOMED CT: Analysis of

Discharge Summary“ (Tarkin 2013). Üks põhjalikumaid SNOMED CT tutvustusi eesti keeles on Kalev Karu artikkel ajakirjas Eesti Arst. Ka käesolevas töös on kasutatud terminite tõlkimisel suures osas Kalev Karu artiklis toodud eestindusi koos nende ingliskeelsete originaalterminitega. (Karu 2011: 466–473)

## 2.3 SNOMED CT ülesehitus

SNOMED CT on keerukas süsteem meditsiini ja tervise valdkonna terminoloogiast ja ontoloogiast. SNOMED CT loogilise mudeli keskseteks komponentideks on mõiste (*concept*), kirjeldus (*description*) ja seos (*relationship*). Loogilise mudeli graafiline kirjeldus on alljärgneval joonisel 1 (Bhattacharyya 2016: 26).



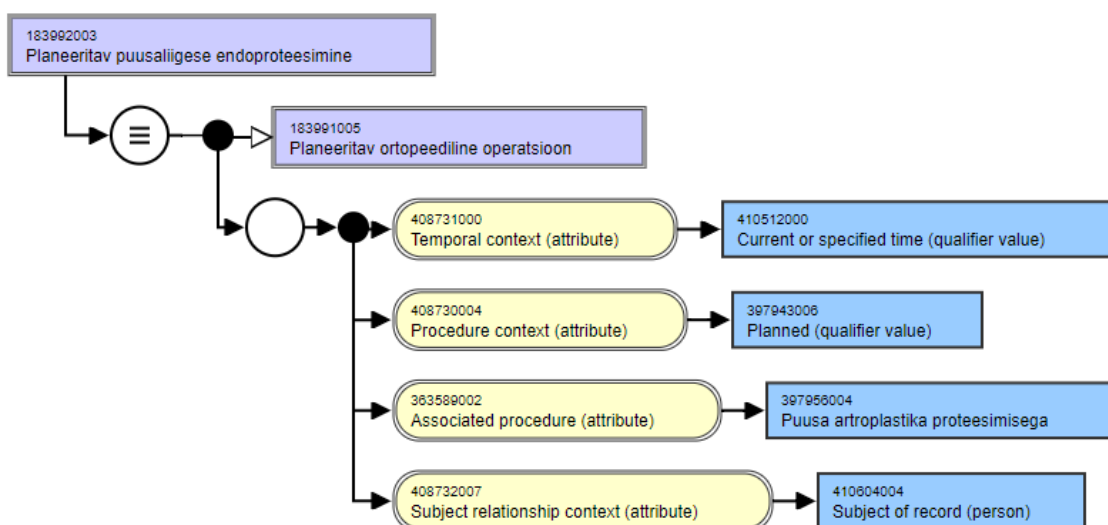
Joonis 1. SNOMED CT loogiline mudel. (Bhattacharyya 2016: 26)

Loogilise mudeli komponentide omavahelisi seoseid reguleerib mõistemudel (ingl *concept model*), mis kirjeldab mõiste kirjeldamiseks ja laiendamiseks võimalikud atribuudid ning nende semantilised väärtuspiirkonnad (Bhattacharyya 2016: 26, 31). Masinloetavat mõistemudelit saab kasutada informatiivsel või normatiivsel eesmärgil MRCM lehitseja või Snowstorm veebiteenuse kaudu (SNOMED 2021h).

### 2.3.1 Mõiste (*concept*)

SNOMED CT definitsiooni järgi on mõiste semantiline ühik ehk kliiniline idee, millele on antud unikaalne kood. Mõiste ei ole samastatav tema tekstilise kirjeldusega, sest sama mõistet on võimalik väljendada mitmel eri moel. Ühele mõistele vastab üks kood, ühele koodile vastab alati üks mõiste. Mõiste võib olla üldisem või täpsem kui teine mõiste (st ülem- või alammõiste), kuid mitte kunagi täpselt sama tähendusega (Bhattacharyya 2016: 27)

Iga mõiste on asetatud hierarhilisse taksonoomiasse. Iga mõiste juurde kuuluvad ka mõiste tekstilised terminid eri keeltes (kirjeldused, ingl *descriptions*) ning komplekt atribuute, mis kirjeldavad selle mõiste seoseid teiste mõistetega (seosed, ingl *relationships*). Seda, et iga SNOMED CT mõiste on seotud mitmete teiste SNOMED CT mõistetega, illustreerib joonis 2, kus kujutatakse graafiliselt mõiste „Planeeritav puusaliigese endoproteesimine“ seoseid teiste mõistetega. (SNOMED 2021i)



Joonis 2. Mõiste graafiline kirjeldus SNOMED CT lehitsejas. (SNOMED 2021i)

SNOMED CT rahvusvahelises versioonis on hetkel üle 350 000 mõiste. Kui kasutajal tekib vajadus uue mõiste järele, on võimalik seda SNOMED Internationalilt taotleda või lisada see enda kohalikku laiendisse. Ka ise lisatud mõiste puhul on oluline mõiste

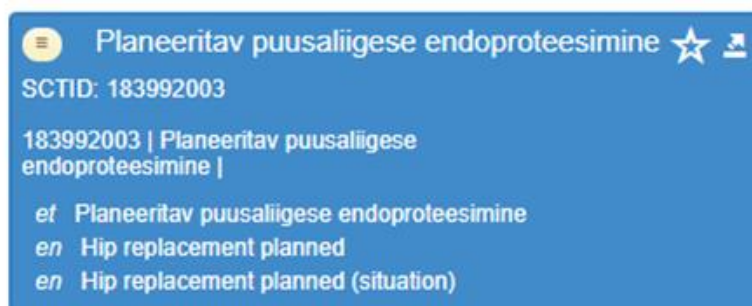
korrektne klassifitseerimine ja atribuutidega varustamine, et mõiste oleks võimalikult hästi mõistetav ka väljaspool kohalikku konteksti. (SNOMED 2021-1)

### 2.3.2 Kirjeldus (*description*)

Igal SNOMED CT mõistel on vähemalt kaks tekstilist kirjeldust:

- FSN (ingl k *Fully Specified Name*) ehk täisnimetus. FSN-i osaks on hierarhia harusse kuulumist tähistav semantiline märgend, mis kirjutatakse sulgudes. Näiteks: *Hip replacement planned (situation)*.
- Aktsepteeritud sünonüüm. Kui sünonüüme on antud keeles/dialektis vaid üks, võib seda pidada eelisterminiks. Kui mõiste juurde kuulub rohkem sünonüüme, tuleb eelistermin ja teised aktsepteeritud sünonüümid üksteisest eristada. Sünonüümide rohkus parandab tekstiotsingutel põhinevate funktsioonide kasutamist infosüsteemides. (Bhattacharyya 2016: 28-29)

Mõistete kohalikku keelde tõlkimine toimub riiklikus SNOMED CT laiendis, mille sisu avaldab SNOMED International ka enda lehitsejas. Järgnev joonis pärineb SNOMED Internationali lehitseja Eesti laiendi versioonist ning näitab tekstilisi kirjeldusi SNOMED CT mõiste 183992003 |Hip replacement planned (situation)| jaoks. Kohalik laiend hõlmab endas alati ka rahvusvahelise laiendi sisu, mistõttu sisaldab ka ingliskeelseid termineid. (SNOMED 2021i)



Joonis 3. Mõistega seotud kirjeldused SNOMED CT lehitsejas. (SNOMED 2021i)

Iga tekstilise kirjelduse puhul määratletakse ka info tema suurtäheortograafia kohta, et tagada tekstilise representatsiooni korrektsus eri kontekstides. (SNOMED 2021-1)

### 2.3.3 Seos (*relationship*)

Seos SNOMED CT loogilises mudelis on kahe mõiste vaheline ühendus, mille eesmärgiks on defineerida või väärtustada mõiste tähendust masintõlgendataval viisil. Seosed on jaotatavad kahte suurde klassi: *is a* seos ehk alammõiste seos ülemmõistega ning *has a* seos ehk mõiste omadust kirjeldav seos. Seoste illustreerimiseks vaatame järgnevalt kahte näidet:

**Näide 1:** 75570004 |Viral pneumonia (disorder)| : 116680003 | Is a (attribute) = | 233604007 |Pneumonia (disorder)|

Atribuutmõiste *is a* tähistab siin viirusliku kopsupõletiku kuulumist hulka “kopsupõletik”.

**Näide 2:** 75570004 |Viral pneumonia (disorder)| : [has a] 246075003 | Causative agent (attribute) | = 49872002 |Virus (organism)|

Erinevalt *is a* seosest ei ole *has a* eraldi kodeeritud mõiste, vaid mõtteline ülemkategoria kõigile omadusi kirjeldavate atribuutmõistete abil loodavate seoste jaoks. Antud näites ütleb seos, et viirusliku kopsupõletiku põhjustaja on viirus. (Bhattacharyya 2016: 30-31)

Atribuutide abil on võimalik ka ise mõisteid defineerida ning vastavalt vajadusele tähendust kitsendada. Näiteks soovides väljendada viiruslikku kopsupõletikku, mille põhjustajaks on konkreetselt *Streptococcus Pneumoniae*, tuleb otsustada, kas rahuldav on kasutada veidi laiemat tähendusega valmis mõistet või on vajalik koostada postkoordinatsiooni abil täpsustatud väljendmõiste (SNOMED 2021i):

**Näide 3:** Prekoordineeritud mõiste 34020007 | Pneumonia caused by Streptococcus (disorder) | omab SNOMED CT seoste hulgas definitsiooni, et põhjustajaks on perekond *Streptococcus*, aga mitte konkreetselt *Streptococcus Pneumoniae*.

**Näide 4:** Postkoordineeritud mõiste võimaldab mõistet rafineerida ning luua väljendmõiste: 75570004 |Viral pneumonia (disorder)| : 246075003 | Causative agent (attribute) | = 58800005 | Genus Streptococcus (organism) |.



Postkoordineeritud mõisteid saab kasutada samaväärselt prekoordineeritud mõistetega ning eelnevas näites toodud pika kirjelduse saab kodeerida lühidalt koode ja märgendeid kasutades 75570004:246075003=58800005. SNOMED CT tööriistad võimaldavad kiiresti kindlaks teha, et näites 4 toodud mõiste sisaldub ka kõigis eelnevates näidetes toodud mõistete hulgas – see on funktsionaalsus, mis teeb SNOMED CT hinnatud tööriistaks analüütikaprojektides. (Bhattacharyya 2016: 33)

SNOMED CT mõistemudel annab ette reegli selle kohta, millega ühest või teisest hierarhiast pärit mõistet võib täiendada ning mis on selle täiendamiseks kasutatava atribuudi väärtuste piirkond. (SNOMED 2021i)

### **2.3.4 Viiteloend (*reference set*)**

Viiteloend (ingl *reference set* või *refset*) ei kuulu SNOMED CT loogilise mudeli koosseisu, vaid on SNOMED CT derivatiiv ning kasutaja poolt kergesti edasiarendatav. Viiteloendi erinevus tavalisest kodeeritud loendist on see, et see ei koosne kood-tekst paaridest, vaid viidetest terviklikele SNOMED CT mõistetele. (SNOMED 2021c)

Viiteloendeid on võimalik SNOMED CT-s koostada erinevatel eesmärkidel. Viiteloendeid kasutatakse keelespetsiifiliste loenditena, kontekstipõhiste loenditena, lihtsate koodiloenditena või erinevate vastendamislõenditena. Viiteloendeid on võimalik koostada rahvusvahelisel, riiklikul või infosüsteemi tasemel. (SNOMED 2021c)

SNOMED CT Eesti riikliku laiendi sisust moodustavad viiteloendid väga suure osa. 2020. aasta novembris välja lastud versioonis on viiteloendeid 23 ning nad sisaldavad kokku ligi 9000 eesti keelde tõlgitud mõistet. (SNOMED 2021i)

## **2.4 SNOMED Internationali tööriistad ja teenused**

SNOMED CT laiendi haldamiseks võib kasutada iseloodud tööriistu, SNOMED Internationali vabatarkvaralisi tööriistu või SNOMED Internationali täisteenust, mis sisaldab nii tööriistu kui ka täiendavat kasutajatuge (SNOMED 2021f).

SNOMED CT terminoloogia kasutamiseks on saadaval palju valmisteenusi ja tarkvarakomponente, mis katavad funktsionaalsuse poolest terve spektri, alates SNOMED CT sisu importimisest andmebaasidesse kuni analüütikatööriistadeni, mis võimaldavad teha ontoloogiaga rikastatud päringuid patsiendi terviseandmete kogumi

pealt. Käesolevas peatükis antakse neist põgus ülevaade, peatudes vaid neil töövahenditel, millel on kanda suurem roll töö praktilises osa kavandamisel ja realiseerimisel.

Kõik SNOMED Internationali poolt pakutavad tööriistad on saadaval vabatarckvaraliste komponentidena ning litsentsitud Apache 2.0 litsentsiga, mis võimaldab nii era- kui ka ärikasutajal neid vastavalt seatud tingimustele tasuta kasutada ja edasi arendada. Kõik tööriistad (sh kasutajaliidesed) on kättesaadavad koodirepositooriumidena SNOMED Internationali GitHubis. (SNOMED 2021e)

SNOMED CT terminoloogiat avaldatakse täisterminoloogiana (sh viiteloendid) kolmel erineval viisil: veebilehitsejas, terminoloogiaserveri teenustena ning masinloetavate failidena (RF2 failid). Kõigi kolme viisi puhul on võimalik kasutada rahvusvahelist versiooni või mõne riigi kohaliku laiendi versiooni. Sealjuures sisaldab iga riigi kohalik versioon ka selle aluseks oleva rahvusvahelise versiooni infot (või viiteid sellele). Nii rahvusvaheline versioon kui ka riiklik laiend avaldatakse kaks korda aastas, kusjuures riigiti on avaldamise aeg erinev. (SNOMED 2021d)

#### **2.4.1 SNOMEDI täisteenus ehk *Managed Service***

SNOMED International pakub enda liikmesriikidele täisteenust laiendi haldamiseks, komponentide lisamiseks ning viiteloendite koostamiseks. Täisteenus koosneb veebipõhistest tööriistadest, kasutajate haldamisest, täiendavast tehnilisest ning sisulisest kasutajatoest ning laiendi avaldamisest kaks korda aastas. Täisteenuse tööriistade hulka kuuluvad:

- sisuloomise töövahend (*Authoring Tool*) mõistete, kirjelduste ja seoste loomiseks (sh esmane valideerimine *Release Validation Network* tarkvaraga);
- viiteloendite ja masstõlkimise töövahend (*Refset and Translation Tool*);
- litsentsihalduse ning laiendijagamise keskkond (*Member Licencing and Distribution System*). (SNOMED 2021f)

Kõik nimetatud tööriistad on vabavartarkvara ning kättesaadavad ka kohandatava tarkvarana SNOMED Internationali GitHub repositooriumides (SNOMED 2021e).

## 2.4.2 Veebipõhised lehitsejad ja terminoloogiaserver

SNOMED CT sisulehitseja on kättesaadav nii integreeritava tarkvaratükina SNOMED Internationali GitHubi keskkonnas kui ka SNOMED Internationali poolt hallatava veebikeskkonnas aadressil <https://browser.ihtsdotools.org>. Selle lehitseja abil on võimalik tutvuda nii rahvusvahelise versiooniga kui ka riiklike laienditega. Samuti on sisulehitseja koosseisus masinloetava mõistemudeli (MRCM) lehitseja. (SNOMED 2021e, 2021i).

Lehitseja võimaldab sirvida mõisteid taksonoomias hierarhiat pidi liikudes, tekstiotsingu põhjal või ka päris keerukate liitpäringute põhjal, milleks kasutatakse spetsiaalset SNOMEDI päringukeelt (SNOMED 2021i). Päringutulemused kuvatakse nii tekstiliste kui ka graafiliste kirjeldustena (joonised 2, 3 ja 4) ning toodud on ka vastendusviited mõnede teiste koodisüsteemi koodidele (nt RHK 10). (SNOMED 2021i)

The screenshot displays the SNOMED CT web browser interface. On the left, a search bar contains the text 'pneumo' and shows '3086 matches found in 0.838 seconds.' Below the search bar is a list of search results with columns for a yellow circular icon, the term name, and its description. The results include: Pneumonia (disorder), Pneumocyte (cell), Pneumouria (finding), Pneumograph (physical object), Pneumovirus (organism), Pneumonitis (disorder), and Haematoloechus (organism). On the right, there are tabs for 'Summary', 'Details', 'Diagram', and 'Expression'. The 'Details' tab is active, showing a hierarchical view of 'Parents' (Lung consolidation (disorder) and Pneumonitis (disorder)) and 'Children (30)' (Abscess of lung with pneumonia (disorder), Aspiration pneumonia (disorder), and Bilateral pneumonia (disorder)). A blue tooltip is visible over the 'Pneumonia (disorder)' entry, displaying its SCTID (233604007) and translations in English ('en Pneumonia' and 'en Pneumonia (disorder)').

Joonis 4. Veebipõhise SNOMED CT lehitseja lihtotsing. (SNOMED 2021i)

SNOMED CT sisulehitseja kasutab Snowstorm terminoloogiaserveri teenuseid. See on avalikuks kasutamiseks mõeldud terminoloogiaserver, mille API liidese kaudu on võimalik pöörduda erinevate laiendite (sh Eesti) eri versioonide poole. Snowstorm

terminoloogiateenus võimaldab SNOMEDI mõisteid pärida ka erinevate seoste ja kirjelduste alusel ning võimaldab ka mõistemudeli päringuid. (SNOMED 2021h)

Päringute vastused on mahukad: nii on mõiste 183992003 |Planeeritav puusaliigese endoproteesimine| kirjeldus JSON objektina üle 1000 rea pikk, sisaldades kõiki kirjeldusi, seoseid ja ontoloogiat. Joonisel 5 on näiteks näha vaid 2% üksikmõiste GET-päringu tulemusest.

```
"descriptions": [  
  {  
    "active": true,  
    "moduleId": "11000181102",  
    "released": true,  
    "releasedEffectiveTime": 20191130,  
    "descriptionId": "296721000181119",  
    "term": "Planeeritav puusaliigese endoproteesimine",  
    "conceptId": "183992003",  
    "typeId": "900000000000013009",  
    "acceptabilityMap": {  
      "71000181105": "PREFERRED"  
    },  
    "type": "SYNONYM",  
    "lang": "et",  
    "caseSignificance": "CASE_INSENSITIVE",  
    "effectiveTime": "20191130"  
  },  
],
```

Joonis 5. Väljavõte Snowstorm terminoloogiateenuse vastusest. (SNOMED 2021h)

Eelnevalt kirjeldatud sisulehitseja ja terminoloogiaserver töötavad SNOMED CT ametlikul versioonil, kus rahvusvahelise versiooni ja laiendite sisu uueneb kaks korda aastas. Soovides näha värskelt loodud uut sisu enne laiendi ametlikku avaldamist, on selleks võimalus Daily Build Browser nimelise sisulehitseja kaudu aadressil <https://dailybuild.ihtsdotools.org>. Selle lehitseja veebiklient on sama, mis ametliku versiooni lehitsejal, ning aluseks olev terminoloogiateenus on analoogne Snowstormi terminoloogiaserveri teenustega. Sisuline uueneb igapäevaselt, kuid SNOMED Internationali tungiv soovitus on kasutada Daily Build lehitsejat vaid informatiivsel eesmärgil. Daily Build versioon ei sisalda uute viitelõendite sisu. (SNOMED 2021g)

### 2.4.3 SNOMED RF2 laiendifailide pakett

Kõige tavalisem SNOMED CT levitamiseviis on masinloetav failipakett, mida saavad registreeritud kasutajad alla laadida litsentsihalduskeskkonnast MLDS. Tegemist on kindla ülesehitusega tekstifailidega, mis asuvad kokkulepitud struktuuriga failide-

kaustade pakettis. Laiendifailide formaadile viidatakse SNOMED Internationali juhendmaterjalides kui spetsiaalsele RF2 (*Release Format 2*) failiformaadile ning oma olemuselt on tegemist tabulaator-eraldusega tekstifailidega. (SNOMED 2021b)

id	effectiveTime	active	moduleId	refsetId	referencedComponentId	acceptabilityId
800df5e0-ee57-4e7d-8427-179592113b31	20191130	1	11000181102	71000181105	275371000181115	900000000000548007
800e2a7d-3d00-46d3-a73b-4d25025eb15f	20191130	1	11000181102	71000181105	295161000181119	900000000000548007
80211cde-65b5-4c1d-b11d-75d1c04a7b61	20191130	1	11000181102	71000181105	221781000181116	900000000000548007
802d394d-ce65-495c-aa30-3732eda9e059	20191130	1	11000181102	71000181105	272191000181114	900000000000548007
803018e2-28ca-42ae-9c24-33a4aa021f49	20191130	1	11000181102	71000181105	295031000181111	900000000000548007

Joonis 6. SNOMED CT RF2 laiendifail: eesti keele viiteloend.

RF2 failid on allalaetavad ühe kokkupakitud kaustana, mis koosneb kolmest identse struktuuriga pakettist:

- *Delta*: viimases versioonis lisandunud, muutunud ja kustutatud kirjed,
- *Snapshot*: kehtiv seis laiendist (kõigi mõistete viimane versioon),
- *Full*: kogu laiendi kehtiv ja ajalooline sisu (SNOMED 2021b: 26).

Iga pakett jaotub omakorda järgmiselt:

- 1) Kaust *Refset* (viiteloend), mis jaguneb kolmeks alamkaustaks:
  - *Content*: viiteloendid, mis koosnevad viidetest SNOMED CT komponentidele;
  - *Language*: keeleloendid, mis koosnevad viidetest kirjeldustele (vt ülal joonis 6);
  - *Map*: viiteloendid, mis sisaldavad vastendusi teistesse koodisüsteemidesse;
  - *Metadata*: viiteloendite metaandmed (nt sõltuvused versioonide vahel);
- 2) Kaust *Terminology* (terminoloogia), mis jaguneb hulgaks failideks:
  - Mõisted;
  - Kirjeldused eri keeltes (Eesti laiendi puhul eesti ja inglise keeles);
  - Seosed ja neist tuletatud OWL väljendid. (SNOMED 2021b: 35–37)

Nii RF2 failide pakett kui ka selles sisalduvad failid on nimetatud kindlate reeglite järgi, mis võtavad arvesse laiendi versiooni ja sisu. Nimetamisreeglid on põhjalikult juhendites lahti kirjutatud (SNOMED 2021b: 28–30).

### **3 Lähteülesande kirjeldus ja analüüs**

Käesolev peatükk annab esmalt laiemat ülevaate hetkeolukorrast, kirjeldades lähemalt probleeme, mis SNOMED CT Eesti laiendi haldamise ja kasutamise raames ette tulevad. Sellele järgneb analüüs lahenduse leidmiseks ning kirjeldus, millistele tingimustele peaks vastama tööriist, mida käesoleva töö raames luuakse.

Analüüsile on lähenetud metoodiliselt ja võimalikult neutraalselt. Kuna töö autor on samal ajal ka nn töö tellija rollis, siis on analüüsi ja arutluse käigus kasutatud teadmisi ja kogemusi, mis on tekkinud isikliku töökogemuse tulemusena ja millele ei ole võimalik viidata. Juhul, kui töö käigus on kasutatud teadmisi, mis on saadud töö kirjutamise käigus kolleegidelt TEHIKus või SNOMED Internationalis, on see viidatud.

#### **3.1 Hetkeolukorra kirjeldus**

SNOMED CT Eesti laiend avaldatakse kaks korda aastas, mai ja novembri viimasel päeval. Laiendi sisulist tööd (mõistete lisamine, tõlkimine, loendite moodustamine) tehakse TEHIKus SNOMED Internationali poolt pakutavate tööriistadega (SNOMED Internationali täisteenus). Erandiks on laiendifailide avaldamiseelne ülevaatus, mille jaoks ei paku eraldi tööriista SNOMED International ja mille jaoks ei ole hetkel tööriista ka TEHIKul.

TEHIK on ka Tervise Infosüsteemi (TIS) haldaja ning avaldab TISis kasutatavaid loendeid, millest paljud on kodeeritud SNOMED CT abil. SNOMED CT koodid on kasutusel ka HL7 CDA standardil põhinevate andmevahetusdokumentides andmeplokkide pealkirjadena. Kõik TISi loendid on avaldatud TEHIKu standardite ja klassifikaatorite publitseerimiskeskuses. (TEHIK 2021a)

TEHIKu 2021–2022 plaanides on uue Teabekeskuse loomine, mis peaks asendama hetkel kasutusel olevat publitseerimiskeskust. Loendite avalikustamiseks on planeeritud nii veebikeskkond kui ka terminoloogiaserver. Teabekeskuse loendite haldamise keskkonna analüüs ja hange on planeeritud 2021. aasta II poolde. (TEHIK 2021b)

Tänases publitseerimiskeskuses ei ole ei haldusmoodulis ega avalikustamismoodulis ühtegi tehnilist seost SNOMED CT tööriistade ega loogikaga, mis tekitab mitmeid probleeme:

- 1) SNOMEDil põhinevaid loendeid hallatakse publitseerimiskeskuses käsitsi, ilma et toimuks koodi või selle tekstilise kirjelduse kontrolli SNOMEDi töövahendite abil. See annab võimaluse vigade tekkeks ning versioonide lahknemiseks.
- 2) Publitseerimiskeskuse loogika suur- ja väiketähe reglementeerimiseks erineb SNOMED CT-st. See võib tekitada asjatut kirjapildi erinevust eri süsteemides.
- 3) SNOMED CT viiteloendid on struktuurilt ja sisult erinevad publitseerimiskeskuse loenditest. Neid hallatakse eraldi, mis tähendab haldajale topelttööd. Samuti on SNOMEDil põhinevate süsteemide arendajatel hädavajalik kasutada nii publitseerimiskeskuse loendit kui ka SNOMEDi viiteloendit ja jälgida, et ei tekiks konflikte kahe sama loendi eri versiooni vahel.

Veel üks ebakõla SNOMED CT ja TIS loendite haldamise praktika vahel on avaldamissageduse erinevus. Publitseerimiskeskuses muudetakse loendeid vastavalt vajadusele, kuid SNOMED CT ametlik versioon avaldatakse kaks korda aastas. Praegune praktika on selline, et kui TISis on vajalik võtta kasutusele uus mõiste, siis valmistatakse see SNOMED Internationali tööriistades ette ning saadud kood võetakse kohe kasutusele. SNOMED CT-s koodi avaldamine toimub aga koos järgmise laiendiga (kuni 6 kuud hiljem). See tekitab probleeme SNOMED CT alusterminoloogia kasutajatele, kelle süsteemidesse satub töödeldamatu kood. Selle vältimiseks tuleks mööda minna SNOMED Internationali täisteenusest ning hakata laiendifailidest tekitama vaheversioone, mida Publitseerimiskeskuse või Teabekeskuse kaudu jagada.

Eelnevatest teemadest eraldi seisev suur probleem on SNOMED CT Eesti laiendi avaldamisele eelnev laiendifailide ülevaatus. Eesti laiendi avaldamise protsessis toimub kuu enne laiendi avaldamist tööriistade lukustamine, seejärel laiendi sisu SNOMED Internationali poolt valideerimine ning laiendifailide koostamine. Umbes kaks nädalat enne uue laiendi avaldamist saadetakse masinloetavad RF2-formaadis laiendifailid TEHIKule üle vaatamiseks ja kontrollimiseks. Selle tarbeks ei ole aga TEHIKus täna ühtegi tööriista. Töö kirjutamise hetkel on TEHIKus avaldamisel neljas SNOMED CT Eesti laiend. Iga uus laiend on mahukam kui eelnevad, mistõttu suureneb ka ülevaadatav andmehulk. Käsitsi või kontoritarkvara vahenditega masinloetavate RF2 formaadis failide ülevaatus on aeganõudev, veaohlik ning infoühiskonna kontekstis ka naeruväärne.

## 3.2 Lahenduse analüüs

Eelnevas peatükis tõstati kolm suurt probleemi:

- 1) SNOMED CT viiteloendite ja Publitseerimiskeskuses/Teabekeskuses avaldatud loendite ühes kohas haldamine ja nende avaldamisel topelttöö minimeerimine;
- 2) TISis uute mõistete kasutuselevõtul tagada, et need asutused, kes kasutavad oma infosüsteemis SNOMED CT terminoloogiat, ei peaks kannatama vigade käes, mida tekitab nn ebakorrektnen SNOMED CT kood (sest kood ei ole veel ametlikult avaldatud);
- 3) Vajalik on luua tööriist avaldamiseelsete RF2 failide ülevaatamiseks.

Probleem number üks tuleb lahendada uue Teabekeskuse analüüsi kontekstis ja ei vaja käesolevas töös pikemat käsitlust, seda eriti arvestades, et diplomitöö maht seda ei võimalda. Seega keskendub analüüs kahele ülejäänud probleemile, mis on mõlemad otseselt seotud SNOMEDi terminoloogiatega ja RF2 formaadis laiendifailide kasutamisega.

### 3.2.1 Terminoloogiatega abil RF2 failide loomine

Selleks, et ise luua RF2 failiformaadis vaheväljalaskeid, on vaja järgmist:

- 1) Luua mõisted sisuloomistööriistas ning kinnitada need projektis kasutamiseks;
- 2) Teha viiteloendi täiendus ning laadida alla vastav viiteloendi fail RF2 formaadis;
- 3) Importida viiteloend andmebaasi;
- 4) Teha viiteloendis viidatud uute komponentide põhjal päringud Daily Build Browseri API suunas, et saada uued mõisted JSON objektidena.
- 5) Töödelda serverrakenduses uued mõisted RF2 failideks (RF2 failide sisu representatsioon relatsioonilise andmebaasi tabelitena on toodud lisa 2).

Sellise funktsionaalsuse kriitilisim koht on see, kas Daily Build Browseri APIst saadud JSON objektid on täielikus vastavuses RF2 failides toodud andmeveergudega ning kui konverteerimist ei tehta otse mälu, siis on vaja ka andmebaasi tabelite ja veergude



vastendamine. Selleks on tarvis analüüsida mõlemat väljundit ning leida seosed. Analüüsi tulemus on toodud lisa 4.

Analüüs näitab, et põhimõtteliselt on võimalik SNOMED CT Eesti laiendis loodud mõisteid enne ametlikku laiendi avaldamist RF2 failideks formuleerida, kuid see ei pruugi olla pikas perspektiivis mõistlik. Värske teadmisega, et SNOMED Internationalil on kavas hakata laiendite avaldamist tegema tihemini (SNOMED 2021k), soovitab autor esialgu oodata, kuni selguvad täpsemad asjaolud selle kohta, kui palju maksaks Eesti laiendi sagedasem avaldamine. Seejärel on võimalik teha uus võrdlev analüüs, milline lahendus oleks funktsionaalselt ja finantsiliselt vastuvõetavam.

Käesolevas töös valmiva prototüübi loomisel arvestatakse võimaliku laiendamisevajadusega, sest vastav funktsionaalsus sobiks selle tööriista juurde hästi, kuid hetkel seda funktsionaalsust ei realiseerita.

### **3.2.2 Avaldamiseelne RF2 failide ülevaatus**

Laiendifailide avaldamiseelne ülevaatus on suhteliselt autonoomne funktsionaalsus, mida on võimalik realiseerida iseseisva rakendusena või integreerida teise süsteemi. Käesolevas töös võetakse eesmärgiks kirjeldada ning realiseerida see funktsionaalsus kui iseseisev töövahend.

**Töövahendi eesmärk** on pakkuda mugavat ja lihtsat võimalust kontrollida SNOMED Internationali poolt saadetavat avaldamiseelset RF2 failide paketti (paketi sisu kirjeldus on toodud peatükis 2.4.3).

**Töövahendi kasutajad** on TEHIKu andmekorraldustalituse töötajad, sh käesoleva töö autor.

**Peamine töövoog.** SNOMED CT Eesti laiendi haldamine toimub SNOMED Internationali tööriistades. Seal luuakse vastavalt vajadusele uusi mõisteid, uusi tõlkeid ning uusi viiteloendeid. Samuti toimub pidevalt viiteloendite sisu täiendamine ja parandamine, mõnevõrra vähem tuleb ette olemasolevate tõlgete parandamist. Kaks korda aastas, mai ja novembri viimasel päeval, laiendi sisu avaldatakse. Kuu enne laiendi avaldamist lukustatakse tööriistad selle päeva seisuga. Kaks nädalat enne laiendi

avaldamise tähtaega saadab SNOMED International TEHIKule ülevaatamiseks uue laiendi sisu RF2 failide paketina. Kirjeldatud tsükkel on graafiliselt kujutatud käesoleva töö lisas 3. Failipaketi ülevaatuseks on ette nähtud kuni 5 päeva. Kontrollida tuleb loodud ja muudetud sisu statistilist ja sisulist vastavust reaalselt tehtud/planeeritud tööle nii mõistete, nende tõlgete ja seoste kui ka viiteloendite lõikes.

**Töövahendi liidestus teiste süsteemidega.** Tööriist peaks suutma olla abiks laiendi sisu valideerimisel – see võib olla moodul loodavas Teabekeskuses, kuid ta võib seista ka eraldi. Organisatsiooni ametliku tööriistana vajab loodav süsteem liidestust TEHIKu kasutajate autentimise süsteemiga ning vastama TEHIKu arendusprofiilile. Loodav süsteem saaks pakkuda lisaväärtust, kui ta suhtleks tulevase Teabekeskusega, tõenäoliselt on vältimatu SNOMED Internationali terminoloogiateenuste kasutamine.

## 4 Tööriista kavand

Tööriista kavandi puhul kirjeldatakse tööriista prototüübi kavandit. Kuna käesoleva töö raames loodav tarkvara ei ole mõeldud asutuse ametlikuks tarkvaraks, on siinsed nõuded veidi erinevad sellest, mida tuleks nõuda reaalses hankes – seda eriti mittefunktsionaalsete nõuete osas.

Prototüüpimise kui analüüsimetoodika eesmärk on visualiseerida funktsionaalsus kasutaja jaoks. Kuigi prototüüpimiseks peetakse ka pelgalt visuaalse kuvandi loomist, on käesolevas töös plaanis teha täisfunktsionaalne prototüüp, mis vastab agiilse evolutsioonilise prototüübi põhimõtetele: prototüüp lähtub ärinõuetest, kuid täpsustab ja parandab neid nõudeid, sest pakub võimalust loodavat süsteemi paremini visualiseerida ning pidevalt testida. Sellisel kujul prototüüp aitab üles leida projekti kitsaskohad ning leida kinnitust, kas katsetatav lahendus on projekti jaoks põhimõtteliselt sobilik. (Girvan, Paul 2017: 155-156) Kavandatav tööriista prototüüp on olemuselt vertikaalne prototüüp, mille eesmärk ei ole mitte niivõrd veebiväljundi tutvustamine kliendile, vaid andme- ja töövoogude defineerimine ning testimine. (Stephens 2015: 288)

Siin kirjeldatud tööriista kavand võimaldab hinnata töö käigus valmiva tarkvara vastavust nõuetele ja püstitatud eesmärgile. Nõuded on kirjeldatud FURPS+ metoodika järgi. See jagab nõuded esmalt funktsionaalseteks ja mittefunktsionaalseteks ning võimaldab nii nõudeid kirjeldada kui ka nende terviklust hinnata. (Stephens 2015: 64–66)

### 4.1 Funktsionaalsed nõuded

F1. Süsteemi on võimalik sisse lugeda RF2 formaadis SNOMED laiendifaile.

F1.1. Süsteemi peab saama lugeda eraldi Delta, Snapshot või Full laiendifaile

F1.2. Süsteemis on võimalik andmebaasi laiendifailide sisu kustutada.

F1.3. Varasem sisu kustutatakse andmebaasist enne uute laiendifailide sisselugemist.

F.2. Kasutajal on võimalik parameetrina ette anda, millises režiimis (Delta, Snapshot või Full) ta andmeid vaadata soovib. Andmete kuvamise eripärad, mis sõltuvad režiimist, lahendatakse vastavalt ärireeglitele, mis on toodud lisa 3.

F3. Süsteemi on võimalik kasutada isikliku kasutajakontoga.

F3.1. Süsteemi on võimalik sisse logida.

F3.2. Süsteemist on võimalik välja logida.

F3.3. Süsteemi on võimalik luua uusi kasutajaid.

F4. Süsteem võimaldab teha mis tahes mõiste kohta üksikpäringut Daily Build terminoloogiateenusest.

F5. Süsteem võimaldab vaadelda laiendifailide sisu komponentide kaupa, arvestades sisselülitatud režiimi ning rikastades andmeid vastavalt lisas 3 toodud ärireeglitele.

F5.1. Süsteem võimaldab vaadata mõisteid sorteeritava-filtreeritava tabelina.

F5.2. Süsteem võimaldab vaadata kirjeldusi sorteeritava-filtreeritava tabelina.

F5.3. Süsteem võimaldab vaadata seoseid sorteeritava-filtreeritava tabelina.

F5.4. Tabelite sisu peab vastama komponentide kaupa ärireeglitele (lisas 3).

F6. Süsteem võimaldab vaadata laiendifailides toodud komponentide statistikat.

F6.1. Iga komponendi (mõiste, kirjeldus, seos) puhul on välja toodud info aktiivsete kirjete, inaktiivsete kirjete kui ka kirjete üldsumma kohta.

F6.2. Kirjelduste puhul on punktis F6.1 kirjeldatud info laiendis kasutatavate keelte kaupa eraldi välja toodud.

F7. Süsteem võimaldab vaadelda laiendifailides olevaid viiteloendeid, arvestades sisselülitatud režiimi ning rikastades andmeid vastavalt lisas 3 toodud ärireeglitele.

F7.1. Süsteem kuvab nimekirja kõigist viiteloenditest.

F7.2. Süsteem kuvab iga viiteloendi kohta statistilise info.

F7.3. Viiteloendi sisu peab olema kuvatud sorteeritava-filtreeritava tabelina.

F7.4. Snapshot režiimis vaadeldav viiteloendi sisu peab olema võrreldav vastava TISi klassifikaatori sisuga.

## 4.2 Mittefunktsionaalsed nõuded

Mittefunktsionaalsed nõuded prototüübile erinevad mõnes punktis sellest, millised oleks nõudmised reaalse töösüsteemi puhul. Mittefunktsionaalsete nõuete kaardistamise aluseks on samuti FURPS+ meetodika: kui eelnevalt kirjeldatud funktsionaalsed nõuded mahuvad FURPS+ meetodika järgi akronüümi esimese tähe F alla, siis akronüümi ülejäänud tähed tähistavad mittefunktsionaalseid nõudeid allpool nimetatud valdkondade kaupa.

### **Kasutatavus (*Usability*)**

- U1. Süsteemi kasutajaliides on selge ja töövoo tundjale intuiitiivselt kasutatav.
- U2. Süsteemi kasutajaliides on ingliskeelne vähemalt SNOMEDi terminite osas.
- U3. Süsteemis ei käidelda konfidentsiaalseid andmeid (v.a sisselogimise andmed).
- U4. Süsteemil on kasutusjuhend; süsteemi API-liides on dokumenteeritud.

### **Töökindlus (*Reliability*)**

- R1. Vigade puhul kuvatakse kasutajale arusaadavad veateated.
- R2. Süsteem nõuab kasutaja autentimist.
- R3. Andmevahetus klient- ja serverrakenduse vahel on krüpteeritud (HTTPS-protokoll).

### **Jõudlus (*Performance*)**

- P1. Süsteem ei ole igapäevases kasutuses. Rakenduse pidev valmisolek ei ole vajalik.
- P2. Süsteemi kasutatakse sisevõrgus (prototüüpi kasutatakse ühes arvutis).
- P3. Süsteemi koormusnõuded on madalad (1–5 üheaegset kasutajat).

### **Toetatavus (*Supportability*)**

- S1. Süsteem teostatakse klient-server rakendusena, mis põhineb REST arhitektuuril.
- S2. Süsteem peab võimaldama SNOMED Internationali RF2 formaadis failide importi.
- S3. Süsteem peab suutma kasutada SNOMED Internationali terminoloogiaserveri APIt.

## **Muud nõuded (+)**

X1. Vaate kujundamisel võetakse eelduseks, et süsteemi kasutatakse eelkõige arvutis ning selle kujundus ei pea sobima väikestele ekraanidele.

X2. Ülevaatlised päringutulemused kuvatakse tabelarvutusprogrammi laadses kompaktses tihedas tabelis, mille veerud on kirjeldatud ärireeglites.

## **4.3 Tööriistade valik**

Arendustööriistade valikul on autoril keskmisest vabamad käed, sest prototüübi loomisel ei ole tingimata vajalik järgida asutuse arendusprofiili või hinnata vahendite pikaajalist perspektiivi. Seega ei ole hädavajalik valida parimat lahendust, vaid mõistlik on valida kiireim lahendus sobilike variantide seast. Kuna tegemist on vertikaalse prototüübiga, mille puhul on ka kood näidislahenduse osa, on oluline, et kood oleks piisavalt selge ja loetav ning lõpplahendus kommenteeritud nii, et reaalse lahenduse tegemisel saaks prototüüpi suhtuda kui dokumentatsiooni, mitte kui kopeerimismaterjali. (Stephens 2015: 290)

Sellest tulenevalt on autor otsustanud enda loodavale prototüübile seada järgmised nõuded:

- arhitektuuriline ülesehitus peab olema ülesande püstituse kohaselt loogiline ja reaalses arenduses taaskasutatav;
- prototüübi arendusvahendid peavad võimaldama objektorienteeritud lähenemist ning kihilist disaini (äri loogika eraldamist andme- ja andmevahetuskihist);
- prototüübi lahenduskäigud peaksid olema pigem konventsionaalsed ja standardsed (*convention over configuration*);
- prototüübi arendamisel tuleb järgida hea jälgitavuse printsiipi:
  - nimetuste järjepidevus koodis, kasutajaliideses ja dokumentatsioonis;
  - kood peab olema kommenteeritud ja liidesed dokumenteeritud.

Autor on otsustanud oma töös kasutada järgmist arendusvahendite komplekti:

- Andmebaas Microsoft SQL;
- Serverrakendus: ASP.NET Core ja .NET Entity Framework (C#);
- Klientrakendus: Aurelia raamistik (TypeScript, HTML, JSON), Bootstrap CSS;
- Välised teenused: SNOMED International terminoloogiateenused;
- Swagger API dokumentatsioon.

Töös kasutatakse projekti eeskuju ja põhjana tarkvara, mis on autori poolt loodud Tallinna Tehnikaülikooli Infotehnoloogia Kolledži aine „Hajussüsteemide ehitamine“ läbimise raames. Programmeerimise käigus ei ole kasutatud muud kõrvalist abi.

## 5 Tehniline realisatsioon

Prototüüp on teostatud klassikalise klient-server võrgurakendusena, mis koosneb Aurelia raamistikul valminud õhukesest klientrakendusest ning ASP.NET Core serverrakendusest. Rakendused suhtlevad HTTPS-protokollil, kasutades REST-põhiseid veebiteenuseid. Et prototüüp ei ole seotud organisatsiooni sisselogimisteenustega, on prototüübi juures ka uue kasutaja registreerimise võimalus.

Klientrakendus on võimalikult vaba ärioloogikast, kuvatavate andmete saamiseks pöördub veebirakendus nii SNOMED Internationali veebiteenuste poole kui ka selle projekti käigus loodud serverrakenduse poole. Otsustus- ja arvutusprotsessid toimuvad valdavalt serverrakenduses. Andmete hoidmiseks kasutatakse MSSQL andmebaasi ning andmete sisselugemiseks SNOMED Internationali poolt ette antud RF2 failiformaati.

### 5.1 Serverrakendus ja andmebaas

Serverrakendus on loodud ASP.NET Core ja Microsoft Entity Framework Core (edaspidi EF) raamistikke kasutades. ASP.NET Core on avatud lähtekoodiga raamistik, mis on disainitud veebirakenduste loomiseks ning mille kasutamine koos EF raamistikuga võimaldab oluliselt tõsta arenduskiirust, rakenduse konformsust ja üldist turvalisust (Arora, Dash 2018: 19, 57). Kontrollrite ülesehituse, autentimise ja autoriseerimise lahenduse puhul on töös järgitud võimalikult standardset lähenemist ja sisseehitatud tööriistu. Andmebaasi loomine on tehtud EF tööriistade abil kood-enne-põhimõttel.

Serverrakenduse lahenduses kasutatakse muuhulgas baasprojekte, mis on maksimaalselt geneerilised ning installeeritud ja imporditud Nuget.org tarkvarapakettidena. Baasprojektid on töö autori poolt loodud 2020. aasta kevadel Tallinna Tehnikaülikooli aine “Hajussüsteemide ehitamine” raames. Baasprojektide eesmärk on kiirendada täislahenduse geneerilise funktsionaalsuse (nt kihtidevaheline objektide lihtvastendus) loomist, et keskenduda ärioloogilisele funktsionaalsusele, andmete haldamisele ning veebiteenustele.

Serverrakenduse ülesehitus põhineb kolmekihilisel struktuuril, mille eesmärk on hoida andmete kiht (*Data Access Layer*) ja ärioloogika kiht (*Business Logic Layer*) üksteisest lahus ja sõltumatud. Kontrollrites ei ole ärioloogilist funktsionaalsust – kontrollierid



kasutavad äriloogetika kihi teenuseid ning ei pöördu otse andmekihi ega andmebaasi poole. Iga kiht näeb ja teab ainult enda naaberkihti hierarhias. Kihilise arhitektuuri eesmärk on tagada parem koodiloetavus, vähendada iga kihi keerukust ning parandada koodi pikaajalist käigushoidmist ja edasiarendamist. (Aroraa, Dash 2018: 18) Prototüübi puhul on olulisemad neist kaks esimest aspekti, kuid analüüsi käigus oli üheks eesmärgiks ka tööriista laiendamisvõimalus, mis toetab samuti kihilise arhitektuuri kasutamise otsust.

Andmebaasi ehitus põhineb lisas 2 toodud skeemil. Samas ei ole tegemist tüüpilise relatsioonilise andmebaasiga, sest andmed pärinevad etteantud failidest, kus seosed on juba loodud, kuid ei pruugi olla terviklikud. Näiteks Delta paketi failide puhul on süsteemis vaid muudatused: kui mõiste (*concept*) on varasemalt loodud ja muutunud/lisandunud on vaid mõistega seotud kirjeldus (tüüpiliselt lisatud eestikeelne tõlge), siis on kirjelduse andmetes küll viide mõistele, kuid mõistete hulgas antud id-ga rida ei ole. Seega seosed tabelite vahel on küll sisuliselt olemas, kuid nende seoste kasutamist reguleerib vaid äriloogetiline teadmine. Snapshot paketi failide puhul on seoseid olemas rohkem, kuid ka seal ei ole kõiki viidatud komponente. Samuti peab arvestama, et tööriista eesmärk on leida puudujääke ja vigu alusmaterjalis, mistõttu ei saa süsteemi tööd teha sõltuvaks alusandmete terviklusest ja loogilisusest. Süsteem peab küll võimaldama vigu leida, kuid mitte ise vea tõttu tööd katkestama.

Teine mainimist vajav aspekt on see, et andmebaasi tabelites kasutatavad identifikaatorid (skeemil SCTID) ei ole käsitletavad primaarvõtmetena. Primaarvõtmeks võib olla komposiitvõti *conceptId* + *effectiveTime* ning seda soovitatakse kasutada andmete salvestamisel. Samas ei talleta meie süsteem andmeid, vaid tegeleb ainult terminoloogia alusfailidega ning neis tehakse viited *conceptId*-le ilma, et kasutataks komposiitvõtit välisvõtmena. (SNOMED 2021j: 16) Kehtivusaja järgi on siiski võimalik need seosed leida, kuid praktiline vajadus võib selle järgi olla vaid üksikutes Full paketi vaadetes.

Serverrakenduse API dokumentatsioon on loodud Swagger API dokumentatsiooni tarkvara kasutades ning see kirjeldab võimalikke API päringuid ning nende vastuseid, kusjuures päringu vastuseks on reeglina andmeedastusobjekt (*Data Transfer Object*).

## 5.2 Klientrakendus ja SNOMED CT terminoloogiateenused

Veebiklient on loodud Aurelia raamistikku kasutades. Ka veebirakendus on üles ehitatud kihiliselt, nii et domeen, teenused/päringud ning vaated on üksteisest eraldi. Rakendus on kirjutatud valdavalt TypeScript keeles, mis on oma olemuselt tüübikindel JavaScript ning toetab ühest küljest kõiki JavaScripti funktsioone ning teisest küljest võimaldab objektorienteeritud lähenemist. (Fenton 2018: 1, 99–100)

TypeScripti kasutamine parandab koodi jälgitavust ning lihtsustab eri kihtide ja rakenduste vahel andmete edastamist. Koodi kirjutamise ajal võimaldab TypeScript paremaid *autocomplete*-võimalusi, tüübikindlaid objektide massiive ning väärtuste automaatset kontrollimist. (Fenton 2018: 6–7)

Klientrakendus pöördub andmete saamiseks kolme erineva serveri API poole: 1) prototüübi baasserverrakendus on kasutatav läbi localhost:5001 pordi ning sealt on võimalik pärida kontrollitavate laiendifailide sisu ja statistikat ning autentimisinfot; 2) SNOMED CT Daily Build API poole pöördutakse siis, kui on vaja pärida mõisteid, mis ei ole veel kättesaadavad Eesti laiendi ametlikus versioonis, ning 3) SNOMED CT Snowstorm terminoloogiateenus võimaldab pöörduda viimase avaldatud rahvusvahelise versiooni poole, mis on Eesti laiendi aluseks (avaldamisintervalle ja allikate valimise loogikat kirjeldatakse lisa 3).

## 5.3 Rakenduse töövoog ja näited

Alljärgnevalt on toodud mõned illustreerivad näited rakenduse töö kohta nii kasutaja vaatest kui ka funktsionaalsust tagava lähtekoodi kohta.

### 5.3.1 Muudatuste statistika töövoog

Laiendi ettevalmistustööde lõppedes genereeritakse SNOMED Internationali tööriistas raportid, mis arvutavad kokku tehtud muudatused komponentide kaupa. Selleks, et saada esimene hinnang, kas kõik mõisted, kirjeldused ja seosed on jõudnud sisuloomistööriistast laiendifailidesse, on projektis loodud statistika vaatamise funktsionaalsus, mille ekraanipilt on toodud alljärgneval joonisel 7.

## Statistics

Component	Active	Inactive	Total	
Concepts	2	0	2	<a href="#">Concept List</a>
Descriptions	4	0	4	<a href="#">Description List</a>
- including Estonian	2	0	2	
- including English	2	0	2	
Relationships	0	0	0	<a href="#">Relationship List</a>

Joonis 7. Klientrakenduse veebitõmmis: menüü ja statistika kuvamise leht.

Joonisel 7 nähtaval lehel vasakul olev “Delta View” hoiab kasutajal meeles ja rakendusel mälus, et vaatluse all on muudatuste failid (Delta pakett). Andmete kuvamise loogika võib kohati erineda vastavalt sellele, milline vaade on sisse.

Kasutajaliideses on statistika päringud implementeeritud ühe baaspäringuna ning muutub vaid pöördumise aadress, näide selle kohta on toodud joonisel 8. Kasutatud on asünkroonset andmete pärimist, sest lehe laadimisel käivitatakse sarnaseid päringuid mitu (ning mõne lehe puhul ka mitme serveri suunas) ning see võimaldab lehe laadimise aega kokku hoida.

```

async getDescriptionStatisticsByLanguage(languageCode: string): Promise<IFetchResponse<IStatistics>> {
  this._baseUrl = "statistics/getDescriptionStatisticsByLanguage/" + languageCode;
  return (await this.getStatistics());
}

```

Joonis 8. Klientrakenduse koodinäide

Muudatuste statistika jaoks on serverrakenduses oma kontrolleri ning spetsiaalne andmeedastusobjekt StatisticsDTO. Erinevate komponentide statistika jaoks (sh viiteloendite statistika küsimine viiteloendi id abil ning kirjelduste küsimine keele koodi abil) on kontrolleri eraldi pöörduspunktid. Alljärgneval joonisel (joonis 9) on näha kontrolleri meetod, mis saab sisendiks keele koodi ning pöördub statistika objekti koostamiseks äri loogika kihi poole.

```

// GET: api/Statistics/GetDescriptionStatisticsByLanguage/et
[Route(template: "[action]/{languageCode}")]
[HttpGet]
new *
public async Task<ActionResult<StatisticsDTO>> GetDescriptionStatisticsByLanguage(string LanguageCode)
{
    var Stat:Statistics = await _bll.Descriptions.CalculateStatsByLanguage(LanguageCode);

    return Ok(_mapper.Map(Stat));
}

```

Joonis 9. Serverrakenduse koodinäide: statistika kontroller

Statistika arvutamine toimub ärioloogikakihi teenustes, kuid statistika ei ole andmebaasiobjekt – see objekt eksisteerib vaid ärioloogikakihis ning veebiteenus. Statistika arvutamine on üks väheseid funktsioone, mis töötab täpselt ühtemoodi nii Delta, Snapshot kui ka Full vaates. Ärioloogika kihi meetod on näha joonisel 10.

```

0+1 usages riindstr *
public async Task<Statistics> CalculateStatsByLanguage(string lang)
{
    var stats = new Statistics();
    var baseList:IEnumerable<Description> = (await Repository.GetAllAsync()).Where(e:Description => e.LanguageCode == lang);
    stats.Total = stats.Active = stats.Inactive = 0;

    var descriptions:List<Description> = baseList.ToList();
    stats.Total = descriptions.Count();
    stats.Active = descriptions.Count(e:Description => e.Active == true);
    stats.Inactive = descriptions.Count(e:Description => e.Active == false);

    return stats;
}

```

Joonis 10. Serverrakenduse koodinäide: ärioloogika kihi statistika arvutamise meetod

### 5.3.2 Viiteloendite ülevaatus

Üks olulisemaid ülevaatusobjekte on laiendi viiteloendid (ingl lühend *refset*). Nende kohta ei teki sisuloomistööriistas raporteid, vaid muudatused dokumenteeritakse jooksva töö käigus. Viiteloendite ülevaatusel algab protsess sellest, et kasutajale kuvatakse viiteloendite nimekiri:

- 1) Kasutajaliides pöördub serverrakenduse API poole, et küsida viiteloendite nimekiri.

- 2) Serverrakenduse kontrolleri poole pöördub äriloogika kihi poole ja saab nimekirja unikaalsetest id-dest, mis on Refset tabelis SCTID väljal ja saadab selle kasutajaliidesele.
- 3) Kasutajaliides küsib iga koodi kohta sellega seotud kirjelduse andmed SNOMED CT terminoloogiateenusest. Kasutajale kuvatakse viiteloendite nimekiri.
- 4) Kui kasutaja vajutab nuppu „Show statistics“, siis pöördutakse uuesti serverrakenduse poole ning küsitakse iga viiteloendi kohta ka selle statistika info (vt joonised 8–10). Kasutajale kuvatakse viiteloendid koos statistikaga nagu näidatud joonisel 11.

SNOMED CT TOOL Snapshot View Concepts Daily Build Search Descriptions Refsets Statistics Relationships Logout

### List of Refsets

Show statistics

Name	FSN	Active	Inactive	Total
Eesti histoloogilise diferentseerumise astme klassifikaator	Estonian tumor histological grading reference set (foundation metadata concept)	16	0	16
Eesti RhD kuuluvuse klassifikaator	Estonian RhD reference set (foundation metadata concept)	3	2	5
Eesti ABO veregrupi klassifikaator	Estonian ABO blood group reference set (foundation metadata concept)	5	0	5

Joonis 11. Klientrakenduse veebitõmmis: viiteloendite nimekiri.

Kasutaja valib viiteloendi, mille andmeid soovib avada.

- 1) Kasutajaliides pöördub serverrakenduse API poole, et küsida viiteloendi sisukirjed.
- 2) Serverrakenduse kontrolleri poole pöördub äriloogika kihi poole ja saab nimekirja kõigist Refset tabeli kirjetest, kus SCTID väljal on küsitud id. Kontrolleri poole edastab küsitud andmed RefsetDTO massiivina.
- 3) Kasutajaliides pöördub uuesti SNOMED CT terminoloogiateenuse API poole ja küsib iga viiteloendi sisuks oleva mõiste kohta selle eestikeelse nimetuse ja ingliskeelse täisnimetuse.
- 4) Rikastatud andmed kuvatakse tabelina, mis on näha joonisel 12.

## Eesti külviuringu vastuse klassifikaator

Show 10 entries

Search

SCT Code <small>tl</small>	Preferred term <small>tl</small>	FSN <small>tl</small>	Effective Time <small>tl</small>	Active <small>tl</small>
10828004	Positiivne	Positive (qualifier value)	20210530	true
260385009	Negatiivne	Negative (qualifier value)	20210530	true
62604006	Kontaminatsioon	Contaminated (qualifier value)	20210530	true
386134007	Kliiniliselt oluline leid	Significant (qualifier value)	20210530	true
371928007	Kliiniliselt ebaoluline leid	Not significant (qualifier value)	20210530	true
264868006	Kasv puudub	No growth (qualifier value)	20210530	true

Joonis 12. Klientrakenduse veebitõmmis: viiteloendi sisu.

### 5.3.3 Teised vaated ja funktsioonid

Lisaks muudatuste statistikale on prototüüpprakenduses ka teisi funktsioone. Järgneval joonisel on näha kirjelduste ülevaatamise veebivaade klientrakenduses. Tabelid on otsitavad, sorteeritavad ning rohkemate andmeridade puhul on andmed jagatud lehekülgedeks. ID-dele vastavad väärtused, mida tabelis kuvatakse, on vastavalt lisas 3 toodud ärireeglitele küsitud kas laiendifailidest või SNOMED Internationali terminoloogiateenuste abil. Seda, kas ülevaatamisel on Delta, Snapshot või Full andmefailid, märgib kasutaja kasutajaliideses.

Joonisel 13 näha olev tabel on otsitav ja sorteeritav, veergudes *Type* ja *Case* näidatakse toorandmetes olevatele id-dele vastavaid ingliskeelseid eelistermineid, mis on küsitud SNOMED CT terminoloogiaserverist (rahvusvaheliste mõistete kirjeldusi Eesti RF2 failipaketis ei ole).

## All Descriptions

Show 10 entries

Search:

Concept <small>tl</small>	SCT Code <small>tl</small>	Term <small>tl</small>	Lang <small>tl</small>	Type <small>tl</small>	Case <small>tl</small>	Active <small>tl</small>
611000181104	304061000181114	Arthroderma melis (organism)	en	Fully specified name	Case sensitive	true
611000181104	304071000181116	Arthroderma melis	en	Synonym	Case sensitive	true
181000181102	298121000181117	Eesti proovimaterjali uuringu paikme klassifikaator	et	Synonym	Case sensitive	true
161000181109	296561000181116	Eesti RhD kuuluvuse klassifikaator	et	Synonym	Case sensitive	true

Joonis 13. Klientrakenduse veebitõmmis: kirjelduste ülevaatus.

Mõistete ja seoste ülevaatus tabelite kuvamine toimub analoogselt kirjelduste ülevaatus tabelile – eri komponentide puhul on tabelivaade veidi erinev, kuid kuvamise loogika on suhteliselt sarnane. Mõistete, kirjelduste ja seoste jaoks on serverrakenduses tehtud eraldi objektid, pöörduspunktid ja teenused.

Prototüüpsüsteemil on ka oma kasutajate registreerimise ja sisselogimise teenus, mis on realiseeritud EF Core standardlahendust kasutades.

RF2 failipaketi sisselugemine on prototüüprakenduses lahendatud serverrakenduses, kus on võimalik rakenduse käimapaneku hetkel käivitada ka meetod andmete sisselugemiseks ettemääratud asukohast.

## 6 Tulemuse analüüs

Töö tulemusena valmis funktsionaalne prototüüp, mida on võimalik kasutada ajutise töövahendina. Prototüübi loomine oli väga kasulik ka analüüsi võtmes, sest täisfunktsionaalse prototüübi loomise käigus tekkis mitmeid küsimusi, mida analüüsi käigus sellise täpsusega ette ei näinud. Olulisemad asjaolud, mis analüüsi käigus tundusid pisidetailid, kuid realisatsiooni käigus osutusid suuremateks küsimusteks, on lühidalt järgmised.

Esiteks arhitektuuri valik. Igas veebivaates kehtivad erinevad reeglid selle kohta, millisest allikast seotud andmeid päritakse: avaldamiseelsed failid ei sisalda rahvusvahelisi mõisteid, kuid kirjeldused võivad neile viidata; Delta failid ei sisalda mõisteid, mis pole muutunud; Snapshot failid ei sisalda varem inaktiveeritud mõisteid jne. Arhitektuuri valides autor alahindas sellise ärioloogika hulka ning selle tulemusena tekkis klientrakendusse rohkem ärioloogikat, kui autor planeeris. Samuti oli Daily Build terminoloogiaserveri võimekus mõnevõrra oodatust madalam ning testimine näitas, et võimaluse korral tuleks eelistada muid allikaid, et vältida masspäringuid Daily Buildi suunas. Tulevikus võiks kõik andmete sidumised toimuda serverrakenduses. See, kas sidumistoiminguid saaks kiirendada kogu rahvusvahelist väljaannet andmebaasi laadides, nõuab täiendavat analüüsi ja professionaalset arenduskeskkonda.

Eelnevaga seotud on ka teine aspekt, mis puudutab teatud väiksemõõdulisi loendeid (*CaseSignificance*, *Type*, *RelationshipType* jne). Kõik selliste väikeloendite elemendid on omakorda SNOMED CT mõisted, mida prototüüplahenduses küsitakse terminoloogiaserveri teenuste abil. Süsteemi toimimist kiirendaks nende kirjade hoidmine andmebaasis, mis tähendaks loendite automaatse sünkroniseerimise vajadust rahvusvahelisest väljaandest.

Kolmandaks aspektiks olid üllatuslikult SNOMEDI koodid, mis on SQL juhendi järgi *bigint* tüüpi, SNOMEDI terminoloogiateenustes aga *string*-tüüpi. Koodid on 6–18-kohalised (SNOMED 2021b: 134) ning ei pruugi seega olla primitiivsete andmetüüpidega töödeldavad. Kuna koodidel numbrilises väärtuses mõõdetavat või võrreldavat sisu ei ole, kasutati lõplikus lahenduses läbivalt *string*-tüüpi, et vältida korduvat konverteerimist. Samalaadne probleem esines ka kuupäevadega (näiteks joonisel 5 on näha, et



*releasedEffectiveTime* ja *effectiveTime* on sama väärtuse, kuid erineva andmetüübiga).  
Ka siin otsustas autor *string*-andmetüübi kasuks.

Prototüüplahendus vastas kavandis (vt peatükk 4) toodud funktsionaalsetele ja tehnilistele nõuetele. Teatud funktsioone oleks võimalik täiendada ja laiendada, et suurendada kasutajamugavust või pakkuda lisaväärtust, kuid nõuetes toodud põhipunktid on täidetud.

Valmislahenduse koodijaotuvust on võimalik hinnata GitHub analüütikatööriistaga, mille järgi on koguprojektis (serverrakenduses ja klientrakenduses kokku) programmeerimiskeeli kasutatud järgnevas mahus:

- 61% C#,
- 30% TypeScript,
- 5% HTML,
- 4% CSS.

Prototüübi töökindlust oli võimalik testida 2021. aasta maikuu laiendi avaldamiseelsetel RF2 failide ülevaatusel. See testimine tõestas, et loodud rakenduse kasutamine lihtsustas ja kiirendas oluliselt failide ülevaatamise protsessi.

## 7 Kokkuvõte

Käesoleva töö eesmärk oli pakkuda lahendus SNOMED CT laiendi avaldamiseelsete RF2-formaadis failide ülevaatamiseks ja tervikluse kontrolliks. Lahendus kätkes endas nii laiemat konteksti hõlmavat analüüsi kui ka täisfunktsionaalset prototüüpi.

Töö esimeses osas anti põgus ülevaade terminoloogiahalduse üldistest põhimõtetest ning veidi põhjalikum ülevaade SNOMED CT olemusest ja avaldamisviisidest. See teoreetiline põhi oli taustsüsteemiks töö teisele osale, mis sisaldas endas SNOMED CT Eesti laiendi avaldamiseks tehtava töö analüüsi ja hetkeolukorra kirjeldust. Töö järgmises osas esitleti planeeritava tööriista kavandit, mida kirjeldati FURPS+ meetodikale tuginedes. Sellele järgnes prototüübi realisatsiooni kirjeldav peatükk, kus kirjeldati loodud prototüüpi nii arhitektuurilise ülesehituse kui ka funktsionaalsete tööprotsesside vaatest. Töö sisulise osa lõpetuseks esitati prototüübi ja selle valmimiseks tehtud programmeerimistöö põgus analüüs, mis juhtis tähelepanu mõnede keerukamatele kohtadele, mis töö käigus selgusid.

Töö tulemusena valminud täisfunktsionaalne prototüüp on töötav tarkvara, mida on võimalik talle püstitatud eesmärgil kasutada. Tegemist ei ole asutuse ametliku tarkvaraga ning ta ei vasta kõigile nõuetele, millele reaalne ja legaalne asutuse tarkvara peaks vastama. Samas on ta funktsionaalsuse ja dokumenteerimise poolest piisavalt põhjalik, et anda selge ettekujutus, milline peaks valmistarkvara olema ning millistele aspektidele tuleks sarnase arenduse planeerimisel ja läbiviimisel erilisel rõhku panna.

Töö autori jaoks oli lõputöö kirjutamine huvitav protsess, mille käigus saadud teadmised on väärtuslikud nii üldharivas plaanis kui ka igapäevases töös. Töö suurim keerukus oli SNOMED CT kui süsteemi toimimise mõistmine arendaja vaatenurgast – see on pädevus, mida Eestis on täna vaid üksikutel arendajatel ja analüütikutel, mistõttu oli prototüüplahenduse loomine enne suuremaid ametlikke arendusi väga kasulik. Töö autor kasutas loodud rakendust ka SNOMED CT Eesti 2021. aasta maikuu laiendi ettevalmistamisel ning sai kinnitust, et selline tööriist on laiendifailide ülevaatamisel oluliseks abiks.

## Kasutatud kirjandus

**Aroraa, Gaurav; Dash, Tadit (2018).** Building RESTful Web Services with .NET Core. Packt Publishing. Birmingham.

**Benson, Tim; Grieve, Grahame (2021).** Principles of Health Terminology. FHIR, HL7 and SNOMED CT. Fourth Edition. Springer Nature Switzerland AG.

**Bhattacharyya, S. B. (2016).** Introduction to SNOMED CT. Springer. Singapore.

**Cimino, J.J. (1998).** Desiderata for controlled Medical Vocabularies in the Twenty-First Century. *Methods Inf med.* 1998; 37: 394-403.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415631/> (17.05.2021)

**Elkin, Peter L. (ed) (2012).** Terminology and Terminological Systems. Springer.

**Fenton, Steve (2018).** Pro TypeScript: Application-Scale JavaScript Development. Springer. New York.

**Girvan, Lynda; Paul, Debra (2017).** Agile and Business Analysis. Practical Guidance for IT Professionals. BCS Learning & Development Ltd.

**Jürimaa, Villu (2016).** Ravitööd toetava infosüsteemi reeglipõhine otsustustoe tugisüsteem. Magistritöö. Tallinna Tehnikaülikool.

**Karu, Kalev (2011).** SNOMED CT – elektroonilise haigusloo loomise vahend. *Eesti Arst* 2011; 90(10): 466–473. <https://eestiartst.ee/snomed-ct-elektroonilise-haigusloo-loomise-vahend/> (17.05.2021)

**Ravimiamet (2019).** Farmaatsiaterminoloogia. Teine, täiendatud trükk. Ravimiamet. [https://www.ravimiamet.ee/sites/default/files/farmaatsiaterminoloogia\\_2019.pdf](https://www.ravimiamet.ee/sites/default/files/farmaatsiaterminoloogia_2019.pdf) (17.05.2021)

**Ross, Peeter (2011).** Data Sharing and Shared Workflow in Medical Imaging. Thesis on Mechanical Engineering E63. TUT Press.

**SNOMED 2021a.** SNOMED International Members. SNOMED International. <https://www.snomed.org/our-customers/members> (17.05.2021)

**SNOMED 2021b.** SNOMED CT Release File Specification. SNOMED International. <https://confluence.ihtsdotools.org/display/DOCRELFMT/> (17.05.2021)

**SNOMED 2021c.** Practical Guide to Reference Sets. SNOMED International. <https://confluence.ihtsdotools.org/display/DOCRFSPG/> (17.05.2021)

**SNOMED 2021d.** SNOMED CT Starter Guide. SNOMED International. <https://confluence.ihtsdotools.org/display/DOCSTART/> (17.05.2021)

**SNOMED 2021e.** SNOMED International GitHub. <https://github.com/IHTSDO> (17.05.2021)

**SNOMED 2021f.** Software and Tools - Managed Service. SNOMED International. <https://www.snomed.org/snomed-ct/software-tools> (17.05.2021)

**SNOMED 2021g.** Daily Build SNOMED CT Browser. Snomed International. <https://dailybuild.ihtsdotools.org> (17.05.2021)

**SNOMED 2021h.** SNOMED Snowstorm Swagger API Documentation. <https://snowstorm.ihtsdotools.org/snowstorm/snomed-ct/swagger-ui.html> (17.05.2021)

**SNOMED 2021i.** SNOMED CT Browser. SNOMED International. <https://browser.ihtsdotools.org/> (17.05.2021)

**SNOMED 2021j.** SNOMED CT SQL Practical Guide. SNOMED International. <https://confluence.ihtsdotools.org/display/DOCSQLPG> (17.05.2021)

**SNOMED 2021k.** Managed Service User Group 21.04.2021. SNOMED International April 2021 Business Meetings. Virtual Conference. 19.-22.04.2021.

**SNOMED 2021l.** SNOMED CT Editorial Guide. SNOMED International. <https://confluence.ihtsdotools.org/display/DOCEG/> (17.05.2021)

**Stephens, Rod (2015).** Beginning Software Engineering. John Wiley & Sons, Indianapolis.

**Tarkin, Kerli (2013).** Possibilities to Improve Semantic Interoperability in Medical Data Exchange in Estonia with SNOMED CT: Analysis of Discharge Summary. Master Thesis. Tallinna Tehnikaülikool.

**TEHIK 2021a.** Standardite ja klassifikaatorite publitseerimiskeskus. Tervise ja Heaolu Infosüsteemide Keskus. <http://pub.e-tervis.ee/> (17.05.2021)

**TEHIK 2021b.** Teabekeskus. Tervise ja Heaolu Infosüsteemide Keskus. <https://tehik.ee/teabekeskus> (17.05.2021)

**WHO 2021.** International Statistical Classification of Diseases and Related Health Problems (ICD). <https://www.who.int/standards/classifications/classification-of-diseases> (17.05.2021)

## **Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

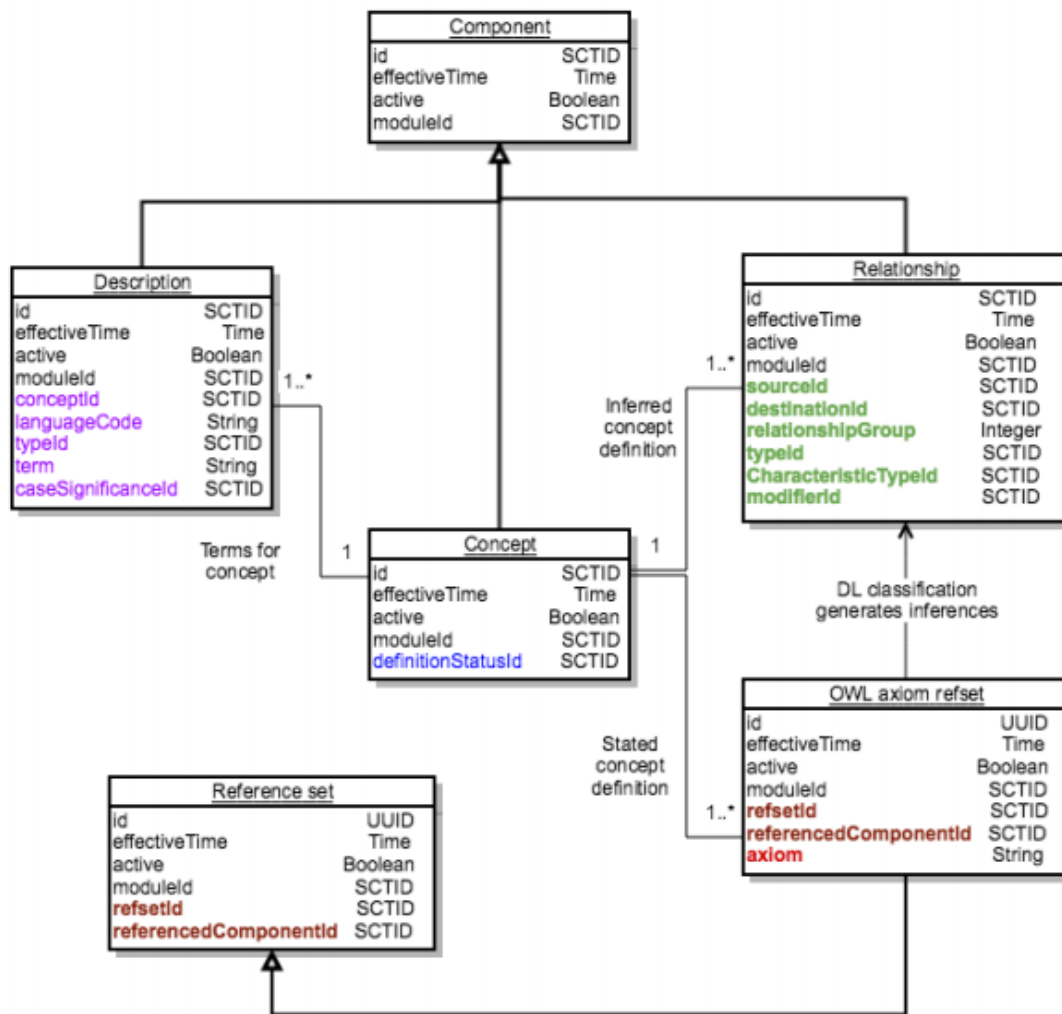
Mina, Rutt Lindström

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „SNOMED CT Eesti laiendi failide ülevaatus tööriist“, mille juhendaja on Priit Rospel,
  - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

17.05.2021

## Lisa 2 – SNOMED CT RF2 failide representatsioon andmebaasi tabelitena

Alljärgneval joonisel on SNOMED CT komponendid ja viiteloendid kirjeldatud kujul, mis aitab neid mõtestada andmebaasi tabelite või RF2 failide kontekstis (SNOMED 2021b: 18).

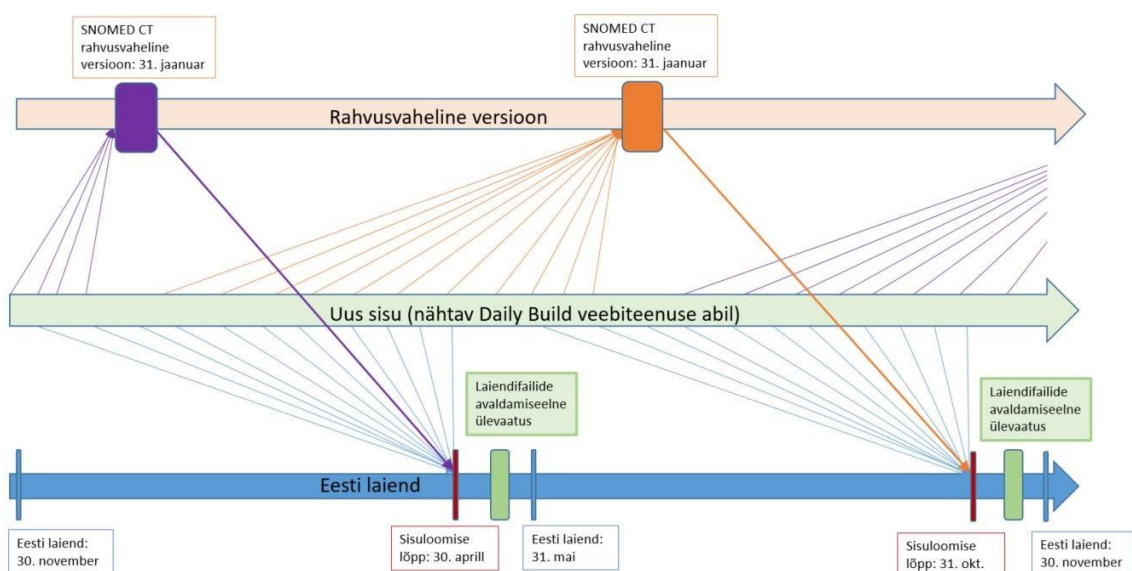


Joonis 14. SNOMED CT ülesehitus SQL ja RF2 jaoks. (SNOMED 2021b: 18)

## Lisa 3 – SNOMED CT Eesti laiendis viidatud sisu ning seotud ärireeglid

### Eesti laiendi sisu kujunemine

Alljärgnevalt (joonisel 15) on graafiliselt kujutatud, millest koosneb Eesti laiendi RF2-failides publitseeritud sisu – seal on Eesti laiendis loodud mõisted ja kirjeldused, mis põhinevad viimase avaldatud rahvusvahelise versiooni sisul. Joonise ülemisel ribal on rahvusvaheline versioon, alumisel ribal Eesti laiend ning keskmisel ribal on näha Daily Build terminoloogiateenusu sisu, mis uueneb sagedamini kui ametlikud versioonid.



Joonis 15. SNOMED CT Eesti laiendi sisu kujunemine ja viitamine.

Sellest tulenevalt on Eesti laiendi failides viidatud komponendid korrektselt kättesaadavad eri allikatest. Olukorra teeb keerulisemaks see, et laiendifailide ülevaatus toimub enne laiendi avaldamist ning et Daily Build terminoloogiateenus, mis on kõige lihtsamini kasutatav allikas, ei pruugi taluda masspäringuid ning seetõttu tuleks hoiduda selle üleekspluaterimisest.

Alljärgnevalt on kirjeldatud tööriista minimaalset andmekoosseisu komponentide lõikes, kusjuures kirjeldatud on ka andmete rikastamisel kasutatavad allikad. Need nõuded ei ole absoluutsed, kuid tagavad võimalikult suure korrektsuse ilma ülearuse keerukuseta.

### **Mõiste (*concept*)**

- Minimaalne andmekoosseis ülevaatus tabelis: mõiste SCTID, mõiste eelistermin eesti keeles, mõiste FSN inglise keeles, mõiste defineeritus (*definition status*) ja olek (kehtiv või kehtetu).
- Mõiste eelistermin ja FSN on alati olemas RF2 failides (kirjeldused).
- Defineerituse mõiste tekstiline kirjeldus viimasest avaldatud rahvusvahelisest versioonist (soovitav eellaadida).
- *Full* režiimi korrektseks kuvamiseks peab arvestama kõigi komponentide kehtivusaegadega ja neid ka kuvama. *Full* režiimi laiendi ülevaatus ajal tavaliselt ei kasutata ning selle osas võib vaadet lihtsustada.

### **Kirjeldus (*description*)**

- Minimaalne andmekoosseis ülevaatus tabelis: kirjelduse SCTID, kirjeldus, keel, kirjelduse tüüp, tõstutundlikkus, kehtivus, seotud mõiste SCTID.
- Kirjelduse tüübile ja tõstutundlikkusele vastavad tekstilised kirjeldused viimasest avaldatud rahvusvahelisest versioonist (soovitav eellaadida).
- *Full* režiimi puhul on soovitav lisada kehtivusajad.

### **Seos (*relationship*)**

- Minimaalne andmekoosseis: seose SCTID, seoses osalev lähtemõiste, seoses osalev sihtmõiste, seose tüüp, seose grupp, kehtivus.
- Seose tüübile vastav tekstiline kirjeldus viimasest rahvusvahelisest versioonist (soovitav eellaadida).
- *Delta* vaade:
  - Lähtemõiste tekstiline kirjeldus valdavalt RF2 failist, selle puudumisel viimasest rahvusvahelisest laiendist ning selle puudumisel viimasest Eesti laiendist (võimalik väikest riski aktsepteerides lihtsustada, kui RF2 failist puudumise korral küsida andmeid otse Daily Buildist).
  - Sihtmõiste tekstiline kirjeldus tavaliselt viimasest rahvusvahelisest laiendist, selle puudumisel Daily Buildist.
- *Snapshot* vaade:
  - Lähtemõiste tekstiline kirjeldus valdavalt RF2 failist, selle puudumisel Daily Buildist.



- Sihtmõiste tekstiline kirjeldus valdavalt viimasest rahvusvahelisest versioonist, selle puudumisel Daily Buildist.
- *Full* vaates tuleb arvestada kehtivusaegadega ning need ka kuvada. Seoste ülevaatomiseks *Full* vaadet reeglina ei kasutata ning selle vaate puhul võib teha mugandusi.

### **Viiteloend (*refset*)**

- Minimaalne andmekoosseis:
  - Viiteloendite nimekiri: viiteloendi nimi eesti keeles, viiteloendi FSN (inglise keeles) ja statistika viiteloendi sisu kohta (aktiivsete, inaktiivsete ja kõigi mõistete arv).
  - Iga viiteloendi kohta: viiteloendis sisalduva mõiste SCTID, vastava mõiste eelistermin eesti keeles, FSN inglise keeles, kehtivuse kuupäev ja kehtivus.
- Viiteloendi nimi *Delta* vaates Daily Buildist, *Snapshot* ja *Full* vaates RF2 failist.
- Viiteloendi statistika RF2 failidest.
- Viiteloendi sisalduvate mõistete kirjeldused Daily Buildist.

## Lisa 4 – Vastavustabel JSON ja RF2 representatsioonide vahel

Tabel 1. Vastavus Snowstorm terminoloogiategenuse vastuse ning RF2 faili vahel.

JSON rada	RF2 faili rada	RF2 tulp	Selgitus
conceptId	Concept	id	
[puudub]	Concept	effectiveTime	saab märkida koostamise kuupäeva
fsn (fsn.term, fsn.lang)		-	sünonüümides
pt (pt.term, pt.lang)		-	sünonüümides
active	Concept	active	alati "true", kui ei ole inaktiveeritud
released		-	kui ei ole veel avaldatud, siis "false"
moduleId	Concept	moduleId	eeldatavasti Eesti
definitionStatus	Concept	(definitionStatus Id)	JSONis tekstina: nt PRIMITIVE
descriptions[x].active	Description-et; Description-en	active	
[puudub]	Description-et; Description-en	effectiveTime	saab märkida koostamise kuupäeva
descriptions[x].moduleId	Description-et; Description-en	moduleId	
descriptions[x].released		-	
descriptions[x].descriptionId	Description-et; Description-en	id	Refset/Language/RefsetLanguage componentId
descriptions[x].term	Description-et; Description-en	term	
descriptions[x].conceptId	Description-et; Description-en	conceptId	
descriptions[x].typeId	Description-et; Description-en	typeId	

descriptions[x].acceptabilityMap (languageId : acceptability)	Refset/Language/RefsetLanguage	acceptabilityId	JSONis keele id ning sõnaline väärtus
descriptions[x].type		-	JSONis sõnaline, aga olemas ka typeId
descriptions[x].lang	Description-et	languageCode	et
descriptions[x].lang	Description-en	languageCode	en
descriptions[x].caseSignificance	Description-et; Description-en	(caseSignificanceId)	JSONis sõnaline, nt CASE_INSENSITIVE
relationships[x].active	Relationship	active	
relationships[x].moduleId	Relationship	moduleId	
relationships[x].released		-	
relationships[x].relationshipId	Relationship	id	id ja relationshipId dubleerivad
relationships[x].sourceId	Relationship	sourceId	concept
relationships[x].destinationId	Relationship	destinationId	concept
relationships[x].typeId	Relationship	typeId	concept
relationships[x].type		-	objekt, mis vastab typeId-le
relationships[x].target		-	objekt, mis vastab destinationId-le
relationships[x].groupId	Relationship	relationshipGroup	RF2 failis 0,1,2
relationships[x].modifier	Relationship	(modifierId)	JSONis sõnaline
relationships[x].characteristicType	Relationship	(characteristicTypeId)	JSONis sõnaline
relationships[x].id	Relationship	id	
[puudub]	Relationship	effectiveTime	koostamise kuupäev