

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond
Informaatika instituut
Teadmussüsteemide õppetool

***Data Documentation Initiative* standardi
kasutuselevõtt Statistikaametis**

Magistritöö

Üliõpilane: Kaia Kulla
Üliõpilaskood: 124494IABM
Juhendaja: Jaak Tepandi

Tallinn
2014

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.

(kuupäev)

(allkiri)

Annotatsioon

Käesolev magistritöö käsitleb statistiliste ja sotsioloogiliste andmete (sh metaandmete) kirjeldamise standardi "Andmete dokumenteerimise algatus" (ingl k *Data Documentation Initiative*; DDI) kasutusele võtmist Statistikaametis (SA) ning selleks vajalike muudatuste tegemist integreeritud metaandmete haldussüsteemis iMeta. DDI standardi kasutusele võtmine on vajalik, et tagada ühtne ja standardile vastav dokumenteeritud andmete kättesaadavaks tegemine (nt teadlastele) nii siseriiklikult kui ka rahvusvaheliselt.

Töö sissejuhatavas osas on kirjeldatud statistika tegemise äriprotsessi ja analüüsitud seda toetavat metaandmete haldussüsteemi iMeta. Seejärel on analüüsitud DDI standardit ja võrreldud selle ühilduvust statistika tegemise äriprotsessi GSBPM mudeliga ning statistiliste andmete ja metaandmete vahetuse standardiga SDMX. Põhjalikum analüüs on toodud DDI standardi ja metasüsteemi iMeta ühilduvuse hindamiseks. Lõpuks on toodud välja erinevused ja nendest lähtuvad ettepanekud DDI standardi integreerimiseks metaandmete haldussüsteemi iMeta.

Töö tulemusena on esitatud metaandmete haldussüsteemi iMeta täiendamiseks vajalik andmete koosseis ja arendusvajadused osas, mida käesolevas töös on käsitletud, ning ettepanekud edasisteks tegevusteks.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 80 leheküljel, 4 peatükki, 15 joonist, 11 tabelit ja 3 lisa.

Abstract

The present Master's thesis analyses the possibility to implement the Data Documentation Initiative (DDI) – a standard for describing statistical and sociological data (incl. metadata) – at Statistics Estonia and outlines the changes that would be required in the integrated metadata management system iMeta before the implementation of the DDI. Implementation of the DDI would ensure that users (e.g. researchers) in Estonia and abroad have access to data that are produced based on harmonised metadata that have been documented according to the standard.

The first part of the thesis describes the statistics production process and the metadata system iMeta, which supports production. This is followed by an analysis of the DDI standard and its compatibility with the Generic Statistical Business Process Model (GSBPM) and with the Statistical Data and Metadata Exchange (SDMX) standard. There is a more detailed analysis of the compatibility of the DDI and the metadata system iMeta. Finally, the main differences are listed together with recommendations for the integration of the DDI standard with the metadata management system.

The outcome of the analysis is an overview of the data required to improve the metadata system iMeta as well as a list of other improvements in the areas covered by the thesis and suggestions for future action.

The thesis is in Estonian and contains 80 pages of text, 4 chapters, 15 figures, 11 tables and 3 appendixes, etc.

Lühendite ja mõistete sõnastik

AS	<i>Analysis System</i> Analüüsisüsteem. Võimaldab teostada erinevaid statistilisi analüüse etteantud reeglite põhjal
ADF	<i>Application for designing forms</i> eSTATi jaoks aruande kirjeldamise rakendus, mis metaandmete haldussüsteemist iMeta küsib statistikatöö kirjeldusi, muutujaid ja klassifikaatoreid
DDI	<i>Data Documentation Initiative</i> Andmete dokumenteerimise algatus. Statistiliste ja sotsioloogiliste andmete (sh metaandmete) kirjeldamise standard [1]
eSTAT	<i>Statistical Reporting Environment</i> Statistiliste aruannete esitamise keskkond
Eurostat	<i>Statistical Office of the European Communities</i> Euroopa Liidu statistikaamet
GSBPM	<i>Generic Statistical Business Process Model</i> Üldine statistika äriprotsessi mudel [2]
GSIM	<i>Generic Statistical Information Model</i> Üldine statistiliste andmete mudel [3]
iMeta	<i>Integrated Metadata Management System</i> Integreeritud metaandmete haldussüsteem
KDM	<i>Survey Defining Module</i> Küsitlisdokumendi defineerimise moodul. KDM on tarkvara arendusplatvormi rakendus, mis on loodud lihtsustama küsimustike, objektimudelite, domeenide ja skriptide haldamist

Klassifikaator	<i>Classification</i> Mittepeidevate, ammendavate ja üksteist vastastikku välistavate kategooriate hulk, mida võib omistada ühele või mitmele mõõdetavale ja/või andmeesitusel kasutatavale muutujale
Metasüsteem	<i>see iMeta</i> vt iMeta
METIS	<i>Statistical Metadata Working Group</i> ÜRO Euroopa majanduskomisjoni statistiliste metaandmete töögrupp
MMX metaandmete repositooriumis	<i>MMX Metadata Repository</i> MMX metaandmete repositoorium on osa MMX metaandmete raamistikust, mudel-orienteeritud <i>Meta-Object Facility</i> 'le vastav metaandmete haldamise keskkond, mis on disainitud metaandmetel juhitud rakendustega (vt ka http://mmxframework.org)
Muutuja	<i>Variable</i> Vaadeldava üksuse tunnus, millel võib olla rohkem kui üks väärtuste hulk
Mõiste	<i>Concept</i> Teadmusüksus, mille moodustab ühene tunnuste kombinatsioon
Neuchâtel TM	<i>Neuchâtel Terminology Model for classifications (version 2.1) and variables (version 1.0)</i> Neuchâteli klassifikaatorite ja muutujate terminoloogia mudel [4]
Registrimuutuja	<i>Registry variable</i> Statistikatöös statistilisi üksikandmeid kirjeldav muutuja
SA	<i>Statistics Estonia</i> Statistikaamet
SDMX	<i>Statistical Data and Metadata eXchange</i> Statistiliste andmete ja metaandmete tõhusaks vahetuseks ja jagamiseks kasutatavate tehniliste standardite ja sisule suunatud juhiste

initsiatiiv ja standard koos IT arhitektuuri ja tööriistadega [5]

SRS

Registry of Statistical Entities System

Statistiliste registrite süsteem, mille koosseisu kuuluvad:
majandusüksuste statistiline register, põllumajandusüksuste register, eGeostat ning rahvastiku, eluruumide ja hoonete statistiline register

**Statistika
infosüsteem**

Statistical Information System

Statistika infosüsteem tegeleb statistiliste andmete kogumise, töötlemise ja levitamisega ning koosneb paljudest alamsüsteemidest

**Statistika tegemise
äriprotsess**

Generic Statistical Business Process

ÜRO Euroopa Majanduskomisjoni statistilise metaandmete töögrupis METIS välja töötatud üldise statistika äriprotsessi mudelist (GSBPM, (versioon 5, detsember 2013) tulenev protsess, milles on kirjeldatud kõik statistika tegemiseks vajalikud etapid ja alametapid, mis tagavad standardiseeritud statistika tegemise

Statistikatöö

Statistical activity

Statistilise info kogumine, säilitamine, teisendamine ja levitamine. Käesoleva magistritöö kontekstis käsitletakse statistikatööna ainult uuringuid, st töid, mille raames üksikobjektidelt (nt isikutelt, ettevõtetelt) kogutakse andmeid valik- või kõikse uuringuga (loendusega). Statistikatöö võib põhineda ka agregeeritud andmetel, mille baasil avaldatakse statistikat (nn raamatupidamise tüüpi tööd, nt sisemajanduse koguprodukti SKP koostamine erinevatel meetoditel).

Statistiline üksus

Statistical unit

Olem, mille kohta otsitakse informatsiooni ja koostatakse statistikat

Tunnus

Characteristic

Objekti või objektihulga mingi omaduse abstraktsioon. Tunnus iseloomustab statistilist üksust üldiselt; konkreetse statistikatöö kontekstis nimetatakse tunnust muutujaks

UNECE

United Nations Economic Commission for Europe

ÜRO Euroopa Majanduskomisjon, mis on loodud selleks, et soodustada majanduslikku koostööd selle liikmesriikide vahel

Uuring

Study unit, see Statistical activity

vt statistikatöö

VAIS

Data Processing System

Vaatluste andmetöötõluse infosüsteem. Kogutud andmete töötlemiseks mõeldud rakendus, mis küsib metaandmete haldussüsteemist iMeta muutujaid (tunnuseid), tehnilisi tunnuseid, klassifikaatoreid ja töötõlõsreegleid ärikeelelisel kujul

VVIS

Observations Information System

Isiku- ning rahva ja eluruumi uuringuteks mõeldud rakendus, mis küsib metaandmete haldussüsteemist iMeta kodeerimissõnastikke, abitekste, statistikatõõde kirjeldusi

Jooniste nimekiri

Joonis 1 GSBPM mudel [2].....	18
Joonis 2 Infosüsteemid ja rakendused statistika tegemise äriprotsessis	19
Joonis 3 Metasüsteemi iMeta üldine arhitektuur	20
Joonis 4 Metasüsteemi iMeta seos statistika infosüsteemi alamsüsteemidega.....	22
Joonis 5 DDI standardi üldine koosseisu ülevaade [10].....	25
Joonis 6 XML nimeruumid, DDI moodulid ja DDI skeemid [9]	26
Joonis 7 DDI elutsükli mudel [14]	31
Joonis 8 Kombinatsioon GSBPM, DDI ja SDMX mudelitest [16].....	33
Joonis 9 DDI mooduli <i>StudyUnit</i> koosseisu lihtsustatud vaade	35
Joonis 10 SA-s kasutatav Neuchâтели muutuja mudeli diagramm [14]	38
Joonis 11 DDI standardis muutuja mudel.....	39
Joonis 12 SA-s kasutatav Neuchâтели klassifikaatori mudeli diagramm [19].....	41
Joonis 13 DDI standardi koodiloendi mudeli diagramm.....	42
Joonis 14 Metasüsteemi iMeta illustreeriv joonis	45
Joonis 15 Mooduli <i>StudyUnit</i> elemendid metasüsteemi iMeta täiendamiseks.....	46

Tabelite nimekiri

Tabel 1 XML nimeruumide ja moodulite lühikirjeldus	27
Tabel 2 GSBPM ja DDI-L omavaheline võrdlus [16]	32
Tabel 3 Metasüsteemis iMeta vajalikud täiendused	47
Tabel 4 Komponent <i>Study Budget</i> elementide võrdlus metasüsteemiga iMeta.....	60
Tabel 5 Komponent <i>Study Unit</i> elementide võrdlus metasüsteemiga iMeta.....	61
Tabel 6 Mooduli <i>LogicalProduct</i> kaardistus metasüsteemiga iMeta	69
Tabel 7 Koodiloendi skeemi võrdlus metasüsteemiga iMeta	71
Tabel 8 Koodiloendi võrdlus metasüsteemiga iMeta	72
Tabel 9 Koodiloendi grupi võrdlus metasüsteemiga iMeta	73
Tabel 10 Kategooria võrdlus metasüsteemiga iMeta.....	75
Tabel 11 DDI standardi mõistete vasted metasüsteemi iMeta mõistetega.....	77

Sisukord

1. Sissejuhatus	13
1.1 Probleemipüstitus	13
1.2 Oodatavad tulemused	15
1.3 Ülesande püstituse lahendamise metoodika	15
1.4 Ülevaade tööst	16
2. Olemasoleva olukorra analüüs.....	17
2.1. Statistika tegemise äriprotsessi kirjeldus.....	17
2.2 Olemasolev infosüsteem.....	18
2.2.1 Metasüsteemi iMeta kirjeldus.....	19
2.2.2 Metasüsteemi iMeta kasutamise eelised ja puudused.....	22
3 Statistiliste andmete dokumenteerimine	24
3.1 DDI standardi raamistik ja struktuur	24
3.1.1 XML nimeruumid ja DDI moodulid.....	25
3.1.2 DDI skeemid.....	29
3.2 DDI standardi sobivus metasüsteemi standardite ja mudelitega	31
3.2.1 DDI elutsükli mudeli kooskõla GSBPM mudeliga	31
3.2.2 DDI standardi kooskõla SDMX standardiga	32
3.2.3 DDI standardi sobivus metasüsteemiga iMeta	34
3.2.3.1 Mooduli <i>StudyUnit</i> võrdlus metasüsteemiga iMeta	34
3.2.3.1.1 Komponenti <i>Study Budget</i> kaardistus metasüsteemiga iMeta.....	35
3.2.3.1.2 Komponenti <i>Study Unit</i> kaardistus metasüsteemiga iMeta	36
3.2.3.2 Mooduli <i>LogicalProduct</i> võrdlus metasüsteemiga iMeta.....	36
3.2.3.3 DDI standardi kooskõla Neuchâтели TM mudeliga	37
3.2.3.3.1 Neuchâтели muutuja mudeli võrdlus DDI standardiga	37

3.2.3.3.2 Neuchâтели klassifikaatori mudeli võrdlus DDI standardiga.....	40
4. Metasüsteemi iMeta täiendused.....	44
4.1 Metaandmete loetelu metasüsteemis iMeta täiendamiseks	44
4.1.1 Neuchâтели TM mudeli ja DDI standardi ühilduvus	49
4.2 Metasüsteemi iMeta vajalikud täiendused.....	49
4.3 Rakendamise kava	50
4.4 Tähelepanekud DDI standardi analüüsist	51
4.5 Standardi DDI versiooni 3.2 töövahendid	52
4.6 Piirangud DDI standardi kasutusele võtmiseks	52
5. Kokkuvõte	53
Summary.....	55
Kasutatud kirjandus	57
Lisa 1 Mooduli <i>StudyUnit</i> kaardistus metasüsteemiga iMeta	60
Lisa 2 DDI mooduli <i>LogicalProduct</i> kaardistus.....	69
Lisa 3 Sõnastik: DDI standardi mõisted, sh vasted metasüsteemi iMeta mõistetega.....	77

1. Sissejuhatus

Käesoleva magistritöö teema on *Data Documentation Initiative* standardi kasutuselevõtt Statistikaametis. *Data Documentation Initiative* e standard "Andmete dokumenteerimise algatus" (*edaspidi* DDI standard) on statistiliste ja sotsioloogiliste andmete (sh metaandmete) kirjeldamise standard.

SA-s on kasutusel metaandmete haldamiseks integreeritud metaandmete haldussüsteem iMeta (*edaspidi* metasüsteem iMeta), mille loomisel lähtuti eesmärgist, et see toetab kogu äriprotsessi – statistika infosüsteemi kõiki alamsüsteeme – ning võimaldab:

- luua metaandmete juhitavat infosüsteemi,
- metaandmete harmoneerimist,
- kirjeldada metaandmeid üks kord ühes kohas kõigiks vajadusteks, tagades ühesuguse metaandmete kasutamise, sõltumata andmete esitamise eesmärgist, vormingust, ajast ja meediast.

Agregeeritud andmete (statistika) ning metaandmete vahetamisel ja esitamisel on rahvusvahelised organisatsioonid, sh Euroopa statistikasüsteem ESS, võtnud kasutusele statistiliste andmete ja metaandmete tõhusaks vahetuseks ning jagamiseks kasutatava tehnilise standardi *Statistical Data and Metadata eXchange* (SDMX, ISO/IS 17369). [6]

Statistiliste üksikandmete dokumenteerimiseks seni ühtne standard puudub, kuid juba mõned aastad on Euroopa Liidu statistikaametis Eurostat ja statistika rahvusvahelistes töögruppides arutatud DDI standardi (alates versioonist 3.1, mis avaldati maikuu 2009) kasutamist. [7] Kavandatakse statistiliste üksikandmete teadlastele kättesaadavaks tegemist rahvusvaheliselt, mis eeldab ka andmete ühtset dokumenteerimist.

Lisaks DDI standardi kasutusele võtmisega seotud rahvusvahelise tegevusele, on SA arutanud Eesti Rahvusarhiiviga statistiliste üksikandmete arhiveerimist. Ka siis langes valik DDI standardile, kuid selle teemaga rohkem edasi ei ole tegeletud.

1.1 Probleemipüstitus

SA-s on loodud metasüsteem iMeta, mis on integreeritud statistika infosüsteemiga (st kõigi

alamsüsteemidega, vt Joonis 2). Metasüsteemis on kirjeldatud erinevaid statistikatöök vajalikke metaandmeid, mis juhivad statistika tegemist, aitavad ühtlustada mõisteid ja anda terviklikku ülevaadet kogutavate ja levitavate andmete kohta.

SA metasüsteem iMeta baseerub Neuchâтели klassifikaatorite ja muutujate terminoloogia mudelil (*edaspidi* Neuchâтели TM mudel), mis tagab statistikatööde ja nende andmestike ühtse dokumenteerimise. Statistika ja selgitavate metaandmete (*reference metadata*) edastamiseks rahvusvahelistele organisatsioonidele on kasutusel statistiliste andmete ja metaandmete vahetuseks mõeldud standard SDMX. Et SDMX on agregeeritud andmete (statistika) standard, siis statistiliste üksikandmete dokumenteerimiseks on valik langenud DDI standardile. [7]

Seega, et vastavalt Euroopa plaanile võimaldada tulevikus statistiliste üksikandmete rahvusvahelist kasutamist, on vaja selgeks teha, kuidas DDI standard on integreeritav olemasoleva metasüsteemiga iMeta.

Käesoleva töö eesmärgiks on selgitada välja DDI standardi kasutuselevõtuga seotud järgmised ülesanded/probleemid:

- teha olemasoleva metasüsteemi iMeta (sh ka Neuchâтели TM mudeli) ja DDI standardi võrdlev analüüs, et selgitada nende integreerimise võimalusi;
- selgitada välja võimalikud arendusvajadused metasüsteemis iMeta;
- milliseid tegevusi lisaks on vaja teha DDI standardi kasutusele võtmiseks.

Kuigi statistiliste üksikandmete ja neid täielikult kirjeldavate metaandmete arhiveerimine on laiem teema, tuleks ühtlasi jälgida, et kavandatavad metasüsteemi iMeta täiendused võimaldaksid edaspidi üksikandmete arhiveerimist Rahvusarhiivis DDI standardi XML vormingus.

SA eesmärk on viia olemasolev metasüsteem iMeta vastavusse DDI standardiga.

Käesolev töö on vajalik eelkõige SA-le, et:

- statistika tegemisel oleksid andmed (sh metaandmed) ühtsed ja üheselt arusaadavad nii kogumisel, töötlemisel, andmeanalüüsil, andmelaos säilitamisel, statistiliste üksikandmete levitamisel kui ka arhiveerimisel;

- oleks loodud eeldus statistikatööde kirjeldamiseks viisil, et kõik statistikatööga seotud asjaolud oleksid dokumenteeritud selliselt, et neid oleks võimalik kasutada sõltumata infotehnoloogilisest platvormist;
- statistilised üksikandmed saaksid arhiveeritud ning oleks võimalik edastada Rahvusarhiivi DDI standardi XML vormingus.

Lisaks eelnevale kaasneb ka statistiliste üksikandmete kasutajatele parem ja kvaliteetsem andmete esitus.

1.2 Oodatavad tulemused

Antud töö peab lahendama punktis 1.1 Probleemipüstitus toodud probleemid. Töö tulemusena on koostatud:

- metasüsteemis iMeta täiendavat kirjeldamist vajavate metaandmete loetelu;
- (võimalik) metaandmete repositooriumi ja metasüsteemi iMeta täiendamise vajaduse kirjeldus;
- (võimalike) metasüsteemi iMeta täienduste mahu hinnang (vajalik ressurss), kui täiendused osutuvad vajalikuks;
- juurutamise kava metasüsteemi iMeta täiendamiseks.

Metasüsteemi iMeta ja DDI standardi integreerimise tulemusena on võimalik dokumenteerida üksikandmeid vastavalt üldtunnustatud standardile, mis tagab vajadusel nende (anonüümitud) rahvusvahelise kasutatavuse, sh teadlaste poolt.

Kirjeldatud muudatuste järel peaks SA-l olema valmisolek väljundite genereerimiseks DDI standardi XML vormingus.

1.3 Ülesande püstituse lahendamise meetodika

Kuna teadaolevalt ei ole täna DDI standardit Eestis kasutusel, siis selle kasutusele võtmiseks SA-s on esmalt vaja saada ülevaade DDI standardist, selle olemusest ja kasutamise loogikast. Eelkõige on vaja aru saada erinevate metaobjektide (statistikatöö, muutuja, klassifikaator) omavahelistest seostest ja nende kasutamise põhimõtetest DDI standardis. DDI standardi mõistmiseks töötas töö autor läbi erinevaid dokumente, mida võis leida DDI standardi veebisaidi (<http://www.ddialliance.org/>) erinevatest sektsioonidest.

Järgnevalt on vaja leida erinevused DDI standardi ning metasüsteemis iMeta rakendatud

standardite ja mudelite vahel, võrreldes standardit eelkõige Neuchâтели TM mudeliga kui metasüsteemi iMeta peamise metaandmete kirjeldamise mudeliga.

Tuvastades erinevused, on vaja hinnata DDI standardi integreerimiseks metasüsteemi iMeta vajalikke täiendusi arendusteks ning tööde mahtu.

Võimalusel kavandada DDI standardi rakendamise plaan ja tegevused.

1.4 Ülevaade tööst

Magistritöö koosneb sissejuhatausest, neljast sisu hõlmavast peatükist, eesti- ja inglisekeelsest kokkuvõttest, kasutatud kirjanduse loetelust ning lisadest.

Esimeses peatükis on ülevaade teema püstitusest ja eesmärgid, mis tuleb magistritöö lõpuks saavutada.

Teises peatükis on ülevaade statistika tegemise põhiprotsessist ja selle seotusest SA infosüsteemidega. Täpsem ülevaade on antud käesoleva töö kontekstis peamisest infosüsteemist - metasüsteemist iMeta, selle seotusest teiste infosüsteemidega, metasüsteemis iMeta kasutatavatest mudelitest ja standarditest metaandmete kirjeldamisel.

Kolmandas osas on ülevaade DDI standardist, selle ülesehitusest ja standardi võrdlus SA-s ning metasüsteemis iMeta kasutatavate standardite ja mudelitega. Võrreldakse täpsemalt metasüsteemi iMeta ja DDI standardit, sh Neuchâтели TM mudelit, ning nendes kasutatavaid metaobjekte. Seejärel võrreldakse metasüsteemi iMeta mooduleid DDI standardi moodulitega ühisosa leidmiseks ja arendusvajaduste väljaselgitamiseks.

Neljandas osas tuuakse välja analüüsi tulemustest tehtavad järeldused, ettepanekud metasüsteemi iMeta andmekoosseisu täiendamiseks ja arendusteks. Lisaks on esitatud ajakava DDI standardi rakendamiseks ning ettepanekud edasisteks tegevusteks.

2. Olemasoleva olukorra analüüs

Aasta-aastalt paraneb SA-s statistikatööde tegemiseks vajalik infrastruktuur. Arendatud on infosüsteemi erinevaid komponente ja kasutusele on võetud erinevaid rahvusvaheliselt tunnustatud mudeleid ja standardeid, mis toetavad statistika tegemise protsessi ja aitavad kaasa kvaliteetse statistika tegemisele.

2.1. Statistika tegemise äriprotsessi kirjeldus

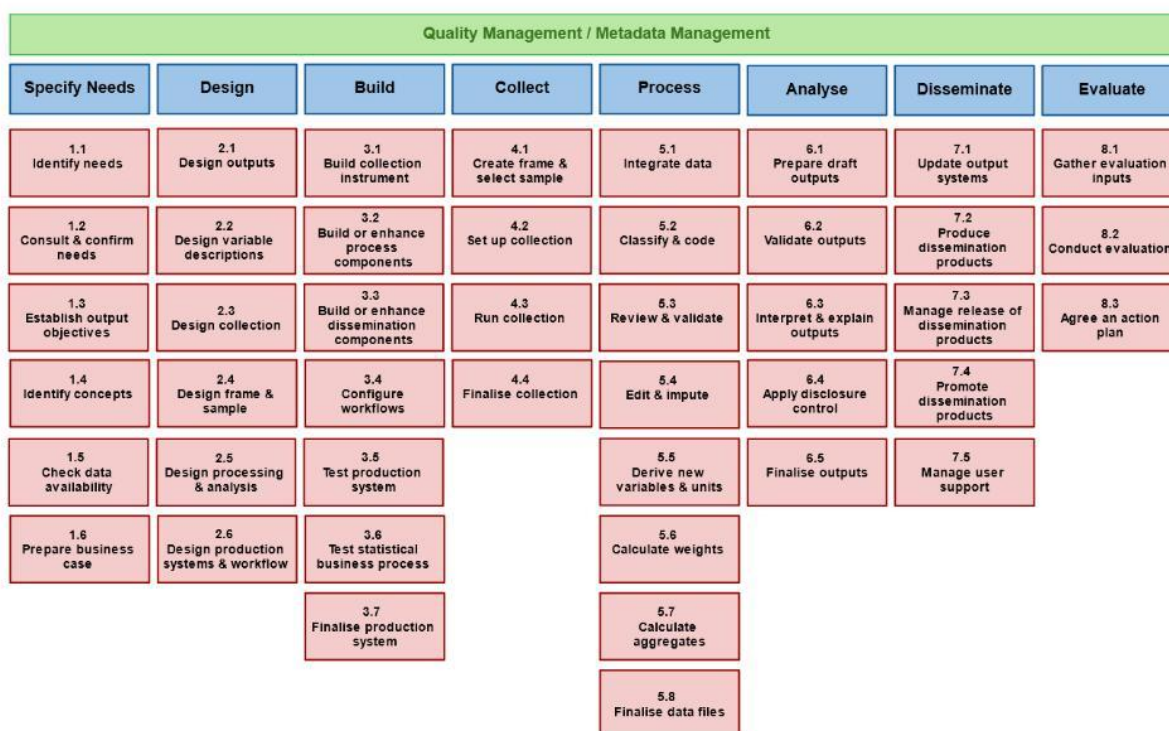
SA on statistika tegemisel rakendanud üldise statistika äriprotsessi mudeli [2] (ingl k *Generic Statistical Business Process Model*, GSBPM, esimene versioon on aastast 2008), mis on üldine statistika äriprotsessi kirjeldav mudel (vt Joonis 1).

GSBPM mudeli järgi jaguneb iga põhietapp omakorda alametappideks, kusjuures alametapid ei järgne mudelis kirjeldatud järjekorras, vaid see mudel on nn maatriks, kus alametappe võib läbida mitut erinevat teed pidi ja mõne regulaarse protsessi korral võib mõned alametapid vahele jätta. [2]

Alates 01.01.2014 on etappide arvu vähendatud küll kaheksani, kuid see muudatus ei puuduta statistika tegemise põhietappe. Varasema GSBPM mudeli versiooniga võrreldes eemaldati arhiveerimise etapp ning lisati see kõiki etappe hõlmavasse andmete ja metaandmete haldamise protsessi, mis väljendab arusaama, et andmeid võib arhiveerida igal statistika tegemise etapil.

GSBPM mudeli järgi hõlmab metaandmete haldamine kõiki põhiprotsessi etappe, alates vajaduste väljaselgitamisest ja täpsustamisest ning lõpetades hindamisega. Enamik metaandmeid kirjeldatakse ära põhiprotsessi kolme esimese etapi jooksul. See hõlmab nii andmete kogumise, analüüsi kui ka andmete levitamise seotud metaandmete kirjeldamist. Hiljem lisanduvad metaandmed uuringu tegeliku läbiviimise kohta ning statistika kvaliteedi kohta.

The GSBPM



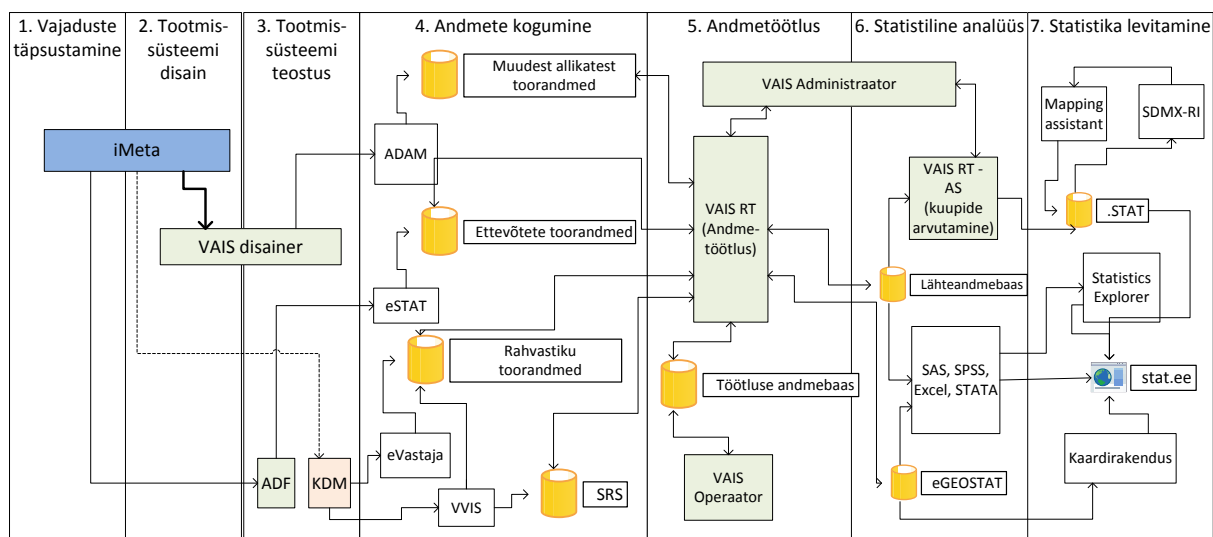
Joonis 1 GSBPM mudel [2]

GSBPM mudeli erinevates etappides ja alametappides kasutatakse erinevaid standardseid meetodeid ja tehnoloogiaid, mis vastavates etappides protsesse juhivad. GSBPM mudeli seos DDI standardiga on kirjeldatud punktides 3.2.1 DDI elutsükli mudeli kooskõla GSBPM mudeliga ja 3.2.2 DDI standardi kooskõla SDMX standardiga.

GSBPM mudeli kasutamise eesmärk on eelkõige muuta statistika tegemine efektiivsemaks.

2.2 Olemasolev infosüsteem

SA infosüsteem koosneb mitmest allsüsteemist ja rakendusest, mis on kasutusel statistika tegemise äriprotsessi erinevatel etappidel ja alametappidel (vt Joonis 2), milledest uuemad arendatud infosüsteemid (vaatluste andmetöötuse infosüsteem VAIS, andmete kogumise rakendus eSTAT, analüüsisüsteem AS) on loodud metaandmetel juhitud.



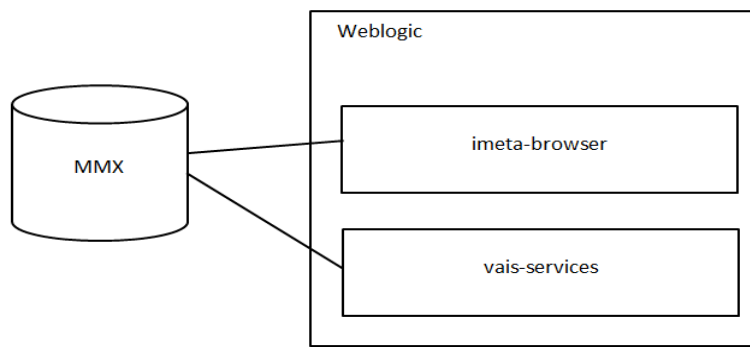
Joonis 2 Infosüsteemid ja rakendused statistika tegemise äriprotsessis

Käesolev töö keskendub eelkõige metasüsteemile iMeta, mis on keskkond, kus kirjeldatakse statistika tegemiseks vajalikud metaandmed ning need salvestatakse MMX metaandmete repositooriumis.

2.2.1 Metasüsteemi iMeta kirjeldus

Metaandmete kirjeldamiseks loodud metasüsteemis iMeta on realiseeritud mitmed metamudelid, mis kõik baseeruvad rahvusvahelistel standarditel või tunnustatud ekspertgruppide poolt väljatöötatud metamudelite spetsifikatsioonidel ja on kohandatud SA vajadustele. Metasüsteemi iMeta rakendus koosneb kolmest suuremast komponendist (vt Joonis 3):

- MMX repositoorium - andmebaas metaandmete hoidmiseks;
- metasüsteemi iMeta veebipõhine kasutajaliides;
- XML servlett, mis on osa VAIS-teenuste veebirakendusest.



Joonis 3 Metasüsteemi iMeta üldine arhitektuur

MMX metaandmete repositooriumis on realiseeritud järgmised metamudelid:

- Neuchâтели terminoloogia mudel; [4]
- ISO/IEC 11179, *Information Technology - Metadata registries (MDR)*; [8]
- EVS-ISO 1087 Terminoloogiatöö sõnastik;
- relatsioonilise andmebaasi mudel (*Relational Database (RDB) Metamodel*), mille abil kirjeldatakse:
- andmebaaside struktuurid;
- välisallikatest hõivatavate andmefailide struktuurid;
- VVISi ja eSTATi küsimustike struktuurid;
- andmetöötlusvoogude mudel (VAIS ETL/XDTL), milles kirjeldatakse VAISi andmetöötlusvood (XDTL *metamodel*);
- WSDL metamudel veebiteenuste kirjeldamiseks;
- rollipõhise juurdepääsu õiguste mudel (RBAC - *Role Based Access Control*), milles kirjeldatakse kasutajaõigused;
- *Data Validation Metamodel*;
- *MMX Core Metamodel*.

Metasüsteemi iMeta rakendus võimaldab ühelt poolt vaadata kõiki repositooriumisse salvestatud metaandmeid ning teisalt lubab hallata statistika tegemise protsessi keskseid metaandmeid, mis metasüsteemi iMeta rakenduses on jaotatud moodulitesse, nagu:

- statistikavaldkonnad;
- statistikatööd ja versioonid, muutujad, kuubid e väljundtabelid;
- mõisted;

- klassifikaatorid, nende versioonid, variandid ja vastavustabelid, sh koodiloendid ja agregeerimis- ja kodeerimistabelid;
- statistiliste üksuste tüübid ja neid kirjeldavad statistilised tunnused;
- mõõtühikute tüübid ja mõõtühikud;
- statistikaga seotud õigusaktid;
- andmebaasi struktuurid (tehnilised tunnused).

Käesoleva töö seisukohalt on oluline Neuchâтели TM mudel, sest metasüsteemi iMeta loomisel lähtuti Neuchâтели TM mudeli loogikast.

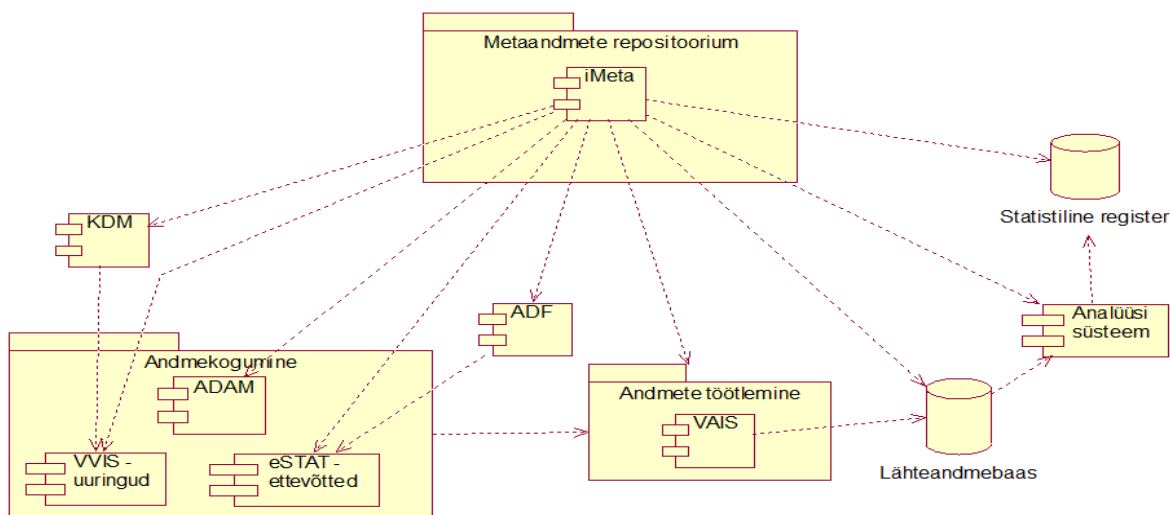
Muutujate/tunnuste ja mõistete süsteemi loomisel on aluseks võetud Neuchâтели muutujate mudel, mida on kohandatud SA vajadustele. Klassifikaatorite andmebaasi loomisel on kasutatud Neuchâтели klassifikaatorite mudelit. Täpsem ülevaade Neuchâтели TM mudelist on toodud punktides 3.2.3.3.1 Neuchâтели muutuja mudeli võrdlus DDI standardiga ja 3.2.3.3.2 Neuchâтели klassifikaatori mudeli võrdlus DDI standardiga.

Metasüsteemiga iMeta (vt Joonis 4) sama MMX metaandmete repositooriumi küljes asuvad ka VAIS rakendused ja automaatne andmehõive infosüsteem ADAM. Ülejäänud infosüsteemid kasutavad metasüsteemis kirjeldatud ja MMX metaandmete repositooriumisse salvestatud metaandmeid.

Analüüsisüsteem AS ja automaatne andmehõive infosüsteem ADAM ei ole vaatamata oma nimele tegelikult eraldi infosüsteemid. Osa funktsionaalsusi on realiseeritud metasüsteemis iMeta ja teine osa VAIS rakendustes.

Statistika tegemise äriprotsessi teise etapi tööde (tootmissüsteemi disainimine) teostamisel on kasutusel küsimustike ja ankeetide kirjeldamise rakendused – küsitlusdokumendi defineerimise moodul KDM uuringute küsimustike disainimiseks ja eSTATi jaoks aruannete kirjeldamise rakendus ADF ettevõtete aruannete disainimiseks.

Kõik mainitud rakendused kasutavad metasüsteemis iMeta ette defineeritud muutujaid ja klassifikaatoreid.



Joonis 4 Metasüsteemi iMeta seos statistika infosüsteemi alamsüsteemidega

SA on metasüsteemi iMeta kasutamisel võtnud eesmärgiks metaandmete standardimise.

2.2.2 Metasüsteemi iMeta kasutamise eelised ja puudused

Metasüsteem iMeta on olnud kasutuses ja pidevas arenduses juba neli aastat. Suurimaks eeliseks on see, et metasüsteemi iMeta abil on võimalik ühtlustada ja standardida metaandmeid ning juba kirjeldatud metaandmeid korduvalt kasutada.

Kõik metasüsteemis iMeta kirjeldatavad tööd koondatakse ühise nimetuse "Statistikatöö" alla. Statistikatöö on üldine raamistik. Statistikatööd ja nende versioonid grupeeritakse statistikavaldkondade järgi.

Statistikatööd ja selle versiooni kirjeldavad mitmesugused tunnused, mida metasüsteemis statistilistest tunnustest eristamiseks nimetatakse atribuutideks. Statistikatööd kirjeldavad atribuudid on hallatavad. Sõltuvalt statistikatöö liigist, kasutatakse statistikatöö kirjeldamiseks ühtesid või teisi atribuute, mis võivad versiooniti olla erinevad.

Statistilist üksust iseloomustavad mitmesugused (statistilised) tunnused. Tunnus iseloomustab statistilist üksust kui liiki üldiselt. Tunnust konkreetse statistikatöö kontekstis nimetatakse muutujaks. Statistikatööle kirjeldatakse selles kasutatavad muutujad.

Statistilise uuringu (statistikatöö liik) väljundiks on statistilised näitajad (statistika). Statistiline näitaja on kirjeldatav koguselise (mõõdetava) muutujaga, mis määrab näitaja sisu, ja klassifitseeriva muutujaga, mille lõikes seda koguselist muutujat esitatakse.

Muutujaid liigitatakse registrimuutujateks (statistilisi üksikandmeid kirjeldavateks

muutujateks) ja kuubimuutujateks (avaldatavat statistikat kirjeldavateks muutujateks) Üldistavalt võib nii registri- kui ka kuubimuutujad jagada mõõdetavaks ja klassifitseerivaks. Et klassifitseeriv muutuja omandab väärtusi mingist etteantud loetelust, siis seotakse klassifitseeriv registri- või kuubimuutuja metasüsteemis hallatava klassifikaatori või muu loendiga.

Kuigi SA metasüsteem iMeta võimaldab üksikasjalikult kirjeldada statistikatöö üksikandmestikku (nii kogutavaid kui ka arvutatavaid registrimuutujaid), avaldatavat statistikat (kuubimuutujaid ja nende baasil avaldatavaid tabeleid), statistika koostamisel kasutatud meetodikat, kvaliteedinäitajaid jpm, on vaja süsteemi täiendada nii, et statistiliste üksikandmete osas saaks sellest moodustada DDI standardile vastavaid väljavõtteid – nii nt teadlastele üksikandmete kasutamiseks, andmete arhiveerimiseks kui ka muudeks vajadusteks.

3 Statistiliste andmete dokumenteerimine

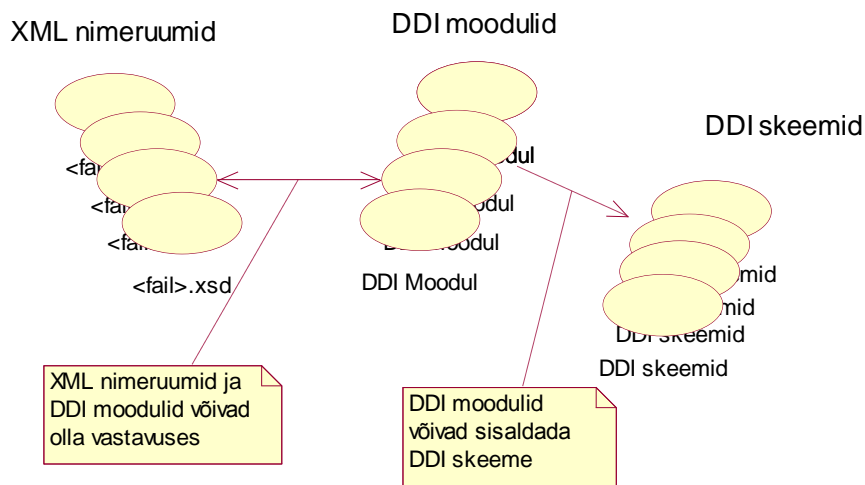
DDI standard on mõeldud statistiliste andmete kavandamiseks ja see võimaldab kirjeldada statistilisi andmeid (sh metaandmeid) kogu nende elutsükli jooksul: kogumiseks, töötlemiseks, analüüsimiseks, levitamiseks, haldamiseks ja arhiveerimiseks ning selle alusel saadud andmestruktuuride kirjeldamiseks sotsiaal-, käitumis-, majandus- ja nendega seotud teadusharudes. [9]

Kuna tegemist on standardiga, mida Eestis töö autorile teadaolevalt käesoleval ajal ei kasutata, siis uuris töö autor kõigepealt põhjalikumalt DDI standardit, selle ülesehitust ja selles kasutatud metaobjekte ning nende vahelisi seoseid, et saada standardist ülevaade ning seejärel võrrelda seda SA metasüsteemiga iMeta ning sellest kasutatava Neuchâтели TM mudeliga.

3.1 DDI standardi raamistik ja struktuur

DDI standardist ülevaate saamiseks on järgnevalt kirjeldatud selle üldist ülesehitust (struktuuri), mis on järgmine (vt Joonis 5): [1] [10]

- XML nimeruumid (*XML Schemas*) - *.xsd* lõpuga failid, mis kirjeldavad XML faili struktuuri, mis seda nimeruum kasutab;
- DDI moodulid (*DDI Modules*) - paketid, mis vastavad elutsükli etapile või on spetsiifilise struktuurilise funktsiooniga;
- DDI skeemid (*DDI Schemes*) - hallatavad struktuurid, mida võib leida mitmest DDI moodulist;
- DDI profiilid (*DDI Profiles*) - DDI elementide ja atribuutide kirjeldus, mida konkreetses profiilis kasutatakse.



Joonis 5 DDI standardi üldine koosseisu ülevaade [10]

DDI standardi peamine funktsionaalsus on, et erinevaid mooduleid ja skeeme saab eraldi hallata ning:

- objektid on identifitseeritavad, versioneeritavad (nt muutujad, kategooriad, üldkogumid, mõisted, organisatsioon, suunamised ja kontrollid jne) ja hallatavad (*maintainables*);
- objektidele saab viidata, sel juhul on kohustuslik märkida haldava asutuse nimi, identifikaator ja versioon;
- saab lisada muid materjale. [11]

Lisaks võib DDI standardi moodulitest moodustada pakette / struktuurilisi mooduleid (nt *DDI Instance* koosneb moodulitest *StudyUnit*, *Group*, *ResourcePackage* ja *LocalHoldingPackage*). [12]

3.1.1 XML nimeruumid ja DDI moodulid

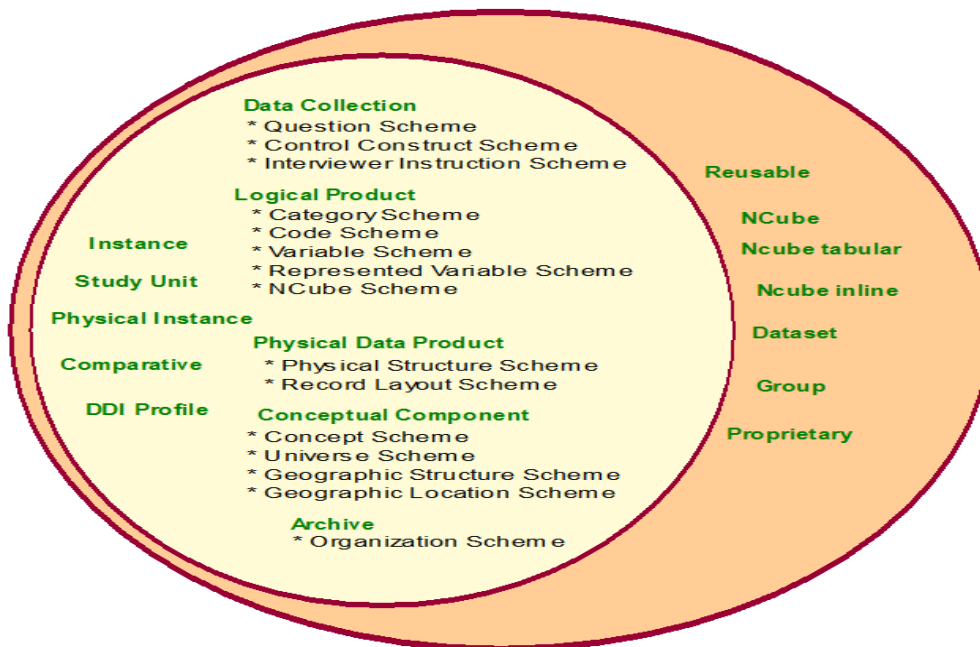
XML nimeruume e XML skeeme on standardis 22, milles on kokku 1154 elementi. XML nimeruume, DDI mooduleid ja DDI skeeme on võimalik iseseisvalt üksteisest sõltumatult hallata ja neid saab omavahel kombineerida vastavalt vajadusele. [1] [10]

DDI moodulid võib jagada põhimõtte järgi järgnevalt:

- skeemipõhised moodulid (sisaldavad hallatavaid skeeme, vt punkt 3.1.2 DDI skeemid);

- mitte-skeemipõhised moodulid (nt *physicalinstance* -- üks-ühene seos andmefailiga; viide andmefaili enda kohta, *Comparative*, *DDIprofile*);
- alammodulid (kasutatakse üksnes teistes moodulites);
- välised XML-skeemid;
- korduvalt kasutatavad (tavaliselt vajaminevad) osad.

DDI moodulid on vastavuses andmete elutsükliga. Mõned DDI moodulid sisaldavad DDI skeeme. Joonis 6 on kujutatud visuaalselt XML nimeruumide ja DDI moodulite ning skeemide omavahelist seost. Rohelise kirjaga on tähistatud XML nimeruumid (*XML Schemas*), heledal taustal roheline kirjaga on samas ka DDI moodulid ning musta kirjaga on tähistatud DDI skeemid.



Joonis 6 XML nimeruumid, DDI moodulid ja DDI skeemid [10]

Tabel 1 on XML nimeruumide, sh DDI moodulite lühiülevaade, millist informatsiooni nad sisaldavad. Sellest lähtuvalt saab hinnata, millise metasüsteemi iMeta mooduliga on neid eelkõige vaja võrrelda.

Tabel 1 XML nimeruumide ja moodulite lühikirjeldus

	XML nimeruum (XML Schemas)	DDI moodul	Nimeruumi ja mooduli lühikirjeldus
1	Arhiiv (<i>Archive</i>)	Arhiiv (<i>Archive</i>)	Sisaldab teavet, mis on seotud andmete ja metaandmete arhiveerimisega (pikemas perspektiivis juurdepääsu ja/või säilitamisega). DDI Arhiiv viitab protsessidele, mitte asukohale.
2	Võrdlusinfo (<i>Comparative</i>)	Võrdlusinfo (<i>Comparative</i>)	Sisaldab sama või sarnast tüüpi metaandmeobjektide (muutujad, küsimused, kategooriad, statistilised üksused jms) võrdlusinformatsiooni, kasutades lähte- ja sihtobjektide vahelist võrdlemist.
3	Kontseptuaalne komponent (<i>Conceptual Component</i>)	Kontseptuaalne komponent (<i>ConceptualCompon ent</i>)	Sisaldab uuringu või uuringute grupi kontseptuaalseid komponente.
4	Andmekogumine (<i>Data Collection</i>)	Andmekogumine (<i>DataCollection</i>)	Sisaldab infot andmepaketi loomiseks andmekogumise ja selleks vajaliku protsessi kirjelduse kohta -- meetodika, tulemused, andmeallikad, kogumisinstrumentid ja protsessid.
5	Andmekogum (<i>Dataset</i>)	-	Andmekogum on alusdokumendi struktuuri (<i>BaseRecordLayout</i>) asendus ja võimaldab metaandmete faili masinhõivamist üksik- või agregeeritud andmete tasandil.
6	DDI profiil (<i>DDI Profile</i>)	DDI profiil (<i>DDIProfile</i>)	Kirjeldab kehtivate DDI objektide alamhulkasid, mida kasutatakse asutuses sihtotstarbeliselt.
7	Grupp (<i>Group</i>)	-	Peamine pakendamise ja avaldamise moodul DDI-s, mis sisaldab mitmeid <i>StudyUnits</i> gruppe.
8	Versioon (<i>Instance</i>)	Versioon (<i>Instance</i>)	<i>DDI Instance</i> on nagu ümbrik (konteiner), mis ümbritseb iga DDI versiooni. Võib sisaldada skeeme - <i>ResourcePackage</i> , <i>LocalHoldingPackage</i> või <i>StudyUnit</i> .

	XML nimeruum (XML Schemas)	DDI moodul	Nimeruumi ja mooduli lühikirjeldus
9	Lähtebaas (Logical Product)	Lähtebaas (LogicalProduct)	Sisaldab muutujate, klassifikaatorite ja kuupide kirjeldust.
10	Füüsilise andmefaili kirjeldus (Physical Data Product)	Füüsiline andmefaili kirjeldus (PhysicalDataProduct)	Moodulis kirjeldatakse andmefailide struktuuri füüsiline säilitamine ja nende sisemiste objektide seosed objektide loogilise (intellektuaalse) kirjeldusega, mis sisaldub moodulis <i>LogicalProduct</i> .
11	Sisene andmekuup (Inline Ncube)	-	<i>BaseRecordLayout</i> – lähtebaasi väljavõte, mis on mõeldud kasutamiseks, kui <i>Ncube</i> on varustatud metaandmetega.
12	Andmekuup (Ncube)	-	<i>BaseRecordLayout</i> lähtebaasi väljavõte, mis on mõeldud kasutamiseks koos <i>Ncube</i> arhiveerimise formaadiga või <i>Ncube</i> koos algandmetega (st <i>NCube</i> komplekt, kus juhtumituvastamist ei kirjeldata <i>Ncube</i> dimensioonina), mis hoitakse välises failis fikseeritud või piiritletud kohtades andmete osas korraldatud nii ühedimensiooniliste ridadena.
13	Tabelikujuline lähtebaasi kirjeldus (Tabular Ncube)	-	<i>BaseRecordLayout</i> lähtebaasi väljavõte, mis on mõeldud kasutamiseks koos tabelitega ja <i>Ncube</i> objekte hoitakse välises failis, kus asukoha andmed on kindlaks määratud kaherealise dimensioonina (rida ja veerg).
14	Omandiõigus (Proprietary)	-	<i>BaseRecordLayout</i> - lähtebaasi väljavõte, mis on mõeldud kasutamiseks, kui andmete kirjed salvestatakse välistesse andmeformaatidesse. Lisatakse viide omandiõiguse kohta.
15	Füüsiline andmefail (Physical Instance)	Füüsiline andmefail (PhysicalInstance)	Sisaldab teavet asjaomase andmetoote kohta (tegelik andmefail).

	XML nimeruum (XML Schemas)	DDI moodul	Nimeruumi ja mooduli lühikirjeldus
16	Korduvalt kasutatav (Reusable)	-	Korduvalt kasutatavad objektid nagu muutujad, koodid jne.
17	Uuring (Study Unit)	Uuring (StudyUnit)	Sisaldab uuringuga seotud metaandmete kirjeldusi.

Lisaks on viis XML nimeruumi, mis on statistikatöö kontekstis nn välised XML nimeruumid, nagu <http://purl.org/dc/dcmitype/>; <http://purl.org/dc/elements/1.1/>, <http://purl.org/dc/terms/>, <http://www.w3.org/1999/xhtml> ja <http://www.w3.org/XML/1998/namespace>, kuid on standardi koosseisus vajalikud. PURL (*Persistent Uniform Resource Locators*) toetab püsivate veebiaadresside kasutamist XML nimeruumis, uuendades veebiaadressi, kui seda peaks muudetama. W3 sätestab reeglid XML skeemis kasutatavatele atribuutidele.

3.1.2 DDI skeemid

DDI moodulites sisalduvad DDI skeemid ja nende lühikirjeldused on järgmised: [1] [10]

Andmekogumine (*DataCollection*)

- Küsimuste skeem (*Question Scheme*) - sisaldab küsimuste ja vastuste valdkondi.
- Küsimuste kontrolliskeem (*Control Construct Scheme*) - küsimustevoo kontroll sõltuvalt vastustest (järjestus, *IfThenElse*, *Loop* jne).
- Küsitlaja instruksioonid (*Interviewer Instruction Scheme*) - juhend küsitluse läbiviimiseks (küsitlus võib olla nii nn näost-näku kui ka veebipõhine) kuidas küsimusi ja vastuseid interpreteerida (selgitused).

Lähtebaas (*LogicalProduct*)

- Kategooria skeem (*Category Scheme*) - kategooriad, mis kindlustavad nummerdatud esindatuse mõistete jaoks ja kasutatakse küsimuste, koodiloendite ja muutujate juures.
- Koodiloendi skeem (*Code List Scheme*) – skeem, mis sisaldab koodiloendeid, mis seovad konkreetse väärtuse kategooriaga ning kasutatakse küsimuste ja muutujate juures.

- Muutuja skeem (*Variable Scheme*) - muutuja skeem sisaldab väärtuste kogumit, mis on kirjeldatud elemendi *ManagedRepresentation* poolt. Neid kasutatakse viidetena, et määratleda muutuja esindatus (*Variable Representation*) ja küsimuste vastuste domeen (*Question Response Domain*).
- Uuringus kasutatud muutuja skeem (*Represented Variable Scheme*) - muutujate loetelu, nende struktuur, esindatus, lähteinformatsioon ja väljendus.
- Kuubi skeem (*NCube Scheme*) - kuupides on andmed dimensioonides, kus on seos individuaalsete andmelahtrite vahel (st statistilised tabelid, korrelatsioonitabelid jms).

Füüsilise andmefaili kirjeldus (*PhysicalDataProduct*)

- Andmete füüsiline struktuur (*Physical Structure Scheme*) - kirjeldatakse andmete üldine füüsiline struktuur (sh salvestusformaadid, andmestiku osad, vaikimisi väärtused ja andmetüübid).
- Andmebaasi struktuur (tehniline tunnus) (*Record Layout Scheme*) - kirjete paigutus näeb ette spetsiifilise lingi muutuja kirjelduse ja kuubi lahtri vahel füüsilise salvestamise asukohaga andmefailis.

Arhiiv (*Archive*)

- Organisatsiooni skeem (*Organization Scheme*) - üksikisikute ja organisatsioonide kirjeldused ja nendevahelised seosed.

Kontseptuaalne komponent (*Conceptual Component*)

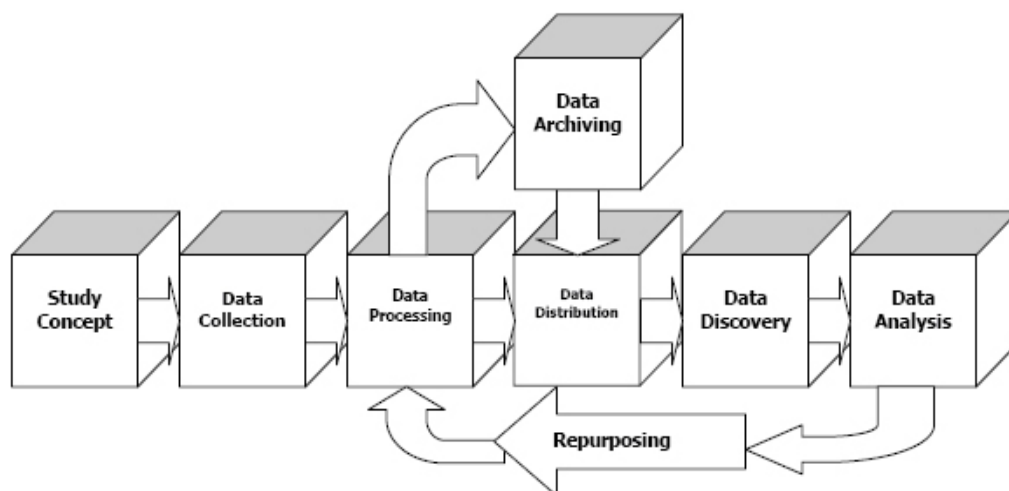
- Mõistete skeem (*Concept Scheme*) - kirjeldab mõisteid ja mõistete gruppe.
- Üldkogumi skeem (*Universe Scheme*) - sätestab üldkogumi hierarhilise struktuuri leitud andmetes.
- Geograafilise struktuuri skeem (*Geographic Structure Scheme*) - määratleb geograafilise hierarhia struktuuri, et kirjeldada geograafilisi piirkondi - riigid, linnad, maa-alad, regioonid.
- Geograafilise asukoha skeem (*Geographic Location Scheme*) - asukohad on spetsiifilised struktuuri tüübi, nime, kodeerimise ja füüsilise asukoha definitsiooni järgi.

3.2 DDI standardi sobivus metasüsteemi standardite ja mudelitega

DDI standardi kasutusele võtmise üheks heaks eelduseks on, et DDI standardi koostajad on väga selgelt deklareerinud, et DDI standard on kooskõlas standardiga ISO/IEC 11179 (ISO/IEC 11179 *Information technology — Metadata registries*, MDR). [13]

Kuna ka metasüsteemis iMeta rakendatud Neuchâтели TM mudeli loomisel, mis on loodud spetsiaalselt statistikasüsteemi jaoks, järgiti selle maksimaalselt võimalikku ühilduvust üldise metaandmete standardiga ISO/IEC 11179 [14], siis on see heaks eeliseks DDI standardi kasutusele võtmisel.

Antud töö seisukohalt on oluline vaadelda nii DDI elutsükli (*DDI Lifecycle*, DDI-L) mudelit kui ka DDI standardit. DDI-L mudel (vt Joonis 7) kirjeldab protsessi raamistikku ja standard andmete koosseisu erinevates protsessi etappides.



Joonis 7 DDI elutsükli mudel [15]

DDI-L mudel toetab tugevalt metaandmete korduvat kasutamist kogu andmete elutsükli jooksul ning samuti toetab see mitmekeelsust – iga dokumendi võib valmis teha vajalikus keeles. [15]

3.2.1 DDI elutsükli mudeli kooskõla GSBPM mudeliga

DDI-L mudel (vt Joonis 7) on üks mudelitest, mis on võetud GSBPM mudeli loomisel aluseks ja see peegeldub GSBPM mudeli struktuuri ülemises kihis (Gregory, 2011). [16] Tabel 2 kajastab kahe mudeli struktuuride võrdlust.

Tabel 2 GSBPM ja DDI-L omavaheline võrdlus [17]

Generic Statistical Business Process Model	DDI 3.2 Combined Life Cycle Model
1 Vajaduste täpsustamine (<i>Specify needs</i>)	Study Concept
2 Tootmissüsteemi disain (<i>Design</i>)	Repurposing (part)
3 Tootmissüsteemi teostus (<i>Build</i>)	
4 Andmekogumine (<i>Collect</i>)	Data Collection
5 Andmetöötlus (<i>Process</i>)	Data Processing (mostly) Repurposing (part)
6 Statistiline analüüs (<i>Analyse</i>)	Data Discovery Data Analysis Data Processing (part)
7 Statistika levitamine (<i>Disseminate</i>)	Data Distribution
8 Arhiveerimine (<i>Archive</i>)	Data Archiving
9 Hindamine (<i>Evaluate</i>)	
Quality Management	
Metadata Management	

DDI-L mudeli ja GSBPM mudeli omavahelist võrdlust on tehtud UNECE poolt [17] ja selle analüüsi tulemusena on tõdetud, et DDI ja GSBPM vahel on erinevusi, mis on peamiselt seotud nende mudelite loomise eesmärgiga, kuid vaatamata eesmärgi erinevustele, on ka palju sarnasusi.

3.2.2 DDI standardi kooskõla SDMX standardiga

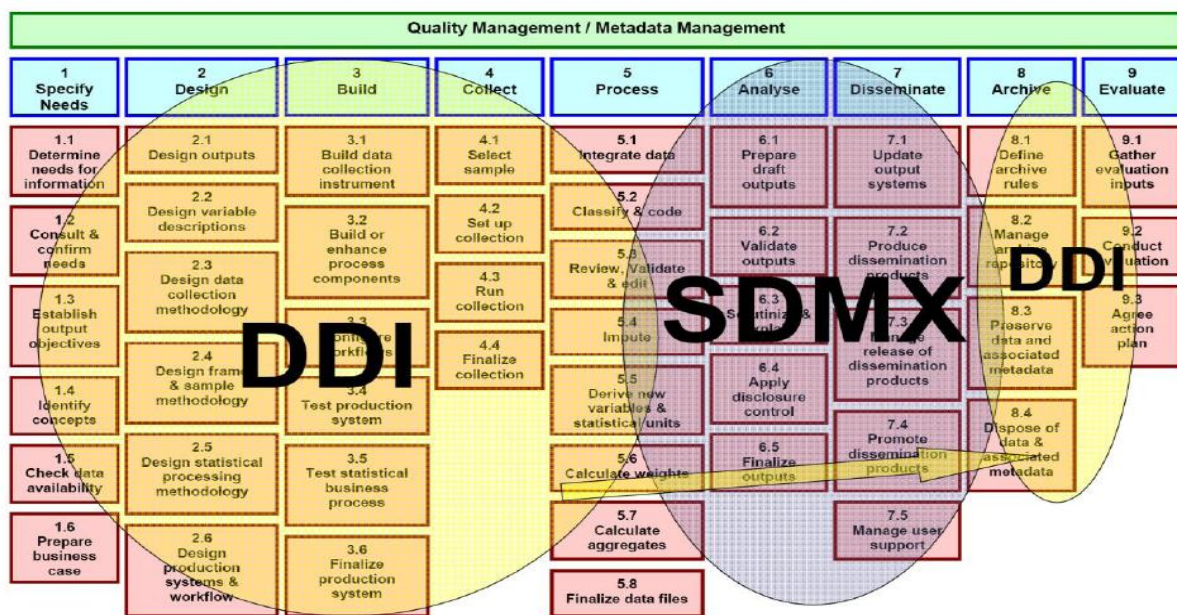
Statistiliste andmete ja metaandmete vahetuse standard SDMX (*Statistical Data and Metadata eXchange*) algatati seitsme rahvusvahelise organisatsiooni poolt [5], et soodustada standardite kasutamist statistilise teabe vahetamisel. Standardi SDMX raskuspunkt on agregeeritud andmetel (*macro data*), kuigi see standard toetab ka üksikandmete (*micro data*) dokumenteerimist. SDMX standard on heaks kiidetud statistikaametite ja Eurostati vahel andmete vahetamiseks ja jagamiseks, mille parimaks tõestuseks on rahva ja eluruumide loenduse tulemuste jagamine Eurostati kodulehel *Census Hub* vahendusel. [18] SDMX algatus seab tehnilised standardid ja sisule orienteeritud suunised, et lihtsustada statistilise andmete, eelkõige agregeeritud andmete, ja metaandmete vahetust kaasaegse infotehnoloogia abil. [5]

SDMX standard kirjeldab agregeeritud andmeid ja pakub selleks standardseid struktuure, nt andmete edastamise mudelit.

Samuti pakub SDMX standard standardiseeritud koodiloendeid, mille alusel andmeid vahetatakse.

DDI ja SDMX standardite sarnasused: [17]

- mõlemad standardid kasutavad sarnast mehhanismi URN tunnuste (identifikaatorite) struktureerimisel;
- mõlemad standardid kasutavad sarnast mudelit ja reeglistikku identifitseeritavate, versioneeritavate ja hallatavate objektide osas;
- mõlemad standardid kasutavad samasuguste kirjete loetlemiseks skeemi (*Schemes*) pakette;
- mõlemad standardid on kavandatud selliselt, et toetada taaskasutamist ning mõlemal on sarnased etalonmudelid.



Joonis 8 Kombinatsioon GSBPM, DDI ja SDMX mudelist [17]

Kui DDI standard pakub rikkalikku üksikandmete kirjeldamise mudelit, siis SDMX standard pakub integreeritud andmevahetusplatvormi statistika väljundi jaoks.

Joonis 8 Kombinatsioon GSBPM, DDI ja SDMX mudelist on kujutatud standardite kasutamist GSBPM mudeli etappide lõigetes - millistel etappidel on ühe või teise standardi kasutamisel eelised.

Kuna mõlemad standardid kirjeldavad andmete kogumeid ja nende struktuure, on neil standarditel ühised metaandmete komponendid nagu mõisted, koodiloendid, dimensioonid ja atribuudid, mõõdustik (mõõtühikute kasutamine) ja agregeeritud andmekuupide struktuur. [15]

3.2.3 DDI standardi sobivus metasüsteemiga iMeta

DDI standardi kasutusele võtmiseks on eelnevalt vaja välja selgitada, mil määral on standard ühilduv juba olemasoleva metasüsteemiga iMeta.

Kuna DDI standard on üles ehitatud XML nimeruumide (XML skeemide), sh moodulite järgi, siis on mõistlik võrrelda DDI standardit metasüsteemiga iMeta just nende samade nimeruumide ja/või moodulite kaupa. Selline lähenemine on mõistlik juba seetõttu, et ka metasüsteemis on kasutusel moodulid ja moodulites on atribuudid teemade järgi (vt punkti 2.2.1 Metasüsteemi iMeta kirjeldus). Paljuski on võimalik intuiitselt moodulite nimetustest tuletada, millega tegemist on, kuid täpsema ülevaate saab siiski mooduli põhjalikul uurimisel.

Tulenevalt punktis 3.1 DDI standardi raamistik ja struktuur toodud lühikirjeldustest, samuti DDI *Alliance* kodulehel [11] toodud soovitudest, on uuringu läbiviimise seisukohalt esimeses järjekorras olulised järgmised DDI moodulid: *StudyUnit*, *LogicalProduct*, *Conceptual Component* ja *DataCollection* ning lisaks veel DDI moodulid *PhysicalDataProduct* ja *PhysicalInstance*.

Punktis 3.2.3.1 Mooduli *StudyUnit* võrdlus metasüsteemiga iMeta on esitatud metasüsteemi iMeta ja DDI standardi võrdlus, alustades esmalt uuringu moodulist *StudyUnit*. Metasüsteemis iMeta on nime järgi kõige sarnasem moodul "Statistikatöö".

DDI moodulites (ja standardis tervikuna) on osa elemente, mis on kohustuslikud kasutada, kuid enamikel juhtudel ei ole elementide kasutamine kohustuslik ja see võimaldab organisatsioonil endal otsustada, milliseid elemente kasutada, ehk millist infot dokumenteerida ja millist mitte.

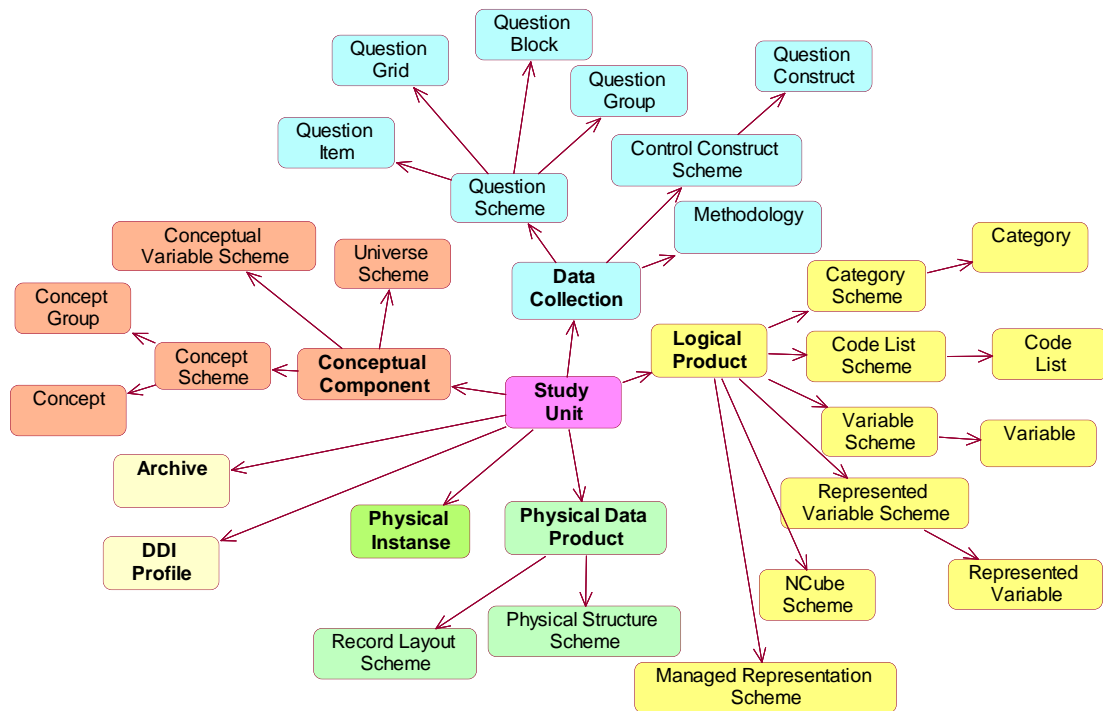
Võrdlus peab andma ülevaate, kas ja kui palju on vaja metasüsteemi iMeta täiendada.

3.2.3.1 Mooduli *StudyUnit* võrdlus metasüsteemiga iMeta

Moodul *StudyUnit* e uuringu moodul on DDI standardi peamine moodul. [11] DDI standardis defineeritakse uuringut kui ühekordset andmekogumi kogumise/hõivamise tegevust, näiteks ühekordne uuring või mitmeaastase korduva uuringu ühekordne iteratsiooni (näiteks üks aasta

pikisuunaline uuring).

Moodul *StudyUnit* koondab endas kõiki uuringu komponente, mis hõlmavad eesmärgi, rahastamise, kvaliteedi aruannete, andmete kogumise ja kogumise meetodite kirjeldusi ja töötlemise tegevusi ja saadud andmete kirjeldusi.



Joonis 9 DDI mooduli *StudyUnit* koosseisu lihtsustatud vaade

Joonis 9 kujutab lihtsustatud ülevaadet, millised moodulid ja skeemid moodulisse *StudyUnit* imporditakse (st millega moodul *StudyUnit* seotud on).

DDI moodul *StudyUnit* ise sisaldab kahte komponenti – *Study Budget* ja *Study Unit*. Komponent *Study Budget* sisaldab omakorda kahte elementi ja komponent *Study Unit* sisaldab 46 elementi.

Mooduli *StudyUnit* kaardistamiseks on eelkõige aluseks võetud metasüsteemi iMeta moodul "Statistikatööd", sest nende sisud on eelduslikult kõige sarnasemad.

3.2.3.1.1 Komponenti *Study Budget* kaardistus metasüsteemiga iMeta

Komponenti *Study Budget* kaheks elemendiks on projekti eelarve kirjeldus (*Description*) ja viide eelarvedokumentidele (*BudgetDocument*). Elementide täpsema sisu ja võrdluse

metasüsteemi iMeta vastavate atribuutidega leiab lisast 1 (vt Lisa 1 Mooduli *StudyUnit* kaardistus metasüsteemiga iMeta).

Võrdluse tulemusena selgus, et komponendi *StudyBudget* elemente käesoleval ajal metasüsteemis iMeta ei dokumenteerita. Töö autor leidis, olles konsulteerinud eelnevalt statistikavaldkonna inimestega, et nende kahe elemendi lisamine metasüsteemi iMeta oleks mõistlik.

3.2.3.1.2 Komponendi *Study Unit* kaardistus metasüsteemiga iMeta

Komponent *Study Unit* sisaldab 46 elementi. Komponendis *Study Unit* sisalduvad uuringu eesmärk, taust, arendamine, uuringuga seotud andmete kogumine ja andmete avaldamine (statistikatooted).

Komponendi täpsema võrdluse metasüsteemi iMeta vastavate atribuutidega leiab lisast 1 (vt Lisa 1 Mooduli *StudyUnit* kaardistus metasüsteemiga iMeta).

Analüüsi tulemusena leidis töö autor, et komponendi *Study Unit* elemente dokumenteeritakse metasüsteemis juba suuremas osas, kuid vaja oleks täiendada seitsme elemendiga.

3.2.3.2 Mooduli *LogicalProduct* võrdlus metasüsteemiga iMeta

Moodul *LogicalProduct* sisaldab 98 elementi, 44 komplekstüüpi (*ComplexTypes*) ja 7 lihsat tüüpi (*SimpleType*) elementi.

Moodul *LogicalProduct* kirjeldab kvantitatiivsete andmete loogilist sisu. Baseerub moodulil *BaseLogicalProduct* ja sisaldab kõiki mooduli *BaseLogicalProduct* karakteristikuid, kuid samuti konkreetseid sisulisi kvantitatiivseid andmeid. Lisaks sisule, mille pärib moodulist *BaseLogicalProduct* (e alusdokumendi struktuur), sisaldab järgmisi DDI skeeme:

- kategooria skeem (*CategoryScheme*),
- koodiloendi skeem (*CodeListScheme*),
- muutuja skeem (*VariableScheme*),
- kuubi skeemid (*NCubeScheme*) ja
- uuringus kasutatud muutuja skeem (*RepresentedVariableScheme*)

nii masin- kui ka inimloetava viitena ja teavet selle kohta, kuidas andmed on organiseeritud komplekselemendis *LogicalRecords* ja nendevahelised seosed (e kuidas andmed on organiseeritud lähtebaasis andmeelementide kogumis ja nendevahelised seosed).

Lisa 2 DDI mooduli *LogicalProduct* kaardistus on toodud võrdlus metasüsteemi iMeta moodulites "Klassifikaator" ja "Tehniline tunnus" olevate atribuutide ja DDI mooduli *LogicalProduct* vastavate elementide vahel.

Töö autor valis metasüsteemi iMeta moodulitest "Klassifikaator" ja "Tehniline tunnus" eelkõige nende võimaliku sisu, mida sai DDI skeemide nimetustest (vt Joonis 9) tuletada, kokkulangevuse tõttu.

3.2.3.3 DDI standardi kooskõla Neuchâтели TM mudeliga

DDI standardi ja Neuchâтели TM mudeli võrdlust ei ole rahvusvahelisel tasemel tehtud, kuid nii DDI standardi kui ka Neuchâтели TM mudeli väljatöötamisel on aluseks olnud ISO/IEC 11179 üldine metaandmete raamistik.

3.2.3.3.1 Neuchâтели muutuja mudeli võrdlus DDI standardiga

Kuivõrd metasüsteemis iMeta, kogu statistikatöö kirjeldamise loogikas ja edaspidi ka DDI standardiga võrdlemise seisukohalt on oluline osa Neuchâтели muutuja mudelil [14], mis käsitleb muutujat mitmel tasandil, siis järgnevalt lühiülevaade sellest mudelist.

Neuchâтели muutujate mudel käsitleb muutujat kui statistikasüsteemi põhiobjekti kahetasandilisena — muutujad väljaspool konkreetset konteksti (tunnus, mõiste) ja muutuja konkreetse statistikatöö kontekstis (muutuja). Mõistet käsitletakse seoses objekti iseloomustava tunnusega vastavalt terminoloogiatöö standardile.

Ühelt poolt on tegemist kontseptuaalse muutujaga e mõiste või tunnusega ning teisalt on tegemist kontseptuaalse muutujaga e muutujaga konkreetse statistikatöö kontekstis.

Muutuja on statistilist üksust iseloomustava tunnuse kaudu seotud mõistega. Koguselise muutujaga on seotud ka mõõtühik. Kõiki nimetatud objekte hallatakse metasüsteemis iMeta.

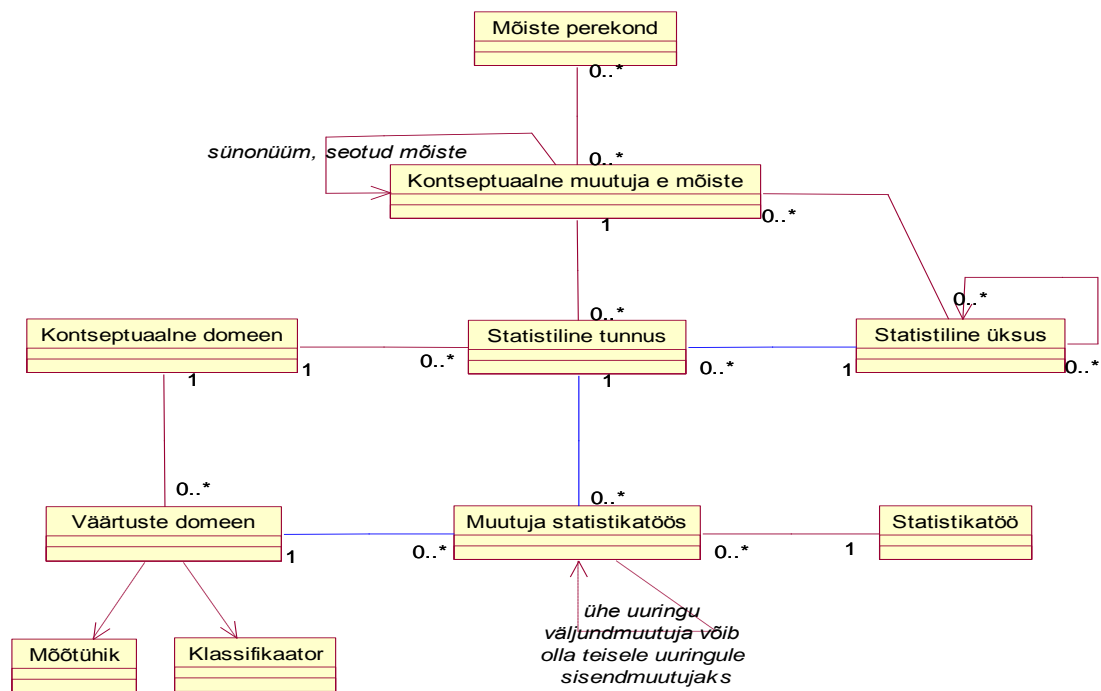
Mõisted, tunnused, muutujad, statistilised üksused ja statistikatööd moodustavad Neuchâтели muutuja mudeli järgi ühe terviku.

SA kasutab lihtsustatud Neuchâтели muutuja mudelit, mis on SA vajadustele kohandatud ja mis kirjeldab järgmisi metaandmete komponente, metaandmete komponentide atribuute ning nendevahelisi seoseid (vt Joonis 10):

- statistikavaldkonnad, statistikatööd ja nende versioonid ning statistikatööde muutujad;

- statistiliste üksuste tüübid, statistilised tunnused;
- mõistete perekonnad ja mõisted;
- klassifikaatorite perekonnad, klassifikaatorid, klassifikaatorite versioonid, klassifikaatorite elemendid, vastavustabelid, kodeerimissõnastikud;
- mõõtühikute tüübid ja mõõtühikud.

Järgneval joonisel (vt Joonis 10) on metasüsteemis iMeta kasutusel olev lihtsustatud Neuchâтели muutuja mudel.



Joonis 10 SA-s kasutatav Neuchâтели muutuja mudeli diagramm [14]

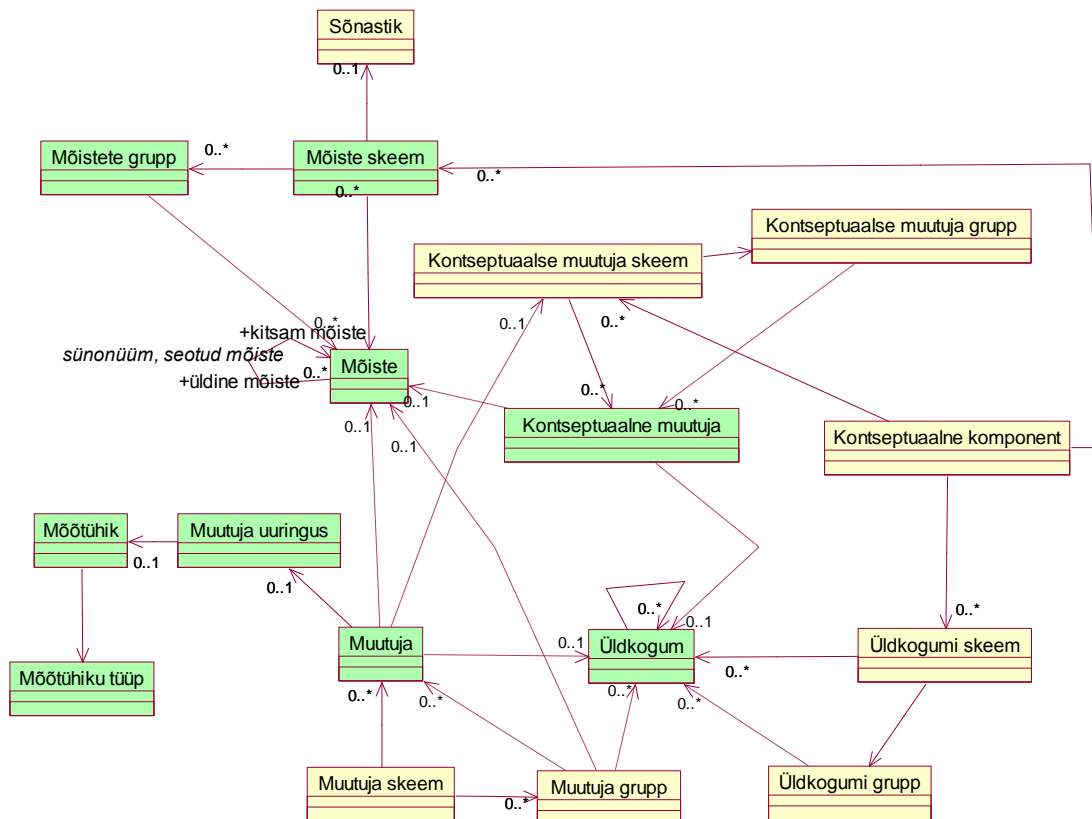
Joonisel (vt Joonis 10) olevad pidevad jooned näitavad olulisust just metasüsteemis iMeta seoste loomise aspektist lähtudes, sest see võimaldab statistikatööde üleselt semantilist koosvõimet ja standardiseeritust tagada. Sinised jooned märgivad kohustuslikku seost metaandmete komponentide vahel ja katkendlik joon näitab võimalikku metaandmete komponentide vahelist seost.

DDI standardi asjakohased mõisted ja nende selgitused on järgmised: [13]

- Mõiste (*Concept*) - kirjeldab mõistet standardi ISO/IEC 11179 järgi.
- Mõistete grupp (*ConceptGroup*) - võimaldab mõistete grupeerimist; grupid võivad olla hierarhilise struktuuriga.

- Mõiste skeem (*ConceptScheme*) - põhjalik mõistete loetelu, mida on andmete küsimisega dokumenteeritud.
- Muutuja (*Variable*) - kirjeldab muutuja struktuuri, mida kohaldatakse andmeelemendi väljendamisel andmekogumis.
- Muutujate grupp (*VariableGroup*) - sisaldab muutujate gruppi, mis võib olla lineaarne või hierarhiline.
- Muutuja uuringus (*VariableRepresentation*) - kirjeldab muutuja esindatust andmekogumis.
- Kontseptuaalne muutuja e mõiste (*ConceptualVariable*) - kirjeldab kontseptuaalset muutujat, mis pakub seost mõiste ja konkreetse üldkogumi (objekti) vahel, mis defineerib seda kui kontseptuaalset mõistet. Võib sisalda alammõisteid ja/või viiteid sarnastele mõistetele

Uurides DDI standardist vastavaid seoseid, siis töö autori hinnangul näevad vastavad seosed välja järgnevalt (vt Joonis 11):



Joonis 11 DDI standardis muutuja mudel

Kahe mudeli võrdluses on näha, et leidub sarnasusi, kuid on ka mõned erinevused. Põhiosa, mis kirjeldab muutujat ja selle seost mõistega, on sama, ja see tuleneb sellest, et mõlemad mudelid toetuvad standardile ISO/IEC 11179.

Peamine erinevus DDI standardis seisneb selles, et DDI standard kasutab erinevaid skeeme, mis kombineerib metaobjekti koodi vastava kirjeldusega ja metaobjekti grupiga, millesse see metaobjekt (antud juhul mõiste, muutuja ja üldkogum) kuulub, kuid see ei muuda metaobjektide vahelist seoste loogikat.

3.2.3.3.2 Neuchâтели klassifikaatori mudeli võrdlus DDI standardiga

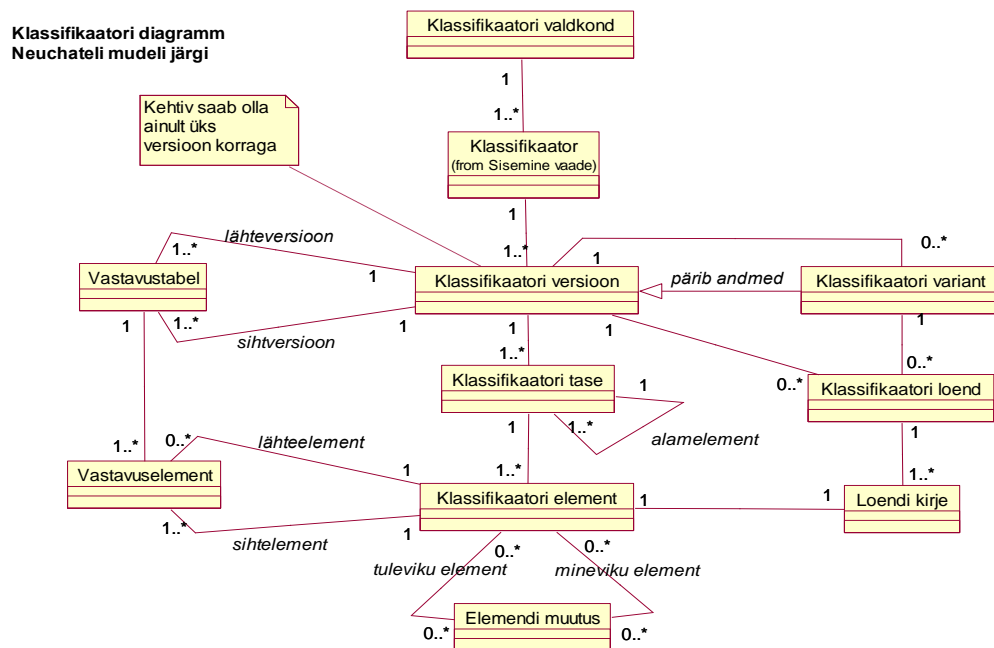
Klassifikaatorite andmebaasi loomisel on kasutatud Neuchâтели klassifikaatorite mudelit (vt Joonis 12). [19]

Klassifikaatorid ja vastavustabelid kui metaandmed on statistikasüsteemi lahutamatu osa ning neid hallatakse samuti metasüsteemis iMeta. Vastavustabelid hõlmavad nii erinevate klassifikaatorite vahelisi vastavustabeleid kui ka sama klassifikaatori erinevate versioonide ja variantide üleminekutabeleid.

Selle mudeli järgi kuuluvad klassifikaatorid valdkondadesse ja klassifikaatorid asuvad nn "katuse" all, mis väljendub klassifikaatori nimetusena. Iga katuse all võib olla mitu klassifikaatori versiooni, kuid ainult üks korraga saab olla kehtiv. Versioonil võib olla mitu kehtivat varianti.

Variant võib olla versioonist laiendatud, grupeeritud või ümbergrupeeritud, kuid oluline tunnus on, et üks klassifikaatori versiooni tase on siduvaks tasemeks klassifikaatori variandiga. Üheks nn eriliseks klassifikaatori variandiks on SDMX standardi standardiseeritud koodiloendid, mille alusel toimub agregeeritud andmete vahetus.

Nii versioonist kui ka variandist võib vastavalt vajadusele luua loendeid, kuid kõik elemendid peavad olema pärit sellest versioonist või variandist, millest see loend on tuletatud.



Joonis 12 SA-s kasutatav Neuchâtel'i klassifikaatori mudeli diagramm [19]

Selleks, et võrrelda Neuchâtel'i klassifikaatorite mudelit DDI standardiga, tuli esmalt võrrelda DDI standardis kasutusel olevaid vastavaid mõisteid.

Selleks tuli lähtuda DDI moodulis *LogicalProduct* sisalduvast kategooria skeemist *CategorySchemes* ja klassifikaatori skeemist *CodeListSchemes* ning kaasata võrdlusinfo moodulist *Comparative* vastavaid elemente, mis annavad vastavuse Neuchâtel'i klassifikaatori mudeli metaobjektidele.

DDI standardi asjakohased mõisted ja nende selgitused on järgmised: [1]

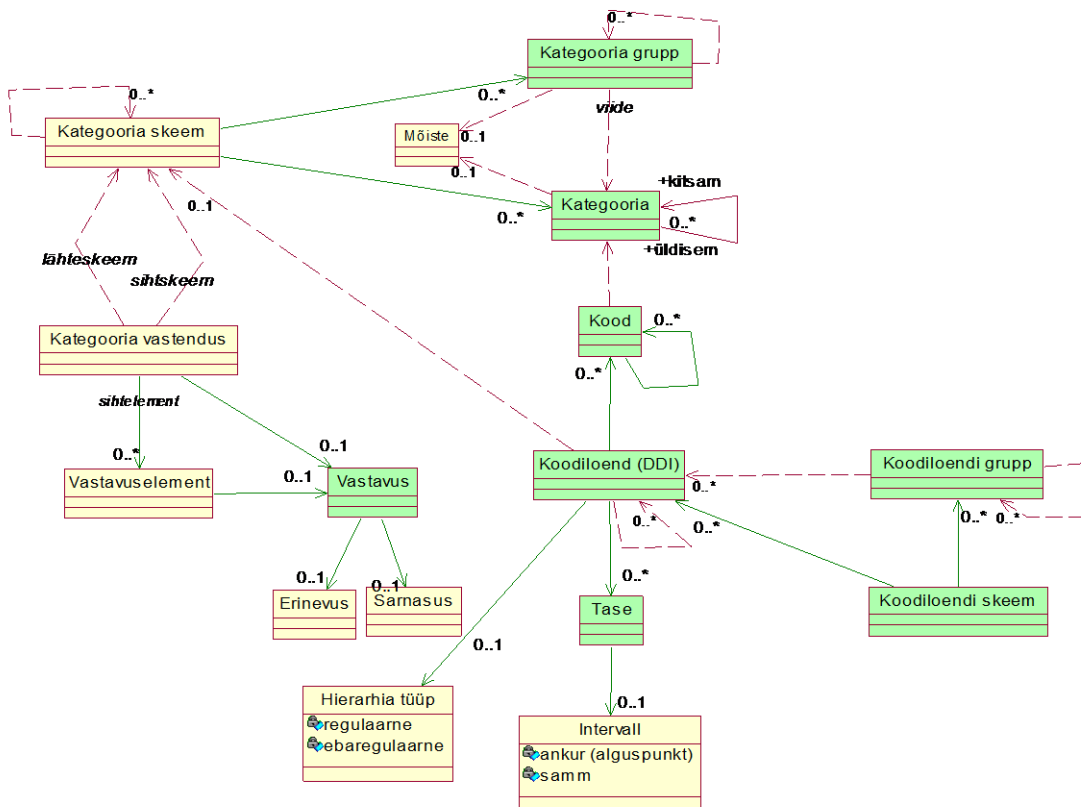
- Koodiloend (*CodeList*) - struktuur, mida kasutatakse, et siduda koodiväärtuste loetelu kindlaks määratud kategooriatega. Koodiloend võib olla lineaarne või hierarhiline.
- Kood (*Code*) - struktuur, mis ühendab koodi unikaalse väärtuse kindla kategooriaga.
- Tase (*Level*) - kasutatakse, et kirjeldada koodiloendi hierarhia tasemeid.
- Koodiloendi grupp (*CodeListGroup*) - koodiloendite grupeerimine kontseptuaalsetel või halduslikel eesmärkidel.
- Koodiloendi skeem (*CodeListScheme*) - skeem, mis sisaldab koodiloendeid, mida kasutatakse, et määrata kindlaks kood vastavalt väärtusele ja vastuste valdkonnale.
- Kategooria (*Category*) - Konkreetse kategooria või vastuse kirjeldus

- Kategooria grupp (CategoryGroup) - sisaldab kategooria kirjelduste gruppi, mis võib olla lineaarne või hierarhilise
- Kategooria skeem (CategoryScheme) - sisaldab kategooriate kirjeldusi, mida kasutatakse küsimuste vastustes ja andmete kirjelduses.

Analüüsi jaoks vajalikud elemendid moodulist *Comparative* on:

- Kategooria vastendus (CategoryMap) - vastendab sama tüüpi objektide kahe erineva kategooria skeemi sisu võrreldavateks kirjeteks nendes kahes skeemis.
- Vastavuselement (ItemMap) - vastendab sama tüüpi kaks üksust kindlaks määratud lähte- ja sihtskeemide vahel.
- Vastavus (Correspondence) - kirjeldab kahe üksuse vahelist sarnasust ja erinevust.

Mõistete selgituste ja spetsifikatsiooni põhjal kujunes vastav klassidiagramm, mis on kujutatud järgneval joonisel (vt Joonis 13).



Joonis 13 DDI standardi koodiloendi mudeli diagramm

Kahe diagrammi võrdluses on näha, et DDI standardis ja Neuchâтели klassifikaatori mudelis on sarnasusi, kuid on ka mõned erinevused. Peamised erinevused, lisaks terminite kasutamise

erinevustele, mis ilmnesid, olid:

- 1) DDI standardis eraldatakse kood kategooriast, Neuchâтели klassifikaatori mudelis ei ole kood eraldi välja toodud, kuid ta asub klassifikaatori elemendi juures.
- 2) DDI standardis on võimalik viidata koodiloendi sees teisele koodiloendile, mida Neuchâтели klassifikaatori mudelis klassifikaatori puhul ei ole võimalik teha;
- 3) DDI standardis on kasutusel koodi intervall, kui tegemist on intervallskaalaga.

Suurim erinevus seisnebki selles, et DDI standardis käsitletakse kategooriaid eraldi ning koodiloendi kood on ühendavaks lüliks, millega konkreetne kategooria seotakse. Ühtlasi viidatakse koodiloendiga kategooriaskeemile, millesse see kategooria kuulub.

4. Metasüsteemi iMeta täiendused

Käesolevas magistritöös võrreldakse DDI standardit SA metasüsteemiga iMeta, standardi integreerimise võimalust metasüsteemi ning metasüsteemi muutmise vajadust ja mahtu tulenevalt analüüsitulemustest.

Antud töö raames on võrreldud Neuchâтели TM mudelit DDI standardis vastavate metaobjektiga ning nende omavahelisi seoseid.

Kaardistatud on moodul *StudyUnit* ja üks osa moodulist *LogicalProduct*, millest sai välja valitud Neuchâтели TM mudeli metaobjektidega võrdlemiseks vajalikud skeemid ja nendega seotud metaobjektid.

Mooduliga *LogicalProduct* seotud elementide osas sai käesoleva magistritöö raames võrreldud mooduli enda elemente ning koodiloendi skeemi (*CodeListScheme*), koodiloendi (*CodeList*), koodiloendi grupi (*CodeListGroup*), ja kategooria (*Category*) elemente metasüsteemi atribuutidega.

Kuna DDI elementide ja metasüsteemi atribuutide võrdlus osutus terminoloogilise erinevuse tõttu märksa mahukamaks, kui eeldada võis, siis mooduli *LogicalProduct* lõplikku võrdlust ei ole tehtud. Seega on edaspidi vaja lisaks kaardistada mooduli *LogicalProduct* ülejäänud skeemid ja elemendid.

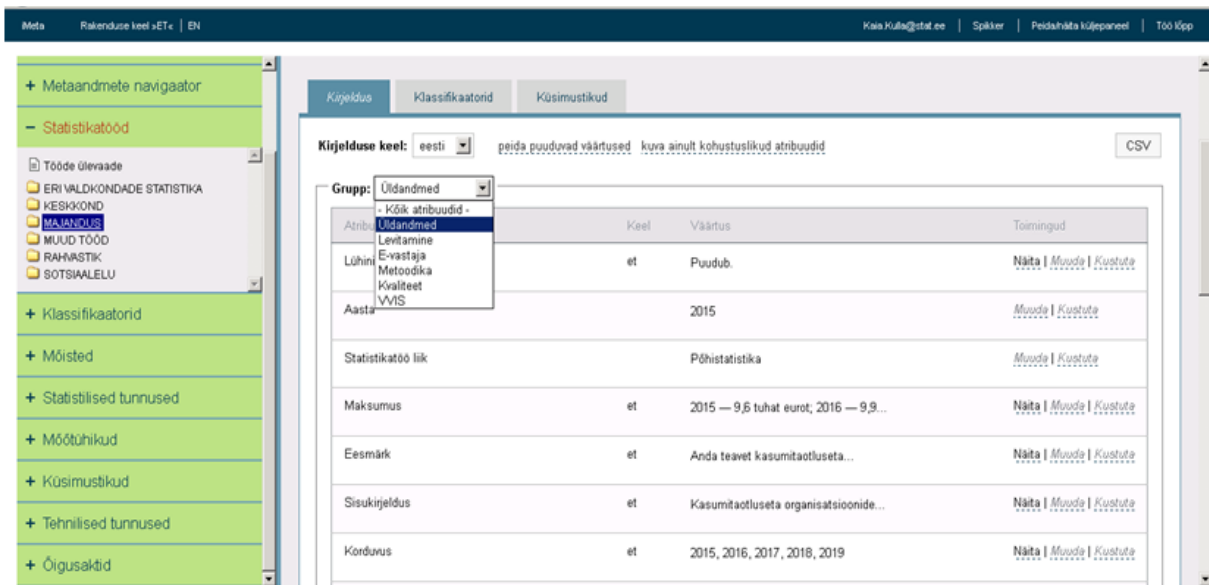
4.1 Metaandmete loetelu metasüsteemis iMeta täiendamiseks

Mooduliga *StudyUnit* võrdlusest selgust, et selle mooduli kasutusele võtmine väga palju muudatusi metasüsteemi iMeta andmekoosseisus kaasa ei too.

Tulenevalt punktides 3.2.3.1.1 Komponenti *Study Budget* kaardistus metasüsteemiga iMeta ja 3.2.3.1.2 Komponenti *Study Unit* kaardistus metasüsteemiga iMeta kajastatud analüüsist tuleks mooduli *StudyUnit* integreerimiseks metasüsteemi iMeta lisada kokku üheksa täiendavat DDI standardi elementi. Samas, DDI standardis võivad elemendid olla nii kompleks- kui ka lihtelemendid.

Arvestades, et osa metasüsteemi iMeta lisatavatest elementidest on DDI standardis komplekselemendid, mis koosnevad omakorda elementidest, siis tegelik metasüsteemi iMeta täiendamise vajadus võib kujuneda teistsuguseks, kui esialgne analüüs näitab.

Seega, metasüsteemi iMeta arenduste käigus tuleb arvestada, et kui tegemist on komplekselemendiga, siis võib tekkida vajadus metasüsteemis iMeta täiendavate atribuudigruppide¹ järele (vt Joonis 14). Vajadus sõltub sellest, kas edasise analüüsi käigus SA peab vajalikuks komplekselemendi elemente kirjeldada (dokumenteerida) või mitte.

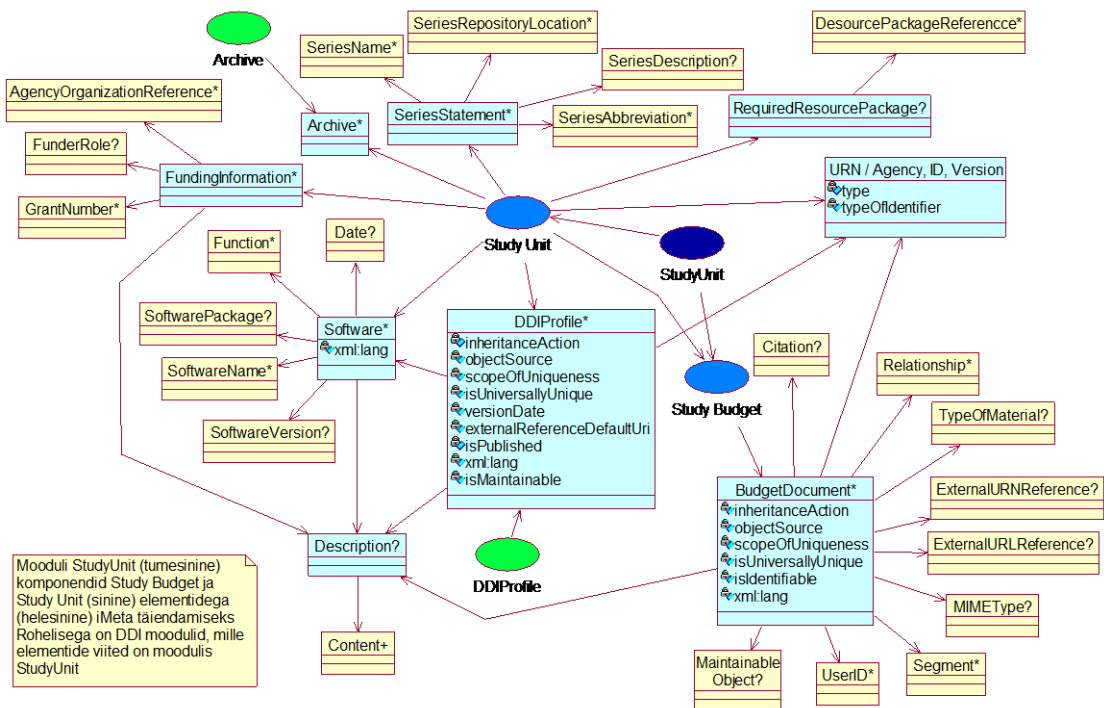


Joonis 14 Metasüsteemi iMeta illustreeriv joonis

Kui väljavalituks osutub lihtelement, siis seda on võimalik lisada juba mõne olemasoleva atribuudigrupi koosseisu täiendava atribuudina.

Joonis 15 Mooduli *StudyUnit* elemendid metasüsteemi iMeta täiendamiseks on toodud helesinisega elemendid, mida tuleks lisada metasüsteemi iMeta.

¹ Metaandmete haldussüsteemis iMeta nimetatakse atribuudiks statistikatööd ja selle versiooni kirjeldavat tunnust, selleks, et eristada seda statistilisest tunnusest



Joonis 15 Mooduli *StudyUnit* elementid metasüsteemi iMeta täiendamiseks

Helekollasega elementide osas tuleb läbi viia täiendav analüüs, kas nende lisamist metasüsteemi iMeta peetakse otstarbekaks või mitte. Kuna tegemist on elementidega, mis ei kuulu vaatlusalusesse DDI moodulisse, siis käesolevas töös neid põhjalikumalt ei analüüsitud. Rohelisega on eraldiseisvad moodulid, mille koosseis tuleb samuti edaspidi eraldi analüüsida.

Mooduliga *LogicalProduct* võrdlusest, sh koodiloendi skeemi (*CodeListScheme*), koodiloendi (*CodeList*), koodiloendi grupi (*CodeListGroup*) ja kategooria (*Category*) võrdlusest selgust, et analüüsitud osas selle mooduli kasutusele võtmine toob kaasa vaid kahe täiendava elemendi lisamise metasüsteemi iMeta.

Kuna DDI standardi kõikide elementide kasutamine ei ole kohustuslik, siis standardi mittekohustuslikest elementidest osutusid väljavalituks need, mille kohta on SA-1 tegelikult info olemas, kuid seni metasüsteemis ei dokumenteerita.

Tabel 3 Metasüsteemis iMeta vajalikud täiendused punktid 1 kuni 10 kajastavad ülevaadet, millised muudatused tuleks metasüsteemis iMeta teha tulenevalt mooduli *StudyUnit* ning selle komponentide *Study Budget* ja *Study Unit* analüüsist ja tabelis 3 punkt 11 kajastab ülevaadet, millised muudatused tuleks metasüsteemi iMeta teha tulenevalt mooduli *LogicalProduct* analüüsist.

Tabel 3 Metasüsteemis iMeta vajalikud täiendused

Jrk nr	Elemendi nimetus	Elemendi tüüp	Elemendi sisu	Sisu kirjeldus [1]	iMeta moodul - ettepanek	Atribuudigrupp
1	<i>r:BudgetDocument</i>	Kompleks	6 atribuuti	Vt http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocu- mentation/	Statistikatöö	Luu eraldi atribuudigrupp Eelarve
2	<i>multiplicity</i> [0 *]		14 elementi			
3	<i>r:Description</i> <i>multiplicity</i> [0 .. 1]	Kompleks	1 element	Sisu (kohustuslik)	Statistikatöö	Eelarve, tarkvara
4	r:URN	Liht	2 atribuuti	Tüüp ja tüübi formaat	Vajalik XML faili loomisel, eraldi atribuudina ei ole vaja lisada (URN =(<i>r:Agency</i> , <i>r:ID</i> , <i>r:Version</i>))	
5	<i>r:Software</i>	Kompleks	1 atribuut	tarkvara keel	Statistikatöö	Luu eraldi atribuudigrupp Tarkvara
			6 elementi	tarkvara nimetus (sh väljaandja), pakett, versioon, kirjeldus, kuupäev (mil väljalase on toetatud tootja poolt) ja funktsionaalsus (mida kasutati töös);		
6	<i>r:SeriesStatement</i>	Kompleks	4 elementi	Asukoht repositooriumis, seeria nimi, lühend ja kirjeldus	Statistikatöö	Üldandmed
7	<i>r:FundingInformation</i>	Kompleks	4 elementi	Viide organisatsioonile, rahastaja, grand number ja kirjeldus	Statistikatöö	Eelarve

Jrk nr	Elemendi nimetus	Elemendi tüüp	Elemendi sisu	Sisu kirjeldus [1]	iMeta moodul - ettepanek	Atribuudigrupp
8	<i>r:RequiredResourcePackages</i>	Kompleks	1 element	Üksikasjalik loetelu, lähtuvalt moodulist <i>ResourcePackages</i> , mis on vajalik, et moodul osadeks jagada.	Vajalik XML faili loomisel, eraldi atribuudina ei ole vaja lisada	
9	<i>a:Archive</i>	Kompleks	9 atribuuti 21 elementi	Vt http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/	Tegemist on eraldi mooduliga, mida antud töö kontekstis ei ole analüüsitud. iMeta seisukohalt tuleks arvestada selle mooduli analüüsi tulemustega	
10	<i>pr:DDIProfile</i>	Kompleks	9 atribuuti 24 elementi	Vt http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/	Tegemist on eraldi mooduliga, mida antud töö kontekstis ei ole analüüsitud	
11	<i>r:ConceptReference</i>	Kompleks	7 atribuuti 6 elementi	Vt http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/	Lisada klassifikaatori valdkonda atribuut "Viide mõistele"	Lisada klassifikaatori elemendile atribuut "Viide mõistele" Atribuudid täidetakse infoga (st lisatakse viide mõistele), kui mõiste on olemas

4.1.1 Neuchâтели TM mudeli ja DDI standardi ühilduvus

Neuchâтели muutuja mudelis oli olemas selge ühisosa DDI standardi vastava modelleeritud mudeliga. Peamised metaobjektide vahelised seosed olid samad, seega DDI standard ühilduvus Neuchâтели muutuja mudeliga sai kinnitust (mõlemal juhul oli eelduseks standardiga ISO/IEC 11179 ühilduvus).

Ka Neuchâтели klassifikaatori mudelil oli võrdluses DDI standardi vastava modelleeritud mudeliga ühilduvus olemas, vaatamata DDI standardis kategooria erinevale käsitlusele võrreldes Neuchâтели mudeli klassifikaatori elemendiga, kuid seda vaid skeemi ülesehitusest lähtudes, mitte sisuliselt.

Arvestades DDI standardi paindlikkust elementide kasutamisel ning teisalt metasüsteemi iMeta ülesehitust, mis võimaldab klassifikaatori elementi, mis vastab DDI standardi kategooriale, kirjeldada DDI skeemis eraldi elemendina, siis töö autori seisukohalt ei vaja selles osas mingeid muudatusi metasüsteemis iMeta teha.

4.2 Metasüsteemi iMeta vajalikud täiendused

Selleks, et DDI standard täielikult integreerida metasüsteemi iMeta, on vaja teha võrdlus kõigi DDI moodulite osas. Lisaks tuleb kaardistada XML nimeruumide elemendid, millised on vajalikud DDI skeemide XML-failide genereerimisel.

Peale selle, et on vaja DDI standard moodulite kaupa kaardistada metasüsteemiga iMeta, on vaja realiseerida ka metasüsteemi iMeta täiendused, mis on vajalikud kõigi juurutatavate moodulite ja XML skeemide jaoks.

Järgnevalt on loetletud täiendused, mis puudutavad tervikuna DDI standardi kasutusele võtmist:

- realiseerida DDI moodulid metaandmete kirjelduste esitamiseks DDI formaadis nii HTML esituses inimloetaval kujul kui ka XML formaadis masinloetaval kujul;
- luua vahend HTML väljundi inimloetavaks esitluseks ja siduda see esialgse (originaal) XML-failiga;
- realiseerida metasüsteemist iMeta DDI standardile vastavate XML skeemide genereerimise vahend (nt XML API vahendusel);
- töötada välja vastavad teenused:
 - küsimustike veebis ja küsitlejarakenduses esitamiseks;

- teadlastele andmetega seotud metaandmete edastamiseks;
- kodulehel üksikandmeid kirjeldavate metaandmete esitamiseks;
- täiendada MMX metaandmete repositooriumi vastavalt.

4.3 Rakendamise kava

Arvestades DDI standardi keerukust ja mahukust, teeb töö autor ettepaneku DDI standardi kasutusele võtmise jagada etappideks ning jaotada metasüsteemi iMeta arendused vastavalt analüüsitud mooduli(te) tulemustele. Sõltuvalt mooduliga kaasnevast arenduste mahust, rakendada moodulid kas ükshaaval või mitu moodulit üheaegselt. Samuti soovib töö autor enne reaalsete arenduste tellimist analüüsida põhjalikult DDI standardit ja selle kasutamise iseärasusi.

Kuna käesolevas töös käsitleti vaid väikest osa DDI standardist, siis täpset ajakava standardi rakendamiseks on keeruline esitada. Vaatamata sellele, teeb töö autor eelnevast tulenevalt DDI standardi rakendamise ajakava suhtes järgmise ettepaneku:

I etapp 01.07.2014 kuni 30.06.2015

- alustada DDI standardi kasutusele võtmise ettevalmistustöödega:
 - ettevalmistustööd seisneksid peamiselt DDI moodulite elementide kaardistamises ja võrdlemises metasüsteemiga iMeta ning vajalike elementide metasüsteemi jaoks väljaselgitamises;
 - moodulite läbivaatamise järjekord tulenevalt punktist 3.2.3 DDI standardi sobivus metasüsteemiga iMeta on: jätkata mooduli *LogicalProduct* analüüsiga, analüüsida moodulid *Conceptual Component*, *DataCollection*, *PhysicalDataProduct*, *PhysicalInstance* ning seejärel ülejäänud moodulid;
 - andmelao relatsioonilise andmebaasi muutmise ja/või täiendamise vajaduste analüüs DDI standardist lähtudes;
- leida võimalus DDI standardi kasutamise koolituseks, kas:
 - kutsudes koolitaja väljastpoolt, kes on tuttav DDI standardiga või
 - saata 1 kuni 2 isikut DDI standardi koolitusele, kes omakorda koolitusel saadud teadmised majasiseselt edasi viiksid.

II etapp 01.07.2015 kuni 30.08.2016

- arenduste tellimine, sh vajalike riigihangete korraldamine ja läbiviimine:
 - metasüsteemi iMeta arendused,
 - DDI standardi XML skeemide genereerimise arendused,
 - teenuste arendused,
 - andmelao arendused,
 - prototüüpide tegemised ja
 - testimine.

III etapp 01.09.2016 kuni 31.12.2016

- metasüsteemi lõppkasutajate koolitused DDI standardi kasutusele võtmiseks.

DDI standardi kasutusele võtmine võiks toimuda alates 2017. aastast.

Paralleelselt nende etappidega ja DDI standardi sujuvamaks kasutusele võtmiseks:

- luua sõnastik, millega töö autor alustus magistritöö tegemise käigus ja mis on vajalik DDI standardi paremaks mõistmiseks, sest mõistete kasutamine DDI standardis ei vasta alati seni SA-s kasutusolevale inglisekeelsele terminoloogiale;
- täiendada metasüsteemis iMeta statistikatööde kirjeldamise juhend täiendavate atribuutide (DDI standardi mõistes elementide) kirjelduste kohta.

4.4 Tähelepanekud DDI standardi analüüsist

DDI standardi metasüsteemiga iMeta sobivuse analüüsimisel kerkisid üles järgmised teemad, mida oleks vaja täiendavalt analüüsida:

- küsitlisdokumendi defineerimise mooduli KDM arendamine või väljavahetamine, et kord kirjeldatud uuringu kirjeldus üks kord kirjeldada ankeeti;
 - selleks, et hinnata, kas küsitlisdokumendi defineerimise moodul KDM võiks välja vahetada, tuleks eelnevalt uurida turul olevaid küsimustike disainimise töövahendeid ja kaaluda nende eeliseid (maksumus, arendaja toetus ja edasiarendused) KDMi arendamise ees;
- analüüsida ja vajadusel rakendada GSIM mudel statistika tegemise protsessi.

4.5 Standardi DDI versiooni 3.2 töövahendid

DDI standardi kasutamisel pakutakse mitmeid töövahendeid küsimustike kirjeldamisel, osad nendest on vabavaralised ja/või tasuta, teised jälle tasulised. Käesoleval ajal on kasutusel järgmised töövahendid, mis toetavad standardit DDI versiooni 3.2:

- *Colectica Excel, Designer* [20]
 - võimaldab lisada Exceli programmile lisamooduli, mille abil on võimalik küsimustikke kirjeldada, kasutades muutujaid ja klassifikaatoreid ning hiljem genereerida kas DDI standardile vastava faili või salvestada Word, PDF või HTMLi;
- *Rogatus Platform* [21]
 - koosneb mitmest uuringu disaini töövahendist (*Qbee, Mbee, Tbee*)
 - toetab standardeid DDI ja SDMX;
 - toetab PAPI (paberankeedi), CAPI (arvutipõhise küsitluse), CATI (telefoniküsitluse) ankeetide tegemist ja *online*'is testimist;
 - ühildub olemasolevate DDI töövahenditega (*Colectica, Questasy*);
 - töövahend on jaotatud kaheks vaateks:
 - uuringu (küsimustiku) koostamiseks (uuringu haldamine)
 - uuringuandmetega töötamiseks (andmete haldamine)
- *queXML*. [22]

4.6 Piirangud DDI standardi kasutusele võtmiseks

Autor ei tuvastanud käesoleva töö tegemise käigus sellised piiranguid, mis seaksid kahtluse alla DDI standardi kasutusele võtmise võimaluse metasüsteemis iMeta. Ainsad piirangud, mis käesoleval ajal võivad tekkida on seotud ressurssidega, sest tegemist on siiski mahuka standardi rakendamisega:

- raha infosüsteemi võimalike muudatuste tegemiseks;
- aeg ja inimesed, et vajalike täiendusi infosüsteemi sisestada, millele lisandub ressurss arenduste testimiseks.

5. Kokkuvõte

Statistikaametil on eesmärk viia olemasolev metasüsteem iMeta vastavusse standardiga DDI, mis on statistiliste ja sotsioloogiliste andmete (sh metaandmete) kirjeldamise standard. Sellest tulenevalt oli käesoleva magistritöö koostamise eesmärk esmalt teha olemasoleva metasüsteemi iMeta (sh Neuchâтели TM mudeli) ja DDI standardi võrdlev analüüs ning välja selgitada nende integreerimise võimalus. Samuti oli vaja välja selgitada, kas ja milliseid arendusi on metasüsteemis iMeta vaja teha, et DDI standard kasutusele võtta.

Selleks, et magistritöö sissejuhatuses püstitatud eesmärki saavutada, tuli töö autoril esmalt uurida põhjalikult DDI standardit, selle ülesehitust, selles kirjeldatud metaobjekte ja nendevahelisi seoseid. DDI standardit ei ole töö autorile teadaolevalt Eestis varem kasutatud ega uuritud, mistõttu sai magistritöös DDI standardist natuke põhjalikum ülevaade antud. Ülevaade sisaldas ka DDI standardi võrdlust SA-s ja metasüsteemis rakendatud mudelite ja standarditega, nagu GSBPM mudel ja SDMX standard.

Töö tulemusena valmis kõigepealt võrdlus DDI mooduli *StudyUnit* ja metasüsteemi mooduli "Statistikatöö" vahel, millest tulenevalt sai koostatud loetelu mooduli *StudyUnit* elementidest, mida on vaja lisada metasüsteemi iMeta.

Teisena valmis DDI mooduli *LogicalProduct*, selles sisalduva koodiloendi skeemi (*CodeListScheme*), koodiloendi (*CodeList*), koodiloendi grupi (*CodeListGroup*) ja kategooria (*Category*) elementide võrdlus metasüsteemi iMeta moodulite "Klassifikaator" ja "Tehniline tunnus" atribuutide vahel. Analüüsitud osas pidas töö autor vajalikuks metasüsteemi iMeta täiendada paari elemendiga.

Seejärel valmis võrdlus Neuchâтели TM mudeli ja DDI standardi vastavate metaobjektide ja nendevaheliste seoste vahel, mille tulemusena selgus nende omavaheline ühilduvus.

Lisaks koostas töö autor ajakava ja tegi ettepaneku edasisteks tegevusteks DDI standardi kasutuselevõtmiseks SA metasüsteemis iMeta.

Käesoleva magistritöö põhjal saab väita, et DDI standardi kasutusele võtmine on võimalik, kuid arvestades selle standardi keerukust ja mahukust, siis on mõistlik selle standardi kasutusele võtmine planeerida pikema perioodi peale. Lisaks on vaja ka DDI standardi ülejäänud moodulid, mida antud magistritöös ei käsitletud, läbi analüüsida ja võrrelda olemasoleva metasüsteemiga iMeta. Oluline on aru saada terminoloogilistest erinevustest, mida kasutatakse metasüsteemis iMeta ja DDI standardis ning leida nende ühisosa.

Summary

Statistics Estonia plans to modify the existing metadata management system iMeta so that it would meet the Data Documentation Initiative (DDI) standard for describing statistical and sociological data (incl. metadata). Therefore, the aim of this Master's thesis was to carry out a comparative analysis of the existing metadata system iMeta (incl. the Neuchâtel classification model) and the DDI standard and determine whether they could be integrated. Another aim was to identify any developments that are required in the metadata management system before the implementation of the DDI.

In order to achieve this objective, the author studied the DDI standard in depth, including its structure, the range of meta objects and their interrelations. To the author's knowledge, the DDI standard has not been used or studied before in Estonia, which is the reason for the thorough overview of the standard in this Master's thesis. The overview includes a comparison of the DDI with the models and standards implemented by Statistics Estonia and in the metadata system (such as the GSBPM model and the SDMX standard).

The author began the analysis by comparing the StudyUnit module of the DDI and the "Statistical activity" module of Statistics Estonia's metadata system. Based on the results, the author prepared a list of StudyUnit elements that need to be added to the metadata system iMeta.

In the second part of the analysis, the author compared the CodeListScheme, CodeList, CodeListGroup and Category elements of the DDI module LogicalProduct with the attributes of the "Classification" and "Technical characteristics" modules of the metadata system iMeta. Based on this analysis, the author found that a few elements should be added to iMeta.

Next, the author compared the Neuchâtel TM model with the corresponding meta objects (and their interrelations) in the DDI standard. The comparison revealed that the DDI standard is compatible with the Neuchâtel TM model.

The Master's thesis includes a suggested schedule and action plan for the implementation of the DDI standard in Statistics Estonia's metadata system iMeta.

The results of this Master's thesis indicate that it would be possible to implement the DDI standard, but a longer period should be planned for its implementation, considering the level of complexity and range of the standard. In addition, the other modules of the DDI (i.e. those not covered by this Master's thesis) need to be analysed and compared with the existing metadata system iMeta. It is essential to understand the differences in the terminology used in the metadata system iMeta and in the DDI standard and to establish the extent of overlap.

Kasutatud kirjandus

- [1] DDI Alliance, „DDI Lifecycle 3.2 XML Schema Documentation,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/>. [Kasutatud 15 03 2014].
- [2] UNECE, „The Generic Statistical Business Process Model (GSBPM),“ [Võrgumaterjal]. Available: <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>. [Kasutatud 23 veebruar 2014].
- [3] UNECE, „Statistical Classification Model,“ [Võrgumaterjal]. Available: <http://www1.unece.org/stat/platform/display/gsim/Statistical+Classification+Model>. [Kasutatud 28 04 2014].
- [4] UNECE, „Neuchâtel Terminology Model,“ [Võrgumaterjal]. Available: <http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=14319930>. [Kasutatud 2014 märts 30].
- [5] „SDMX veebisait,“ [Võrgumaterjal]. Available: <http://www.sdmx.org>. [Kasutatud 15 03 2014].
- [6] „ISO 17369:2013 Statistical data and metadata exchange (SDMX),“ [Võrgumaterjal]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52500. [Kasutatud 20 02 2014].
- [7] UNECE, EUROSTAT, OECD, „DDI-SDMX Integration and Implementation,“ [Võrgumaterjal]. Available: <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP5.pdf>. [Kasutatud 23 02 2014].

- [8] „ISO/IEC 11179, Information Technology - Metadata registries (MDR),“ [Võrgumaterjal]. Available: <http://metadata-standards.org/11179/>. [Kasutatud 30 märts 2014].
- [9] DDI Alliance, „Data Documentation Initiative veebisait – What is DDI?,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/what>. [Kasutatud 23 veebruar 2014].
- [10] W. Thomas, „Göteborgs Universitet, SND Svensk Nationell Datatjänst,“ [Võrgumaterjal]. Available: <http://snd.gu.se/sites/snd.gu.se/files/DDITrainingWorkshop.pptx>. [Kasutatud 28 02 2014].
- [11] DDI Alliance, „DDI veebisait - Getting Started with DDI,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/getting-started#introduction> . [Kasutatud 15 02 2014].
- [12] DDI Alliance, „DDI Technical Specification, Part II: User Guide, versioon 3.2,“ [Võrgumaterjal]. Available: http://www.ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/HighLevelDocumentation/DDI_Part_II_UserGuide.pdf. [Kasutatud 15 03 2014].
- [13] DDI Alliance, „DDI veebisait - DDI Specification,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/specification>. [Kasutatud 23 02 2014].
- [14] UNECE, „Neuchâtel Terminology Model, PART II: Variables and related concepts objects types and their attributes,“ [Võrgumaterjal]. Available: <http://www1.unece.org/stat/platform/download/attachments/14319930/Neuchatel%20Model%20V1.pdf?version=1&modificationDate=1265695890041&api=v2>. [Kasutatud 28 03 2014].
- [15] O. D. F. Arofan Gregory, „The Data Documentation Initiative (DDI): An Introduction for National Statistical Institutes,“ [Võrgumaterjal]. Available: http://odaf.org/papers/DDI_Intro_forNSIs.pdf . [Kasutatud 23 02 2014].
- [16] Thomas, Wendy (University of Minnesota), Arofan Gregory (Open Data Foundation), Alistair Hamilton (Australian Bureau of Statistics), „Metadata standards to support

- controlled access to microdata,“ [Võrgumaterjal]. Available: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/41_Thomas.pdf. [Kasutatud 30 03 2014].
- [17] U. Steven Vale, „Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/resources/publications/working/others/ExploringRelationshipBetweenDDI-SDMX-GSBPM.pdf>. [Kasutatud 15 03 2014].
- [18] „Eurostat info space,“ Eurostat, [Võrgumaterjal]. Available: https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Census_Hub. [Kasutatud 20 04 2014].
- [19] UNECE, „Neuchâtel Terminology Model, Classification database object types and their attributes,“ [Võrgumaterjal]. Available: http://www1.unece.org/stat/platform/download/attachments/14319930/Part%20I%20Neuchatel_version%202_1.pdf?version=1&modificationDate=1265695896952&api=v2. [Kasutatud 30 03 2014].
- [20] DDI Alliance, „DDI Tools,“ [Võrgumaterjal]. Available: <http://www.ddialliance.org/resources/tools>. [Kasutatud 30 03 2014].
- [21] Ingo Barkow, David Schiller, „EDDI Conference System, EDDI13 – 5th Annual European DDI User Conference,“ [Võrgumaterjal]. Available: <http://www.eddi-conferences.eu/ocs/index.php/eddi/eddi13/paper/view/118>. [Kasutatud 28 04 2014].
- [22] „queXML veebisait,“ [Võrgumaterjal]. Available: <http://quexml.sourceforge.net/node/1>. [Kasutatud 24 04 2014].

Lisa 1 Mooduli *StudyUnit* kaardistus metasüsteemiga iMeta

Tabelites on järjekorranumbri juures kasutatud informatsioonina elemendi kohustuslikkuse või mitte dokumenteerimise jaoks järgmisi sümboleid järgmiste tähendustega:

+ kohustuslik; [] kohustuslik unikaalne; * valikuline; ? valikuline unikaalne.

NB! DDI standardis ei ole elementide kasutamine enamikel juhtudel kohustuslik ja see võimaldab organisatsioonil endal otsustada, milliseid elemente kasutada (e millist infot dokumenteerida), milliseid mitte.

Tabel 4 Komponent *Study Budget* elementide võrdlus metasüsteemiga iMeta

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
1*	Eelarve dokument (<i>r:BudgetDocument</i>)	Eelarve dokument. Viide ühele või enamale välisele eelarvedokumendile.	Puudub, lisada
2?	Kirjeldus (<i>r:Description</i>)	Eelarve kirjeldus ükskõik millise avaldatud dokumendi jaoks, mis võib sisaldada viidet välisele eelarve dokumendile.	Puudub, lisada

Kokkuvõte: komponendi *Study Budget* elemente käesoleval ajal metasüsteemis iMeta ei dokumenteerita. Töö autor leidis, olles konsulteerinud eelnevalt statistikavaldkonna inimestega, et nende kahe elemendi lisamine metasüsteemi iMeta oleks mõistlik.

Komponendi *Study Unit* elementide kaardistus metasüsteemiga iMeta

Komponent *Study Unit* sisaldab 46 elementi.

Tabel 5 sisaldab komponent *Study Unit* elementide võrdlust metasüsteemi iMeta atribuutidega.

Tabel 5 Komponent *Study Unit* elementide võrdlus metasüsteemiga iMeta

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
1[]	URN (r:URN)	URN olemi sobitamine DDI URN mustriga, mis on seotud atribuudi väärtusega <i>typeOfIdentifier</i> (normatiivne või ebasoovitav).	Puudub, lisada
2[]	Asutus (r:Agency)	Asutuse registrikood, vajadusel alamasutuse koodiga (nt osakonna lühinimega), mis on eraldatud omavahel punktiga (nt sa.mo).	Üldandmed / Organisatsioon, Kontaktisiku struktuuriüksus
3[]	ID (r:ID)	Objekti ID. See peab vastama DDI identifikaatorile lubatud struktuurile ja peab olema unikaalne.	Olemas on süsteemisene ID, mis on unikaalne
4[]	Versioon (r:Version)	Objekti versiooni number. Identifitseeritavale objektile on see tema versioneeritava vanema versiooni number selle mitteadministratiivsete tuvastatavate metaandmete loomise või muutmise hetkel.	Puudub, ei ole versioneerimist, kuid iga uus versioon saab uue ID, mis on unikaalne,
5*	Kasutaja ID (r:UserID)	Kasutaja identifikaator, mis võimaldab tuvastada kasutaja, kes atribuute haldab või muudab.	Olemas süsteemisene
6*	Kasutaja atribuutide paar (r:UserAttributePair)	Kasutaja määratletud süsteemispetsiifiline objekti omadus, mis on väljendatud võti-väärtus paarina. Kuna see on spetsiifiline eraldiseisva süsteemi kontrollitud sõnavara	Puudub, esialgu ei lisa

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
		kasutamise jaoks, siis võtme kasutamine on tungivalt soovitatav.	
7.?	Versiooni eest vastutaja / Viide vastutajale (<i>r:VersionResponsibility / VersionResponsibilityReference</i>)	Isik või organisatsioon haldavas asutuses, kes vastutab versioon muutuste eest. Kui on oluline säilitada kuuluvus versioonide eest vastutajate ja tema organisatsiooni vahel, siis see võib sisaldada vastavat märget. See on eelkõige mõeldud sisemiseks kasutamiseks. / Viide eelnevale.	Üldinfo / Kontaktisiku nimi, Kontaktisiku ametinimetus
8.?	Versiooni põhjendus (<i>r:VersionRationale</i>)	Tekstiline põhjendus/otstarbe kirjeldus versiooni muutuse ja kodeeritud väärtuse jaoks, et varustada süsteemis või organisatsioonis sisemine töötlemisprotsess vastava lipuga. NB! Versioneerimine saab toimuda ainult objektide osas, millel on omanikuks määratud DDI asutus. Kui luua teise asutuse objektist oma kohalik eksemplar praeguse või tulevase muutmise tarbeks, siis tuleks kasutada elementi <i>BasedOnObject</i> .	Puudub, esialgu ei lisa
9.?	Baasobjekt (<i>r:BasedOnObject</i>)	Kasutatakse, kui luuakse uus objekt, mis baseerub olemasoleval objektil või objektidel, mida haldavad erinevad asutused või kui uus objekt ei ole lihtsalt olemasoleva objekti versioon, vaid see on suurem muudatus, kuid oleks vaja säilitada seos olemasoleva objektiga.	Puudub, esialgu ei lisa
10*	Märkus (<i>r:Note</i>)	Võimaldab lisada täiendavat informatsiooni ID-ga objektidele.	Üldandmed / Märkus
11*	Tarkvara (<i>r:Software</i>)	Viide tarkvarale, mida kasutatakse metaandmete loomiseks ja/või haldamiseks. Seda elementi saab korrata, et kirjeldada	Puudub, lisada

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
		erinevad tarkvarad ja mitmed funktsioonid	
12*	Metaandmete kvaliteet (<i>r:MetadataQuality</i>)	Metaandmete kvaliteedi hindamine hallatava objekti kohta, näiteks transkribeerimise kvaliteet, täielikkus, toimetamise staatus jne	Kvaliteet / Kvaliteedi hindamine
13?	Viide avaldajale (<i>r:Citation</i>)	Bibliograafilised andmed esitatakse kasutades DDI struktuuri (vastavus <i>Dublin Core</i> objektidele). Ükski nendest objektidest ei ole vajalik tootmisprotsessi toetamiseks. Äärmiselt soovitatav on, et vähemalt pealkiri oleks esitatud.	Üldandmed / Statistikatöö, Aasta
14?	Lühikokkuvõte (<i>r:Abstract</i>)	Sisu lühikokkuvõte, kus kirjeldatakse uuringu olemus ja andmete kogumise ulatus, selle sisu eriomaduste kohta. NB! Üksikasjalik teave uuringu eesmärgi ja struktureeritud kaetuse kohta tuleb kirjeldada eesmärgi (<i>Purpose</i>) ja kaetuse (<i>Coverage</i>) elementides. Sisu lühikokkuvõte toetab sama sisu jaoks mitmekeelseid versioone, samuti on sisu vorming vabalt valitav.	Üldandmed / Sisukirjeldus
15*	Eelarvevahendite allikas (<i>r:AuthorizationSource</i>)	Uuringu eelarvevahendite allikas. Siin on võimalik viidata terviktekstile (seadusele, määrusele või muu volitusnormile).	Üldandmed / Tellija liik, Õigusaktid
16?	Viide üldkogumile (<i>r:UniverseReference</i>)	Viide üldkogumi formuleeringule, milles kirjeldatakse isikute rühma või muid elemente, mis on uurimisobjektiks ja millele analüütilised tulemused viitavad. Vanus, rahvus ja elukoht aitavad tavaliselt piiritleda antud üldkogumit, kuid mõned mitmed tegurid võivad olla seotud, nagu sugu, sissetulek jne. Üldkogum võib koosneda ka muudest elementidest peale inimeste, näiteks	Metoodika / Statistiline üldkogum

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
		eluasemed, kohtuasjad, surmad, riigid jne. Üldkogumit võib kirjeldada kui "kõikehõlmav" või "eksklusiivne". See <i>StudyUnit</i> tasandil viitamine on tavaliselt <i>UniverseScheme</i> ülemine kiht.	
17*	Järjeuuring (<i>r:SeriesStatement</i>)	Uuringud, eriti uuringute moodulid, võidakse omavahel ühendada üheks ühtseks uuringuks. Uus uuring kuulub mõlemasse sarja.	Puudub, lisada
18*	Viide kvaliteediraportile (<i>r:QualityStatementReference</i>)	Viide kvaliteediraportile. Viide seotud uuringu üldise kvaliteediga metaandmete või andmete kvaliteediraportile. Kvaliteediraport võib olla seotud väliste kvaliteedistandarditega.	Kvaliteet / Kvaliteedidokumendid
19*	Kvaliteedinäitajate loetelu / Viide kvaliteedinäitajate (<i>r:QualityStatementScheme / r:QualityStatementSchemeReference</i>)	Skeem, mis sisaldab kvaliteedi formuleeringuid, mis on seotud meetodika, metaandmete või andmete kvaliteediga. Kvaliteedi formuleeringud võivad olla seotud väliste kvaliteedistandarditega. Viide skeemile <i>QualityStatementScheme</i> .	Kvaliteet / Kvaliteedi tagamine
20*	Järelhindamine (<i>r:ExPostEvaluation</i>)	Uuringuprotsessi hindamine, mis sageli toimub pärast uuringu lõppu. Need võivad hõlmata selliseid küsimusi nagu uuringu ajastus, küsimuste järjestus uuringus, kulu / eelarve küsimusi, asjakohasust, uuringu institutsiooniline või õiguslik korraldus jne. Kui uuring on osa seeriast või rühmast (st tegemist on kestevuuringuga), võib see hõlmata algstaadiumi hindamist ning seetõttu muutuda hiljem vastavalt kestevuuringu järgmistele seeriatele.	Kvaliteet / Kvaliteedi tagamine
21*	Rahastamisfondide info (<i>r:FundingInformation</i>)	Sisaldab andmeid uuringu rahastamise kohta, sealhulgas informatsiooni grantide jms kohta.	Puudub, lisada

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
22*	Uuringu eelarve (<i>StudyBudget</i>)	Siin kirjeldatakse uuring kogu eelarve. Seda võib korrata eelarve eraldiseisvate tegevuste jaoks. See sisaldab struktureeritud kirjeldust ja ühte või mitut eelarvedokumenti, mis on kirjeldatud <i>OtherMaterial</i> tüübis	Üldandmed / Maksumus
23?	Eesmärk (<i>r:Purpose</i>)	Uuringu eesmärk, miks uuring toimus. See peaks sisaldama üksikasjalikku teavet esmase uuringu küsimuste või hüpoteeside kohta, samuti teavet võimalike õiguslike alusandmete kogumise kohta, nagu seadused, mis nõuavad rahvaloenduse andmete kogumist jaotamise eesmärgil. Juriidiline või muu volitus, tuleb esitada üksikasjalikult Tellija (<i>AuthorizationSource</i>) väljal. Eesmärk toetab sama sisu jaoks mitmekeelseid versioone, samuti on sisu vorming vabalt valitav.	Üldandmed / Eesmärk
24?	Kaetus (<i>r:Coverage</i>)	Kirjeldab uuringuüksuse kaetust. Täpne informatsioon aktuaalse (<i>Topical</i>), ajalise (<i>Temporal</i>) ja ruumilise kaetuse (<i>Spatial Coverage</i>) kohta sisaldub siin NB! Kaetus sellel tasandil peaks kaasama kõik madalama taseme moodulid või osad. Madalama taseme kirjelduse eesmärk on piirata kaetust siin kirjeldatu piires.	Metoodika / Sektorite kaetus
25*	Analüüsiüksus (<i>r:AnalysisUnit</i>)	Võimaldab kasutada kontrollitud sõnastikku (<i>Controlled Vocabulary</i>), milles on loetletud kõik analüüsitavad üksused, mida uuringus kasutatakse. Vajadusel korrata, et kirjeldada mitu analüüsiüksust.	Metoodika / Statistiline üksus
26?	Analüüsitav	Lühikokkuvõte uuringus analüüsitava	Metoodika /

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
	analüüsiüksus (<i>r:AnalysisUnitsCovered</i>)	analüüsiüksuse kohta. Kasutab elementi <i>InternationalString</i> , et toetada mitut keelt.	Freim
27*	Andmete liik (<i>r:KindOfData</i>)	Kirjeldada lühidalt, millist liiki andmeid dokumenteeritakse uuringu lähtebaasis (<i>LogicalProduct</i>) (toodetes). Näiteks uuringuandmed, rahvaloenduse / loendamise andmed, administratiivsed andmed, mõõtmisandmed, hindamise andmed, demograafilised andmed jne Toetab välise kontrollitud sõnastiku kasutamist	Metoodika / Uuringu liik
28*	Muu materjal (<i>r:OtherMaterial</i>)	Sisaldab asjakohaseid viiteid uuringuüksust (<i>Study Unit</i>) puudutavale muule materjalile kas DDI või muus vormis.	Üldandmed / Muud kokkulepped
29?	Vajalike ressursside paketid (<i>r:RequiredResourcePackages</i>)	Täpsustatakse viidates komponendile <i>ResourcePackages</i> 'ile, mis on nõutav uuringu otsustamiseks. See nimekiri on informatiivne ja aitab luua täielikku metaandmete ülekandmist või luua arhiveerimise pakette. Kasutatakse peamiselt pärast seda, kui versioon (uuringudokument) on suhteliselt stabiilne ja avaldatud.	Puudub, lisada
30*	Piirangud (<i>r:Embargo</i>)	Keelustamise (embargo) info uuringuüksuse kohta. Selles jaos võib üksikutest muutujatest teha viiteid keelustamise infole.	Metoodika / Konfidentsiaals uspoliitika
31*	Kontseptuaalne komponent / Viide kontseptuaalsele komponendile (<i>c:ConceptualComponent</i> / <i>r:ConceptualComponent Reference</i>)	Dokumenteeritakse mõistete sisu ja mõistete, üldkogumi, andmeelementide, geograafilise struktuuri ja geograafilist asukoha relatsiooniline struktuur, mida kasutatakse uuringus / Viide <i>ConceptualComponent</i> moodulile	Metoodika / Mõisted ja määratlused (kui see ei kata seda osa ära, siis võtta mõistete andmebaasist)

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
32*	Andmekogumine / Viide andmekogumisele (<i>d:DataCollection/r:DataCollectionReference</i>)	Dokumenteeritakse andmekogumistegevused, mida selles uuringus kasutati; andmete kogumise protsessi arendused, küsimused, küsimuste suunamised (konstrueeritud kontrollide abil), andmete kogumise instrument, juhendid, töötlemise lõpptulemused ja juhiseid, metoodika, jne / Viide olemasolevale <i>Data Collection</i> moodulile	Metoodika / Andmete kogumine
33*	Andmete kirjeldus / Viide andmete kirjeldusele (<i>l:BaseLogicalProduct /r:LogicalProductReference</i>)	Dokumenteeritakse uuringus produtseeritud andmete loogiline (mõistuspärane) sisu. NB! <i>l:BaseLogicalProduct</i> on asendusgrupi peaelement ja asendatakse sobiva asendusgrupi liikmega / Viide olemasolevale <i>Logical Product</i> moodulile	Metoodika / Andmete kirjeldus
34*	Tehniline tunnus / Viide tehnilisele tunnusele (<i>p:PhysicalDataProduct/r:PhysicalDataProductReference</i>)	Dokumenteeritakse uuringuga esitatud andmete füüsiline struktuur / Viide olemasolevale <i>Physical Data Product</i> moodulile.	Tegemist on iMetas tehnilise tunnusega, mida kasutatakse lähteandebaasi loomisel
35*	Füüsiline dokument / Viide füüsilisele dokumendile (<i>pi:PhysicalInstance /r:PhysicalInstanceReference</i>)	Sisaldab infot füüsilise andmepaketi (<i>Data Product</i>) eksemplari (tegelike andmefailide) kohta. See on moodulis <i>PhysicalDataProduct</i> sisalduv täielik dokumentatsioon, mis on konkreetse faili spetsiifiline ja toimib välise andmefaili kirjeldusena. / Viide olemasolevale <i>Physical Instance</i> moodulile.	Puudub, SA-s on ühesugune andmestruktuur olemas, tegemist on lähtebaasiga
36*	Arhiiv / Viide arhiivile (<i>a:Archive /r:ArchiveReference</i>)	Arhiiv toimib paketina, mis sisaldab nii püsivat kui ka lühiajalist teavet. Kirjeldatakse arhiivispetsiifilist materjali, sh kirjetega	Puudub, lisada

Jrk nr	iMeta atribuut (DDI element)	Selgitus	Olemasolu iMetas / kommentaar
		talletatud informatsiooni kui ka elutsükli teavet. Metaandmed pakettis jagunevad elementideks, mis on püsivad (peaksid alles jääma uuringuüksusega kogu tema elutsükli jooksul) ja lühiajalised (asjakohased üksnes uuringuüksuse jaoks nii nagu oleks kuuluksid need arhiivi).	
37*	DDI profiil / Viide DDI profiile (<i>pr:DDIProfile / r:DDIProfileReference</i>)	Sisaldab DDI objekte, mida on kasutatud konkreetses uuringus.	Puudub, lisada

Kokkuvõte: komponendi *Study Unit* elemente suuremas osas metasüsteemis iMeta dokumenteeritakse, vaid moodulit "Statistikatöö" oleks vaja täiendada seitsme elemendi võrra.

Lisa 2 DDI mooduli *LogicalProduct* kaardistus

Tabel 6 Mooduli *LogicalProduct* kaardistus metasüsteemiga iMeta

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
1*	Lähtebaasi nimi (<i>LogicalProductName</i>)	Lähtebaasi nimi, mida võib väljendada erinevates keeltes. Võib elementi korrata, et väljendada näiteks erinevaid nimetusi erinevate süsteemide jaoks.	Tehnilised tunnused / Tabeli nimi
2*	Lühinimi (<i>r:Label</i>)	Lähtebaasi väljakuvatav nimetus. Võib väljendada erinevates keeltes ja näiteks erineva pikkuse piiranguga nimetuste korral.	Tehnilised tunnused / Tabeli nimi
3?	Kirjeldus (<i>r:Description</i>)	Lähtebaasi sisu kirjeldus ja eesmärk. Võib väljendada mitmes keeles ja toetab struktureeritud sisu kasutamist.	Tehnilised tunnused / Kommentaar
4?	Kaetus (<i>r:Coverage</i>)	Kirjeldab ajalist, ruumilist ja aktuaalset kaetust.	Kasutada statistikatöö juures olevat infot (Mooduli <i>StudyUnit</i> p 24)
5*	Andmete seosed / Viide andmete seostele (<i>DataRelationship / r:DataRelationshipReference</i>)	Sisaldab sõnalist kirjeldust, kuidas faili loogiline sisu on üksteisega seotud programmeerimise eesmärgil. Näiteks, märkides, et on olemas leibkond, perekond ja isik, kus leibkond on tähistatud muutujaga, unikaalne pere on muutuja_H ja muutuja_F kokkuliidetuna ja unikaalne isik leibkonnas muutuja_H ja muutuja_P kokkuliidetuna. / Viide moodulile <i>DataRelationship</i>	Olemas lähteandmebaasis

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
6*	Muu materjal (<i>r:OtherMaterial</i>)	Kirjeldatakse muud materjalid, millel on konkreetne seos lähtebaasiga. Näiteks trükitud andmete sõnastik või lähtebaasi trükitud väljund (publikatsioon).	Puudub, esialgu ei lisa
7*	Kategooria skeem / Viide kategooria skeemile (<i>CategoryScheme /r:CategorySchemeReference</i>)	Sisaldab kategooriate kirjeldusi, mida kasutatakse küsimuste vastustes ja andmete kirjelduses. Nendevaheline seos ja koodi väärtused on kirjeldatud koodiloendi skeemis / Viide eelnevalt kirjeldatud kategooriate skeemi sisemisele või välisele XML dokumendile.	Klassifikaator / Kirjeldus – iMetas ei halda eraldi kategooriatena
8*	Koodiloendi skeem / Viide koodiloendi skeemile (<i>CodeListScheme /r:CodeListSchemeReference</i>)	Loetletakse kasutatavad koodide väärtused, mis esindavad kategooriaid muutuja või küsimuste jaoks. Samuti kirjeldatakse alluvussuhted kategooriate vahel, kasutades koodskeemi / Viide eelnevalt määratletud koodiskeemi sisemisele või välisele DDI dokumendile.	Klassifikaator / Kood, tase - vastab osaliselt
9*	Uuringus kasutatud muutuja skeem / Viide kasutatud muutuja skeemile (<i>RepresentedVariableScheme/r:RepresentedVariableSchemeReference</i>)	Uuringus kasutatud muutujate skeem on muutujate korduvalt kasutatavate osade alus (tuumik). (<i>RepresentedVariable</i> vastab GSIM'is <i>Represented Variable</i> 'le.)	Statistikatöö / registrimuutuja
10*	Muutuja skeem / Viide muutuja skeemile (<i>VariableScheme /r:VariableSchemeReference</i>)	Sisaldab muutujate ja muutujagrupi valikuid / Viide eelnevalt kirjeldatud muutujaskeemile kas sisese või välise DDI dokumendi jaoks.	Sisaldub muutuja mudelis
11*	Kuubi skeem / Viide	Sisaldab <i>NCubes</i> ja <i>Ncube</i> gruppide	Sisaldub muutuja

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
	kuubi skeemile (<i>NCubeScheme / r:NCubeSchemeReference</i>)	valikuid. / Viide varem kirjeldatud <i>Ncube</i> skeem kas sisese või välise asjakohase DDI versiooni jaoks.	mudelis

Koodiloendi skeem (*CodeListScheme*)

Skeem, mis sisaldab koodiloendeid, mida kasutatakse, et määrata kindlaks kood vastavalt väärtusele ja vastuste valdkonnale.

Tabel 7 Koodiloendi skeemi võrdlus metasüsteemiga iMeta

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
1*	Koodiloendi skeemi nimi (<i>CodeListSchemeName</i>)	Koodiloendi skeemi nimi. Võib väljendada erinevates keeltes, Korrates elementi saab väljendada erineva sisuga nimesid, näiteks erinevad nimetused erinevatele süsteemidele.	Klassifikaatorid / Klassifikaatori nimi
2*	Silt (<i>r:Label</i>)	Koodiloendi skeemi kuvanimi. Võib väljendada mitmes keeles. Korrates elementi saab väljendada erineva sisuga siltide jaoks, nt sildid erineva pikkuse piiranguga.	Klassifikaatorid / Lühinimi
3?	Kirjeldus (<i>r:Description</i>)	Koodiloendi skeemi sisu kirjeldus ja eesmärk. Võib väljendada mitmes keeles ja toetab struktureeritud sisu kasutamist.	Klassifikaatorid / Kirjeldus
4*	Koodiloendi skeemi viide (<i>r:CodeListSchemeReference</i>)	Võimaldab lisada viite teisele koodiloendi skeemile.	Puudub, XML-faili genereerimisel arvestada klassifikaatori nimega
5*	Koodiloend / Viide koodiloendile (<i>CodeList /</i>	Struktuur, mida kasutatakse, et siduda koodiloendi väärtused täpselt kindlaks määratud kategooriatega. Võib olla lineaarne	Klassifikaatorid / Klassifikaator, tegemist nn klassifikaatori

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
	<i>r:CodeListReference</i>)	või hierarhiline. / Viide olemasolevale koodiloendile lisatakse viidates koodiloendi skeemile.	"katusega" Neuchâтели TM mudeli järgi
6*	Koodiloendi grupp / Viide koodiloendi grupile (<i>CodeListGroup</i> / <i>CodeListGroupReference</i>)	Koodiloendite grupeerimine kontseptuaalsetel või halduslikel eesmärkidel. / Viide olemasolevale koodiloendi grupile.	Klassifikaatorite valdkond

Koodiloend (*CodeList*)

Koodiloend on struktuur, mida kasutatakse, et siduda koodiväärtuste loetelu kindlaks määratud kategooriatega. Koodiloend võib olla lineaarne või hierarhiline.

Tabel 8 Koodiloendi võrdlus metasüsteemiga iMeta

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
1*	Koodiloendi nimi (<i>CodeListName</i>)	Koodiloendi nimi. Võib väljendada erinevates keeltes. Korda elementi, et väljendada erineva sisuga nimesid, näiteks erinevad nimetused erinevatele süsteemidele	Klassifikaatorid / klassifikaatori versiooni nimi
2*	Silt (<i>r:Label</i>)	Koodiloendi kuvatav silt. Võib väljendada erinevates keeltes. Korda erineva sisuga silte, näiteks silt erineva pikkuse piiranguga	Klassifikaatorid / klassifikaatori versiooni lühinimi
3?	Kirjeldus (<i>r:Description</i>)	Koodiloendi sisu kirjeldus ja eesmärk. Võib väljendada mitmes keeles ja toetab struktureeritud sisu kasutamist.	Klassifikaatorid / klassifikaatori versiooni kirjeldus
4?	Soovitav andmetüüp (<i>r:RecommendedDataType</i>)	Andmetüüp, mida andmete esitaja soovib. Toetab kontrollitud sõnavara kasutamist, mida DDI soovibki kasutada.	Puudub, ei lisa esialgu

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
5*	Viide koodiloendile (<i>r:CodeListReference</i>)	Võimaldab kaasata teist koodiloendit viidates sellele. Hoolitseda tuleb selle eest, et oleks tagatud koodiloendi kõigi koodide unikaalsed väärtused.	Puudub, ei lisa, SA infosüsteem ei võimalda
6?	Viide kategooria skeemile (<i>r:CategorySchemeReference</i>)	Viide kategooria skeemile, eeldades, et kõik viidatud kategooriad järgnevate koodidega on osa sellest, v.a juhul, kui skeemi viide kirjutatakse üle <i>CategoryReference</i> 'i koodi väljal.	Puudub, ei lisa esialgu
7?	Hierarhia tüüp (<i>HierarchyType</i>)	Hierarhilise koodiloendi korral defineeritakse hierarhia tüüp: regulaarne või ebaregulaarne.	Puudub, ei lisa, võimalik kirjeldada klassifikaatori kirjelduses
8*	Tase (<i>Level</i>)	Kasutatakse, et kirjeldada koodiloendi hierarhia tasemeid	Klassifikaatorid / klassifikaatori elemendi tase
9*	Kood (<i>Code</i>)	Struktuur, mis ühendab koodi unikaalse väärtuse kindla kategooriaga	Klassifikaatorid / klassifikaatori elemendi kood

Koodiloendi grupp (*CodeListGroup*)

Koodiloendite rühmitamine kontseptuaalsel või administratiivsel eesmärgil. Võib olla hierarhiline. Metasüsteemis vastab sellele klassifikaatori valdkond.

Tabel 9 Koodiloendi grupi võrdlus metasüsteemiga iMeta

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
1?	Koodiloendi grupi tüüp (<i>TypeOfCodeListGroup</i>)	Koodiloendi grupi tüübi lühike tekstiline identifitseerimine. Toetab välise kontrollitud sõnavara kasutamist	Klassifikaatorite valdkonna kirjeldus

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
2?	Koodiloendi grupi nimi (<i>CodeListGroupName</i>)	Koodiloendi grupi nimi. Võib väljendada mitmes keeles. Korda elementi, et väljendada erineva sisuga nimesid, näiteks erinevad nimetused erinevatele süsteemidele	Klassifikaatori valdkonna nimetus
3*	Silt (<i>r:Label</i>)	Koodiloendi grupi kuvanimi. Võib väljendada erinevates keeltes. Korrata elementi erineva pikkusega siltide piirangute korral.	Klassifikaatorite valdkonna nimetus
4*	Kirjeldus (<i>r:Description</i>)	Koodiloendi grupi sisu kirjeldus ja eesmärk. Võib väljendada mitmes keeles ja toetab struktureeritud sisu kasutamist.	Klassifikaatorite valdkonna nimetus
5?	Viide üldkogumile (<i>r:UniverseReference</i>)	Viide üldkogumi selgitusele, mis kirjeldab isikuid või muid objekte, mille kohta koodiloendi grupi sisu käib.	Puudub, ei lisa esialgu
6*	Viide mõistele (<i>r:ConceptReference</i>)	Viide mõistele, mida koodiloendiga väljendatakse selles grupis.	Puudub, lisada seos mõistega, kui mõiste on olemas
7?	Teema (<i>r:Subject</i>)	Teema kasutamine grupi tasemel võimaldab seostada objekte teemapõhistena või identifitseerida taaskasutatava grupiobjektide teema tunnuseid.	Puudub, ei lisa esialgu
8*	Võtmesõna (<i>r:Keyword</i>)	Kui selle rühma jaoks on märksõnad loetletud, siis on äärmiselt soovitatav, et loetletud märksõnad leiduksid ka ülempaketi <i>TopicalCoverage</i> elemendis, kui see grupp on lisatud otse või viitega moodulisse, mis sisaldab kirjeldavat elementi.	Puudub, ei lisa esialgu
9*	Koodiloendi viide / Koodiloendi grupi viide (<i>r:CodeListReference /</i>	Viide koodiloendi koostisele / Viide koodiloendi grupi koostisele. See võimaldab koodiloendi grupi pesastamist.	Puudub, XML failis arvestada klassifikaatori versiooni nimetusega /

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
	<i>CodeListGroupReference</i>)		klassifikaatori valdkonna nimetusega

Kategooria (*Category*)

Konkreetses kategooria või vastuse kirjeldus. OECD statistika mõistete sõnastikust: objekti üldmõiste klassifikaatori igal tasandil, tavaliselt kategooriate, sektsioonide, alamsektsioonide, divisjonide, allüksuste, gruppide, alagruppide, liikide ja alaliikide lühiseletus. Lisaks standardi nimele, sildile ja kirjeldusele võib kategooria sisaldada viidet mõiste määratlusele, anda teavet selle kohta, kuidas kategooria tekitati ja märkida, kui see on puuduva väärtuse tüüp.

Metasüsteemis iMeta ei hallata eraldi kategooriat, küll on kategooriad klassifikaatori elemendina dokumenteeritud.

Tabel 10 Kategooria võrdlus metasüsteemiga iMeta

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
1*	Kategooria nimi (<i>CategoryName</i>)	Kategooria nimi. Võib väljendada erinevates keeltes. Võib korrata elementi, kui vaja väljendada nimesid erinevalt, näiteks erinevad nimetused erinevatele süsteemidele	Klassifikaatori elemendi nimetus
2*	Silt (<i>r:Label</i>)	Kuvatav silt kategooriale. Võib väljendada mitmes keeles. Võib korrata elementi erinevate siltide jaoks, näiteks sildid erineva pikkuse piirangutega või erinevate rakenduste jaoks	Klassifikaatori elemendi lühinimi
3?	Kirjeldus (<i>r:Description</i>)	Kategooria sisu ja eesmärgi kirjeldus. Võib väljendada mitmes keeles ja toetab struktureeritud sisu kasutamist. NB! kategooriate sisu võrdlus toimub sisu	Klassifikaatori elemendi kirjeldus

Jrk nr	Atribuut (element)	Atribuudi selgitus	iMeta grupp ja kommentaar
		kirjeldust kasutades.	
4?	Viide mõistele (<i>r:ConceptReference</i>)	Viide defineeritavale mõistele.	Puudub, lisada seos mõistega, kui mõiste on olemas
5?	Põlvnemine (<i>Generation</i>)	Kirjeldab, millest kategooria tuleneb või on moodustatud. Protsess kategooriate koosseisu määramise kirjeldamiseks.	Puudub, esialgu ei lisa
6*	Viide alamkategooriale (<i>SubCategoryReference</i>)	Viide ühele või mitmele kategooriale, mille suhtes olemasolev kategooria omab laiemat definitsiooni. Võimaldab viidata kitsamale kategooriale ja võimaldab defineerida spetsiaalselt või osalist suhet.	Olemas tase, hierarhilisel klassifikaatoril viide tuletatav tasemetest

Kokkuvõte: vaadeldud osas mooduli *LogicalProduct* e lähtebaasi kohta käivast infost on metasüsteemis olemas lähtebaasi nimi, mida ühtlasi ka kuvatakse välja, ja kommentaar selle kohta. Esialgu metasüsteemi täiendamiseks DDI elementidega detailsemalt vajadused puuduvad.

Koodiloendi skeemist puudub ainult üks DDI standardi element. XML-faili genereerimisel kasutada klassifikaatori nime.

Koodiloendi grupist ja kategooriast valis töö autor esialgu metasüsteemi täiendamiseks ühe elemendi.

Koodiloendist on puuduolevaid elemente neli, kuid töö autor ei pea vajalikuks ka neid praegu metasüsteemi lisada.

Lisa 3 Sõnastik: DDI standardi mõisted, sh vasted metasüsteemi iMeta mõistetega

DDI standardis kasutatav terminoloogia (inglise keeles) ei vasta alati seni SA-s kasutatava inglisekeelse terminoloogiaga. Seetõttu on mõistlik olulisemad mõisted välja tuua koos vastava SA-s kasutatava eestikeelse mõistega, et edaspidi lihtsustada kasutajatel DDI standardi dokumentatsioonist arusaamist.

Tabel 11 DDI standardi mõistete vasted metasüsteemi iMeta mõistetega

Mõiste DDI standardis	Vaste, sh metasüsteemis iMeta
Analysis Unit	Analüüsiüksus
Archive	Arhiiv
BaseRecordLayout	Alusdokumendi struktuur
Budget Document	Eelarvedokument
Category	Kategooria (iMeta: klassifikaatori element)
Category Scheme	Kategooria skeem
Category Group	Kategooriate grupp
Code	Kood
Code List	Koodiloend (iMeta: klassifikaator)
Code List Scheme	Koodiloendi skeem
Code List Group	Koodiloendi grupp (iMeta: klassifikaatori valdkond)
Code Scheme	Koodi skeem
Code Item	Koodi element
Comparative	Võrdlusinfo
Concept	Mõiste
Concept Scheme	Mõistete skeem

Conceptual Component	Kontseptuaalne komponent
Conceptual Variable	Kontseptuaalne muutuja e mõiste
Contextual Variable	Töö konteksti põhine muutuja
Control Construct Scheme	Küsimuste kontrolliskeem
Data Collection	Andmekogumine
Data File	Andmefail
Data Product	Andmepakett
Data Set	Andmekogum
DDI Modules	DDI moodulid
DDI Profile	DDI profiil
Element	Element (iMeta: atribuut)
Geographic Location Scheme	Geograafilise asukoha skeem
Geographic Structure Scheme	Geograafilise struktuuri skeem
Inline Ncube	Sisene andmekuup
Instance	(Asjaomane) versioon, instants
Interviewer Instruction Scheme	Küsitaja instruksioonid
Item Value	Näitaja
Logical Product	Lähtebaas
LogicalRecord	Andmeelementide kogum (lähtebaasis)
Managed Representation Scheme	Hallatavad esindusskeemid
Ncube	Andmekuup
NCube Logical Product	Kuubi kirjeldus
NCube Scheme	Kuubi skeem
NCube Structure	Kuubi skeem – komponendid: dimensioon, atribuut, mõõtühik (<i>measure</i>)

Organization Scheme	Organisatsiooni skeem
Physical Data Product	Füüsilise andmefaili kirjeldus
Physical Instance	Füüsiline (tegelik) andmefail
Physical Structure Scheme	Andmete füüsiline struktuur
Proprietary	Omandiõigus
Question Scheme	Küsimuste skeem
Record Group	Dokumentide kogum
Record Item	Dokumendi kirje
Record Layout	Dokumendi struktuur
Record Layout Scheme	Andmebaasi struktuur (tehniline tunnus)
Represented Variable	Uuringus kasutatud muutuja
Represented Variable Scheme	Uuringus kasutatud muutuja skeem
Response Domain	Loendite valdkond
Reusable	Korduvalt kasutatav
Study Unit	Uuring (iMeta: statistikatöö versioon)
Tabular Ncube	Tabelikujuline lähtebaasi kirjeldus
Universe Scheme	Üldkogumi skeem
Value Location	Väärtuse asukoht failis (veeru ja rea ristumiskoht)
Variable	Muutuja
Variable Scheme	Muutuja skeem
VariableRepresentation	Muutuja uuringus (muutuja statistikatöös)
XML Schemas	XML nimeruum (ka XML skeemid)