

Ontoloogiade loomise meetoodika.

Versioon Ver 4
21.10.2011

Hele-Mai Haav
Küberneetika Instituut

1. *Metoodika eesmärk ja objekt.*

Käesolevas dokumendis esitatakse riigi infosüsteemi (IS) valdkonnaontoloogiade loomise meetoodika, mille raames käsitletakse selle eesmärke, objekti ja esitatakse soovituslik protsesside jada ontoloogiade arendamiseks. Käesolevas dokumendis esitatud meetoodika on aluseks dokumendis „Nõuded RIHA ontoloogiatele“¹ toodud nõuetele vastavate ontoloogiade loomisele.

Metoodika eesmärgiks on pakkuda meetodeid ja tehnikaid riigi infosüsteemis kasutatavate andmete semantika (struktuuri ja tähenduse) *kooskõlastatud* kirjeldamiseks kasutades ontoloogiaid esitatuna OWL keeles.

Metoodika objektiks on andmekogu andmeobjektide ja andmeteenuste kirjelduste semantiliseks rikastamiseks vajalike ontoloogiade loomine. Tegelikult moodustavad riigi ISi valdkonnaontoloogiad ontoloogiade võrgustiku, milles eri valdkondade ontoloogiad on omavahel seotud kas viidete, modulaarsuse, versioonide või muude seoste abil. Selline ontoloogiade võrgustik seab vastavad nõuded ontoloogiade loomise meetoodikale. Ontoloogiade haldamise osa muutub sel juhul märksa keerulisemaks kui iga eraldiseisva ontoloogia korral.

Ontoloogiade loomis- ja arendusprotsessid on keerulised ja nende läbiviimiseks on olemas mitmeid ontoloogiade loomise üldisi metodoloogiaid. Ontoloogiade loomise metodoloogiade vaatlus näitab, et pole olemas ühte standardset ontoloogiade loomise ja arenduse meetoodikat (Gruninger & Fox, ²1995; Uschold & Gruninger, 1996;³ Jones et al., 1998;⁴ Fernandez-Lopez, 1999;⁵ Corcho et al., 2003⁶).

¹ H-M. Haav, Nõuded RIHA ontoloogiatele, Tallinn, 2010

² Gruninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. *IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Quebec, Canada.

³ Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 11(2), 93-155.

⁴ Jones, D., Bench-Copon, T., & Visser, P. (1998). Methodologies for ontology development. In: J. Cuenca (Ed.), *Proceedings of ITi and KNOWS Conference of the 15th IFIP World Computer Congress*, Budapest, August 1998, pp. 62-75. Retrieved November 20, 2004, from <http://www.iet.com/Projects/RKF/SME/methodologies-for-ontology-development.pdf>

⁵ Fernandez-Lopez, M. (1999). Overview of methodologies for building ontologies. In: *Proceedings of the IJCAI-99 workshop on ontologies and problem-solving methods*, Stockholm, Sweden. Retrieved November 20, 2004, from <http://www.lsi.upc.es/~bejar/aia/aia-web/4-fernandez.pdf>

⁶ Corcho, O., Fernandez-Lopez, M., & Gumez-Perez, A. (2003). Methodologies, tools and languages for building ontologies: where is their meeting point? *Data & Knowledge Engineering*, 46, 41-64.

Hea ülevaate ja hinnangu erinevatele metoodikatele võib leida Gómez-Pérez et al. raamatust⁷.

Seega tuleb käesolevat metoodikat vaadelda kui ontoloogiate loomise üldiste meetodite kohandamist riigi ISi semantilise koosvõime tagamise vajadustele, nõuetele ja spetsiifikale. Metoodika loomisel on üldistatud ontoloogiate loomise koolitustel kasutatud metoodikaid ja arvestatud riigi ISi esindajate kui koolitatavate tagasisidega.

2. Ontoloogiate arendusprotsess.

Ontoloogiate arendusprotsess on protsess, mille käigus kasutaja vajadused teisendatakse ontoloogia või ontoloogiate kujule. Selle protsessi üldiseks eesmärgiks on välja selgitada ja defineerida ontoloogiate arenduseks vajalikud tegevused.

Ontoloogiate arendusprotsess ei määra vastavate tegevuste järjekorda ontoloogiate loomisel. Seda tehakse ontoloogia elutsükli mudeli abil, mis määrab kindlaks millal teatud tegevused tuleb läbi viia.

Elutsükli mudeliteks kohandatakse tavaliselt tarkvaraarenduses levinud mudeleid, mis kirjeldavad tarkvara arendamiseks läbiviidavate tegevuste erinevaid organiseerimise viise. Nendeks on tarkvaraarenduses kasutusel kose, iteratiivne, evolutsioonilise prototüüpimise, ekstreemprogrammeerimise mudel jt.

Näiteks METHONTOLOGY⁷ ontoloogia arenduse raamistik pakub välja 3 gruppi ontoloogiate arenduseks vajalikke tegevusi: haldus, arendus, toetus.

METHONTOLOGY kasutab evolutsioonilise prototüüpimise (evolving prototypes) elutsükli, milles iga uue ontoloogia prototüübi loomine algab ajakava väljatöötamisega, sellele järgneb ontoloogia spetsifikatsiooni välja töötamine ja teised ontoloogia arendustegevused nagu kontseptualiseerimine, formaliseerimine, realiseerimine ja käitlemine. Iga tegevuse juurest võib pöörduda eelmiste tegevuste juurde selleks, et midagi täpsustada või muuta. Nende tegevustega paralleelselt toimuvad haldustegevused (dokumenteerimine, versioneerimine jms) ja toetustegevused (teadmushõive, hindamine, integreerimine jms).

Ontoloogiate arendus on meeskonnatöö. Selle liikmed peaksid esindama kasutajaid, valdkonnaeksperte, teadmus- või ontoloogiainsenere ja vajadusel ka tarkavara arendajaid.

Järgnevalt esitatakse riigi ISi tarbeks loodud ontoloogiate arendusprotsessi põhiliste tegevuste loetelu ja sõnastik. Riigi ISi valdkonnaontoloogiate arendusprotsessi elutsükli kirjeldamisel lähtutakse sellest sõnastikust. Toodud sõnastik on loodud kasutades osaliselt NEON projektis toodud vastavat tegevuste sõnastikku.⁸

3. Arendusprotsessi tegevuste sõnastik.

Mitte-ontoloogiliste ressursside taasarendus (Non Ontological Resource Reengineering) – olemasoleva mitte-ontoloogilise ressursi (sõnastik, andmebaas jms) leidmine ja teisendamine ontoloogiaks.

Mitte-ontoloogiliste ressursside taaskasutus (Non Ontological Resource Reuse) – olemasolevate mitte-ontoloogiliste ressursside (andmebaasid, sõnastikud jms) kaasamine ontoloogia arendusprotsessi.

Ontoloogia annoteerimine (Ontology Annotation) - ontoloogia kirjelduse rikastamine lisainformatsiooniga (näiteks kommentaarid, metaandmed jms).

⁷ Gómez-Pérez, A., Fernández-López, M., Corcho, O., Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Springer, 2004

⁸ NEON projekt, www.neon-project.org

Ontoloogia areng (Ontology Evolution) – ontoloogia muutmine säilitades selle mittevastuolulisuse.

Ontoloogia dokumentatsioon (Ontology Documentation) – ontoloogia loomise käigus genereeritud dokumentide ja selgitavate märkuste kogum. Näiteks, ontoloogia spetsifikatsiooni ja kontseptualisatsiooni dokumendid, ontoloogilised pühendumused (kokkulepped kasutada vastavat ontoloogiat) jms.

Ontoloogia evalueerimine (Ontology Evaluation) – ontoloogia tehnilise kvaliteedi kontrollimine.

Ontoloogia hindamine (Ontology Assessment) – ontoloogia vastavuse kontroll ontoloogia kasutaja nõuetele (näiteks kasutatavus, kasulikkus, abstraktsiooni tase, kvaliteet jms).

Ontoloogia realiseerimine (Ontology Implementation) - ontoloogia esitamine OWL keeles vastavalt selle keele süntaksile.

Ontoloogia integratsioon (Ontology Integration) - mingi ontoloogia kasutamine teise ontoloogia sees.

Ontoloogia joondamine (Ontology Alignment) - ontoloogiate vaheliste vastavuste tuvastamine ja nende kasutamine. Sünonüümiks on ontoloogiate seostamine (Ontology Mapping).

Ontoloogia konfiguratsiooni haldamine (Ontology Configuration Management) – ontoloogia dokumentatsiooni ja OWL keelse esituse kõigi versioonide talletamine ja nende muutuste kontrollimine ning haldamine.

Ontoloogia kontseptualisatsioon (Ontology Conceptualization) – valdkonna teadmushõive käigus saadud informatsiooni (andmete ja teadmiste) organiseerimine ja struktureerimine valdkonna mõistmiseks vajalikeks teadmismudeliteks, mis vastavad ontoloogia spetsifikatsiooni dokumendile. See tegevus ei sõltu ontoloogia realiseerimise viisist.

Ontoloogia kvaliteedi tagamine (Ontology Quality Assurance) – kõigi ontoloogia arendusprotsessi tegevuste ja tulemite rahuldava kvaliteedi kindlustamine.

Ontoloogia mestimine (Ontology Merging) – uue ontoloogia loomine kahest või enamast ülekattuvast lähteontoloogiast või ontoloogia moodulist.

Ontoloogia modulariseerimine (Ontology Modularization) – suurema ontoloogia jagamine mooduliteks selle taaskasutamise ja käitlemise hõlbustamiseks.

Ontoloogia muutmine (Ontology Modification) – ontoloogia muutmine arvestamata ontoloogia mittevastuolulisuse nõuet.

Ontoloogia pöördarendus (Ontology Reverse Engineering) - Ontoloogia võimaliku kontseptuaalse mudeli tuletamine ontoloogia realisatsiooni põhjal.

Ontoloogia sobitamine (Ontology Matching) – erinevate ontoloogiate komponentide vaheliste seoste ja sõltuvuste leidmine.

Ontoloogia spetsialisatsioon (Ontology Specialization)- ontoloogia laiendamine sügavuti.

Ontoloogia spetsifikatsioon (Ontology Specification) – nõuete kogum, millele loodav ontoloogia peaks vastama.

Ontoloogia taasarendus (Ontology Reengineering) – olemasoleva realiseeritud ontoloogia kontseptuaalse mudeli parendamine ja selle taasrealiseerimine.

Ontoloogia taaskasutus (Ontology Reuse) – olemasoleva ontoloogia või ontoloogia mooduli kasutamine mingi teise probleemi lahendamisel. Näiteks, olemasoleva ontoloogia kasutamine uue ontoloogia loomisel, mingi teise ontoloogiapõhise rakenduse loomisel, ontoloogiate joondamisel jms.

Ontoloogia valideerimine (Ontology Validation) – ontoloogia definitsioonide tähenduste ja kontseptualiseeritava valdkonna mudeli vastavuse hindamine. Kas ontoloogia on vastavuses valdkonnaga? Kas me loome õiget ontoloogiat?

Ontoloogia verifitseerimine (Ontology Verification) – ontoloogia ja tema spetsifikatsiooni vastavuse kontrollimine. Kas ontoloogia vastab spetsifikatsioonis esitatud nõuetele? Kas me loome ontoloogiast õigesti?

Ontoloogia versioneerimine (Ontology Versioning) – ontoloogia muutuste käsitlemine luues ja hallates erinevaid ontoloogia versioone.

Ontoloogia võrdlus (Ontology Comparison) – erinevuste leidmine kahe või enama ontoloogia või ontoloogia mooduli vahel.

Ontoloogia õppimine (Ontology Learning) – teadmushõive meetod, mis baseerub mittestruktureeritud (näiteks tekst), osaliselt struktureeritud (näiteks HTML lehed) ja struktureeritud (näiteks andmebaasid) andmeallikate kontseptuaalseteks struktuurideks või mudeliteks teisendamise (pool)automaatsetel meetoditel.

Ontoloogia ümberstruktureerimine (Ontology Restructuring) – ontoloogia algses kontseptuaalses mudelis sisalduva teadmuse korrigeerimine ja reorganiseerimine ning puuduva teadmuse kindlaks tegemine.

4. Riigi ISi valdkonnaontoloogiade arendusprotsess ja elutsükli mudelid.

Käesolev metoodika esitab riigi ISi valdkonnaontoloogiade arendusprotsessi vastavalt riigi erinevate ISide valdajate ja arendajate käsutuses olevatele taaskasutatavatele ressursidele. Näiteks võivad sellisteks ressursideks olla andmekogu loomise määrad, andmebaasi kontseptuaalne skeem, andmekoosseisu esitus xmi failina, andmeteenuste kirjeldused jms.

Sobivaks ontoloogiade elutsükli mudeliks iga üksiku valdkonna ontoloogia loomisel on valitud iteratiivne mudel, mille korral hakatakse ontoloogiad looma lähtudes osalisest nõuetele vastamisest. Iga järgmise iteratsiooni korral täiustatakse ontoloogiad nii, et lõpuks on kõik nõuded täidetud. Seda elutsükli mudelit on edukalt kasutatud äriregistri ontoloogia⁹, Registrate ja Infosüsteemide Keskuse (RIK) ning Maa-ameti valdkonna ontoloogiade loomisel.

Riigi ISi valdkonnaontoloogiade võrgustiku kui terviku loomisel lähtutakse evolutsioonilisest prototüüpimise mudelist, mille omapäraks on nõuetele vastava osalise tulemi loomine st riigi ISide korral osalise ontoloogiade võrgustiku loomine. Saadud ontoloogiade võrgustiku prototüüpi hinnatakse nõuete suhtes. Esialgne prototüüp võimaldab ka täpsustada kasutajate võimalikke ebaselgeid nõudeid.

Ontoloogiade arendusprotsessi esitamisel tuuakse välja kasutatavad ressursid (sisend), tegevuste jada ja tulem.

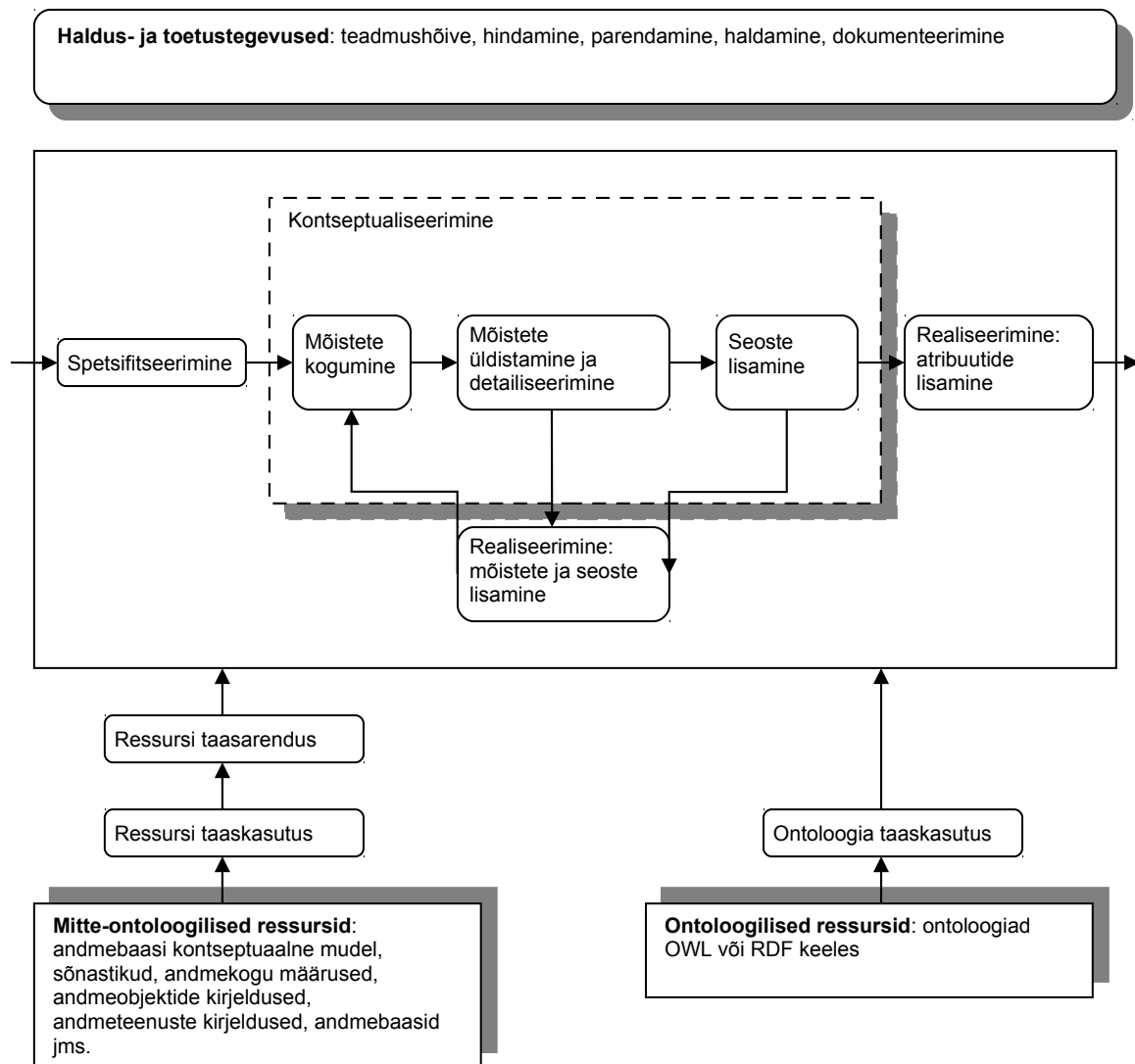
4.1. Kasutatavad ressursid.

Kasutatavateks ressursideks on üldised ja spetsiifilised valdkonna teadmised, mitte-ontoloogilised ja ontoloogilised ressursid. Mitte-ontoloogilised ressursid on andmebaasi kontseptuaalne mudel, sõnastikud, andmekogu põhimäärus vm andmekogu õiguslikult reguleerivad õigusaktid, andmeobjektide kirjeldused, andmeteenuste kirjeldused, andmebaasid jms. Ontoloogilised ressursid on ontoloogiade kirjeldused, mis paiknevad ontoloogiade teekides või repositooriumides. Reeglina on nendele juurdepääs läbi veebi. Näiteks, RIHAs juba paiknevad riigi ISide ontoloogiad võivad olla aluseks mingi uue ontoloogia loomisel.

⁹Leive Turi, Ontoloogia arendamine Eesti Äriregistri näitel, magistritöö, Tallinn, 2010
Äriregistri ontoloogia RIHAs: <https://riha.eesti.ee/riha/main/val/ariregister/1>.

4.2. Tegevuste jada.

Ontoloogiate loomise tegevuste jada on esitatud joonisel 1. Seal on toodud haldus- ja toetustegevused, mis toimuvad paralleelselt arendusprotsessi tegevustega kogu arenduse jooksul. Joonisel on esitatud ka mitte-ontoloogilised ja ontoloogilised ressursid ning nende seos arendusprotsessiga läbi vastavate tegevuste.



Joonis 1. Riigi ISi valdkonnaontoloogiate loomise protsess

Ontoloogiate arendusprotsessid on joonisel 1 näidatud nende järgnevuses moodustades järgmise jada: spetsifitseerimine, kontseptualiseerimine ja realiseerimine. Vastavalt eespool kirjeldatud iteratiivsele elutsükli mudelile toimub varane kontseptualisatsiooni realiseerimine. Esimesel iteratsioonil keskendutakse ontoloogia põhimõistetele ja nendevahelistele seostele, järgmise iteratsiooni korral lisatakse uusi mõisteid ja lõpuks lisatakse mõistete indiviidide atribuudid. Seejärel kontrollitakse kas ontoloogia komponendid katavad andmeteenuste sisend/väljundparameetrid. Kui ei, siis lisatakse vastavad ontoloogia komponendid kuni kõik nõuded (vt Nõuded RIHA ontoloogiatele) on rahuldatud.

Järgnevalt esitatakse iga arendustegevuse täpsem kirjeldus kasutades punktis 3 toodud sõnastikus esitatud tegevusi.

4.2.1. Teadmushõive.

Ontoloogiate loomine algab valdkonna teadmushõivega. Kuigi teadmushõive toimub kogu ontoloogia arenduse jooksul, tuleb enamus teavet hankida protsessi alguses, vajaliku teabe maht kahaneb protsessi jätkudes.

4.2.2. Ressursside taaskasutus.

Nagu eespool mainitud jagunevad ressursid mitte-ontoloogilisteks ja ontoloogilisteks. Riigi ISi valdkonnaontoloogiate loomisel on võimalik kasutada mõlemaid ressursse kogu ontoloogia arendusprotsessi jooksul. Kõige vajalikum on mitte-ontoloogiliste ressursside taaskasutus ontoloogia spetsifitseerimise ja kontseptualiseerimise etapil. Ontoloogiliste ressursside taaskasutus on mõeldav ka realiseerimise käigus (näiteks ontoloogia import). Vastavalt ressursi tüübile toimuvad ka tema taaskasutuse tegevused.

Mitte-ontoloogiliste ressursside puhul läbitakse ressursi taaskasutus ja seejärel selle taasarendus. Näiteks, mingi ISi domeeni mudelit saab edukalt kasutada ontoloogiasse kuuluvate klasside määramiseks. Ontoloogia indiviidide andmetüübiomaduste (datatype properties) määratlemiseks saab kasutada xmi faili kujul esitatud andmeobjektide kirjeldusi.

Ressursi taaskasutuse käigus analüüsitakse ressursi (andmebaasi mudel, andmebaas jms) kasutatavust ontoloogia loomisel. Näiteks on xmi faili kujul esitatud andmeobjektide kirjeldustest otsene kasu riigi ISide ontoloogiate loomisel. Kui on otsustatud, millised ressursid võiksid olla kasulikud ontoloogia loomisel, siis tuleb läbi viia vastava ressursi taasarendus. Selle protsessi käigus teisendatakse mitte-ontoloogiline ressurss ontoloogiaks, mida saaks taaskasutada loodava ontoloogia arendusprotsessis.

Ontoloogiliste ressursside puhul toimub mingi sobiva ontoloogia taaskasutus. Näiteks, kui on juba RIKis tehtud kinnistusraamatu ontoloogia, siis võib juhtuda, et seda saaks kas osaliselt või täielikult taaskasutada Maa-ameti maakatastri ontoloogia loomisel. Teine näide võiks olla geomeetria ontoloogia, mis on ühine paljudele Maa-ameti valdkonna ontoloogiatele.

Ontoloogia taaskasutuse võib jagada järgmisteks alamtegevusteks:

1. Sobivate nõuetele vastavate kandidaatontoloogiate otsing üldistest ontoloogia teekidest ja repositooriumidest nagu näiteks Swoogle¹⁰ ja WATSON¹¹. Riigi ISi valdkonnaontoloogiaid saab otsida RIHAst.
2. Leidud kandidaatontoloogiate hindamine nõuete seisukohast (ontoloogia sisu, keel, formaalne keel, dokumentatsioon, kvaliteet jms).
3. Sobiva ontoloogia valik.
4. Ontoloogia taaskasutuse viisi otsustamine järgmiste valikute seast:
 - Ontoloogiat taaskasutatakse sellisel kujul nagu ta on.
 - Ontoloogia nõuab taasarendust ja alles pärast seda saab teda taaskasutada.
 - Ontoloogiad mestitakse uueks ontoloogiaks.

4.2.3. Spetsifitseerimine.

Paralleelselt teadmushõivega tuleb kindlaks teha nõuded, millele ontoloogia peab vastama. Nõuded esitatakse ontoloogia spetsifikatsiooni dokumendis. Lisas 1 on toodud näide taolisest dokumendist. Spetsifitseerimise etapis määratakse ontoloogia eesmärk ja skoop, kirjeldatakse

¹⁰ Swoogle; <http://swoogle.umbc.edu/>

¹¹ Watson; <http://watson.kmi.open.ac.uk/WatsonWUI/>

kasutusvõimalusi, andmeallikaid ja esitatakse tähtsamate mõistete loetelu. Arendusetapis aitab spetsifikatsioon kitsendada ülesannet ning edaspidi annab kolmandatele isikutele esmase ülevaate ontoloogiast.

4.2.4. Kontseptualiseerimine ja varajane realiseerimine.

Spetsifitseerimise etapile järgneb ontoloogia kontseptualiseerimise etapp. Kontseptualiseerimine etapis luuakse vastavalt hõivatud teadmiste ja spetsifikatsiooni dokumendile valdkonna mõisteid ning nende seoseid kirjeldav teadmismudel (vt Lisa 2) ontoloogia esitamiseks OWL keeles.

Ontoloogia kontseptualiseerimisele järgneb ontoloogia realiseerimine (keerukate ontoloogiatega puhul ka enne seda formaliseerimine), mille tulemuseks on ontoloogia kirjeldus OWL keeles. Käesolevas metoodikas vaadeldakse kontseptualisatsiooni ja realiseerimist koos, sest kasutatakse iteratiivset elutsükli mudelit.

Kontseptualiseerimiseks saab kasutada mitmeid strateegiaid: ülevalt alla, keskelt üles-alla, alt üles.

- *Ülevalt-alla* strateegia korral alustatakse kõige üldisematest valdkonna mõistetest ja liigitakse spetsiifilisemate mõistete poole. Probleemiks on see, et võib juhtuda, et ontoloogia sisaldab palju mittevajalikke ülemise taseme mõisteid või tekib liiga palju kasutuid abstraktsioonitasemeid.
- *Keskelt üles-alla* strateegia korral käsitletakse alguses ainult valdkonna põhilisi mõisteid, hiljem liigitakse nii nende üldistamise kui spetsialiseerimise (täpsustamise) suunas. See meetod sobib nii alama taseme kui ka ülema taseme ontoloogiatega loomiseks, sest ontoloogia ei lähe liiga detailseks ega teki mittevajalikke üldisi mõisteid.
- *Alt-üles* lähenemise korral toimub mingi rakendusega seotud andmeväljade tähenduste üldistamine. Positiivne aspekt on see, et loodavad mõisted kirjeldavad vaadeldavaid rakendusi suhteliselt täpselt. Negatiivne pool on see, et meetod on üsna töömahukas ja raske on leida alamas taseme mõistete tähendusi katvaid üldisi mõisteid, detailsuse aste võib liiga suureks minna, ka on risk luua vastuolulisi ontoloogiaid. Samuti võib tekkida ontoloogia, mida järgmiste teenuste/rakenduste korral on raske kui mitte võimatu kasutada, sest ta on liiga spetsiifiline.

Võttes arvesse, et riigi ISi valdkonnaontoloogiaid hakkavad tegema valdkonda tundvad inimesed (valdkonnaekspertid) oma valdkonna andmekogu ja ISi põhjal, on kesketest mõistetest ontoloogia ehitamine kõige loomulikum. Keskseid mõisteid tähistavad reeglina sõnad, mida kasutatakse tavaliselt kui räägitakse infosüsteemi andmetest. Ülevalt alla meetodit kasutades võib tee andmeteenuste sisend- ja väljundandmete tasemele olla suhteliselt pikk. Samas alt üles meetodi korral on keeruline näha ülemise astme mõisteid, mille tulemusena on tunduvalt aeganõudvam eristada mõisteid ja atribuute. See omakorda tähendab, et nõuetele vastava ontoloogiiani jõudmiseks kulub rohkem aega.

Kontseptualiseerimine ja realiseerimine on käesolevas metoodikas jagatud järgmisteks väiksemateks tegevusteks:

1. **Mõistete kogumine.** Leitakse valdkonna kesksed mõisted. Neid võiks olla esialgu 7-10. Kesksete mõistete leidmiseks tuleks läbi mõelda, mis eesmärgil infosüsteem on loodud, millistest andmetest räägitakse, kui teemaks on andmebaasi sisu, ja analüüsida võimalusel andmete esinemissagedust X-tee andmeteenuste parameetritena. Sellise lähenemisega on võimalik leida enamkasutatavad mõisted, millest moodustub ontoloogia tuumik.
2. **Mõistete detailiseerimine ja üldistamine ühe taseme võrra.** Kõik üldistavad ja detailiseerivad mõisted ei pea kuuluma X-tee andmeteenuste sisend-väljundandmete

hulka, nende eesmärgiks võib olla ka valdkonna arusaadavuse parandamine kolmandate isikute jaoks. Samas tuleks siiski silmas pidada skooopi ning piirduda vähima vajaliku hulga mõistetega selle katmiseks.

3. **Peamiste seoste määramine ja lisamine.** Kui tuumikmõisted koos võimalike üldistuste ja detailidega on määratud, siis võib lisada mõned peamised seosed nende vahel.
4. **Varajane realiseerimine.** Ontoloogia kontseptualisatsiooni realiseerimise tulemuseks on ontoloogia kirjeldus OWL keeles. Varajase kodeerimise eeliseks on loogikavigade avastamine ja ettekujutuse saamine tulevase ontoloogiast, selle keerukusest ja vastavusest spetsifikatsiooni nõuetele. Masinloetava ontoloogia loomise vahenditel nagu näiteks Protégé¹² on reeglina olemas sisseehitatud või lisatavad moodulid, mis kasutades kirjeldusloogikat, võimaldavad avastada loogilisi vastuolusid (näiteks Pellet¹³, Fact++¹⁴ jm).
5. **Iteratiivne arendus.** Vaadatakse üle X-tee andmeteenuste sisend- ja väljundandmed. Võib märgistada andmeväljad, mis on kaetud juba olemasolevate mõistetega või edaspidises etapis lisatavate mõiste atribuutidega. Sel juhul saab uute mõistete otsimisel keskenduda kasutamata andmeväljade analüüsile. X-tee andmeteenuste parameetrite analüüsi tulemusena on mõistlik korraga ontoloogiasse lisada ja realiseerida umbes 5 mõistet ning nende seosed. See tagab, et kogu protsess on pidevalt kontrolli all ning tõenäoliselt ei teki mõisteid, mis kuskile alamklassi või seosesse ei kuulu. Iteratiivse arendamise tulemusena jääb uute mõistete hulk järjest väiksemaks ja seda vähem on vaja lisada neid eraldi kontseptuaalsele mudelile, st võib piirduda vaid realiseerimisega.
6. **Andmetüübiomaduste lisamine.** Ontoloogia kirjeldusse lisatakse loodud mõistete indiviidide andmetüübiomadused (atribuudid), mis reeglina vastavad ISI andmekogu andmeobjektidele. Näiteks, Objekti_kood, Isikukood, Nimi, Eesnimi jms.
7. **Ontoloogia komponentide annoteerimine realiseerimisel.** Kui kõik komponendid, mõisted, seosed ja atribuudid on ontoloogias kirjeldatud ning automaatsed kontrollid loogilisi vigu ei tuvasta, on vaja lisada masinloetavasse ontoloogiasse ontoloogia enda ja kõigi tema komponentide inimkeelsed kirjeldused (nn annotatsioonid). Kirjeldused peavad olema vastavuses Semantika juhise¹⁵ nõuetega. Lisaks ontoloogia masinloetavale kujule OWL failina on kasutajasõbralik koostada ka dokumentatsioon. See hõlbustab ontoloogia haldust, taaskasutust ja edasiarendust.

Ontoloogia kontseptuaalse mudeli loomiseks on hea kasutada mingit teadmiste organiseerimise ja struktureerimise viisi. Soovitatakse luua terminite ja mõistete sõnastikud, esitada mõistete vahelised seosed ning määrata indiviidide atribuudid. Seda protsessi võib nimetada ka kontseptualiseerimise dokumenteerimiseks. Lisas 2 on toodud fragment kontseptualiseerimise protsessi dokumenteerimisest aadressiandmete süsteemi (ADS) andmekogu ontoloogia¹⁶ jaoks. Lisas 2 esitatud tabelid 1-4 on soovituslikud ontoloogia kontseptualisatsiooni esitamisel. ADS ontoloogia taksonoomiline struktuur tehti realiseerimise käigus ja see on esitatud Lisas 3.

¹² Protégé; <http://protege.stanford.edu/>

¹³ Pellet; <http://www.clarkparsia.com/pellet>

¹⁴ Fact++; <http://owl.man.ac.uk/factplusplus>

¹⁵ Semantika juhise; http://www.riso.ee/et/files/Semantika_juhis_v1.1.pdf

¹⁶ ADS ontoloogia dokumentatsioon on loodud Maa-ameti töötaja Mall Kivisalu poolt

4.2.5. Ontoloogiaarenduse toetustegevused.

Ontoloogiaarenduse toetustegevused (teadmushõive, dokumenteerimine, konfiguratsiooni haldus, hindamine, valideerimine, parendamine) peavad toimuma kogu ontoloogia arendusprotsessi jooksul. Nende tegevuste vajadus on erinevate arendusprotsessi tegevuste ajal erinev.

Hindamine ja parendamine on kõigi etappide lahutamatuks osaks. Loogilisi vastuolusid aitavad leida tarkvara automaatsed loogikakontrollid. Sisuliste vigade avastamiseks on tarvilik konsulteerida soovitatavalt mitme valdkonda tundva spetsialistiga. Mida varasemas faasis vead leitakse, seda vähem mõjutab nende parandamine kogu ontoloogia struktuuri.

Ontoogiatega halduse eesmärgiks on ontoloogia kirjelduse kaasajastamine. Ontoloogia muutub ajas vastavalt uutele nõuetele. Võib tekkida vajadus mestida sama sisuga mõisteid ja omadusi. Samuti võib tekkida vajadus restruktureerida ontoloogia taksonoomilist struktuuri kui mõiste alammõistete arv muutub liialt väikeseks või suureks, kui ülemmõiste eemaldatakse või lisatakse uusi alammõisteid. Eelnev kehtib analoogiliselt ka omaduste kohta. Riigi Isi valdkonnaontoloogiatega võrgustiku haldamine on märksa keerulisem kui iga üksiku valdkonna ontoloogia haldamine. Halduse tulemusena luuakse seosed erinevate ontoloogiatega sama sisuga mõistete ja omaduste vahel. Samas tuleb garanteerida kogu ontoloogiavõrgustiku aktuaalsus. See tagatakse konfiguratsiooni haldusega, mis hõlmab dokumentatsiooni ja ontoloogiatega kirjelduste versioneerimist ning nende muutuste juhtimist.

5. Tulem.

Ontoloogiaarenduse tulemiks on vastavat ISI valdkonda esitav ontoloogia kirjeldatuna OWL keeles. Ontoloogia peaks vastama spetsifikatsioonis esitatud nõuetele ja ontoloogia kirjeldus dokumendis „Nõuded RIHA ontoloogiatega“ toodud nõuetele. Erinevate tegevuste käigus loodud dokumentatsioon (spetsifikatsioon, kontseptualisatsiooni dokument jms) on samuti ontoloogiaarenduse väljundiks.

Riigi ISI valdkonnaontoloogiatega arenduse tulemiks on valdkonnaontoloogiatega võrgustik, mis talletatakse RIHAs.

Aadressiandmete süsteemi (ADS) ontoloogia spetsifikatsioon.

Eesmärk:

Ontoloogia eesmärgiks on esitada Aadressiandmete süsteemi poolt hõivatud valdkonna ontoloogia riigi IS semantilise koosvõime saavutamiseks veebiteenuste kirjelduste semantilise rikastamise abil nimetatud ontoloogia mõistetega

Kasutusstsenariumid ja kasutajad:

- Kasutajad on ISide disainerid ja arendajad, RIHA arendajad.
- Kasutusstsenarium: mingi teise ISi arendaja soovib luua uut veebiteenust. Kasutades RIHA funktsionaalsust esitab ta päringu, milles märgib uue veebiteenuse sisendile, väljundile või mõlematele vastavad mõisted. Vastuseks saab listi võimalike (komponent)veebiteenuste kirjeldustega.

Skoop:

Aadressiandmete süsteemis olemasolevate ja tulevaste veebiteenuste liideste sisend ja väljundparameetritele ning ADS andmekogu põhiaandmete väljadele vastavad mõisted.

Formaalsuse tase:

Formaalne, esitatud OWL keeles.

Taaskasutatavad ontoloogiad ja teised infoallikad:

- ADS määrus (Vabariigi Valitsuse 20.detsember 2007.a. määrus nr 251),
- avaliku teabe seadus,
- IS loomise dokumendid: ADS spetsifikatsioon, andmebaasiskeemid, olemasolevate WSDLs kirjeldatud veebiteenuste kirjeldused,
- geomeetria ontoloogia.

Põhimõisted:

Mõiste	Selgitus
Aadressiobjekt	Maaga seotud objekt, millele on määratud koha-aadress või millele koha-aadressi esitamise nõue tuleneb õigusaktist
Koha aadress	Objekti asukohta kirjeldav tekst
Aadressipunkt	Aadressi ruumilist asukohta väljendav punkt. Peab asuma ühe aadressiobjekti ruumikuju ala sees või joone peal, punktobjekti korral vastavas punktis millel on seos selle koha-aadressiga
Ruumiaadress	Aadressiobjekti arvutuslik (analüütiline) aadress, mis saadakse aadressiobjekti ruumikujude ja haldus- ning asustusüksuste ruumikujude analüüsimisel.

Aadressiandmete süsteemi (ADS) ontoloogia kontseptualisatsiooni esitus.

Tabel 1. Terminite sõnastik (fragment)

Nimi	Lühend	Kirjeldus	Ontoloogia komponendi tüüp
Aadressiobjekt	Objekt	Maaga seotud objekt, millele on määratud koha-aadress või millele koha-aadressi esitamise nõue tuleneb õigusaktist	Mõiste
Objekti kood	<i>ADS_OID</i>	<i>ADS-i poolt hallatav aadressiobjekti identifikaator (ADS_OID), identifitseerib objekti üle kõikide liikide.</i>	<i>Atribuut</i>

Tabel 2. Mõistete sõnastik (fragment).

Mõiste nimi	Indiviidi atribuudid	Seose nimi
Aadressiobjekt	<i>Objekti kood</i> <i>Objekti versiooni kood</i> <i>Paritoluregistri objekti kood</i> <i>Objekti liik</i> <i>Objekti versiooni algus</i> <i>Objekti versiooni lõpp</i> <i>Objekti versiooni olek</i> <i>Objekti taisaadress</i> <i>Objekti lahiaadress</i> <i>Objekti kehtestamise kuupäev</i> <i>Objekti kehtetuks muutmise kuupäev</i>	Aadressiobjekt_omab_Koha-aadress Aadressiobjekt_omab_Objekti-tsentroid Aadressiobjekt_omab_Ruumiaadress Aadressiobjekt omab Ruumikuju <i>(Viide geomeetria ontoloogiale)</i>

Tabel 3. Seoste kirjeldused (fragment)

Seose nimi	Lähtemõiste nimi	Kardina alsus	Sihtmõiste nimi	Pöördseos
Aadressiobjekt_omab_Koha_aadress	Aadressiobjekt	(M,N)	Koha-aadress	
Aadressiobjekt_omab_Ruumiaadress	Aadressiobjekt	(1,N)	Ruumiaadress	

Tabel 4. Indiviidide atribuutide kirjeldused (fragment)

Atribuudi nimi	Lähtemõiste nimi	Väärtuse tüüp	Väärtuste vahemik
Aadressi_identifikaator	Koha_aadress	String	
Koodaadress	Koha_aadress	String	
Objekti_kood	Aadressiobjekt	String	
Objekti_liik	Aadressiobjekt	String	
Objekti versiooni_algus	Aadressiobjekt	dateTime	
Objekti versiooni_kood	Aadressiobjekt	Integer	

ADS ontoloogia mõistete taksonoomia (fragment).

Alam/ülemmõiste seos on joonisel märgitud termini “is-a“ abil. Termin “Thing” tähistab kõige ülemist abstraktset mõistet.

