

ABSTRAKTI

*Mikko Lappalainen*

# ***Yläontologiat tiedon jakamisen mekanismeina : Yleinen suomalainen ontologia YSO ja yhdistetyn tiedon visio Suomessa***

*Yhteystiedot: Mikko Lappalainen, Kansalliskirjasto, mikko.lappalainen@helsinki.fi*

Suljetuissa ympäristöissä toimivien tietämispohjaisten järjestelmien sisällöt rakentuvat usein laajan implisiittisen informaation varaan: asioista käytetään tietylle rajatulle joukolle ymmärrettäviä nimityksiä, jotka voivat poiketa radikaalistikin yleisessä käytössä, tai toisen tietojärjestelmän käytössä olevista merkityksistä. Informaatiotieteiden ontologiat syntyivät alun perin ratkomaan tätä siiloutuneen tiedon ongelmaa. Tiedon jakamisen mahdollistamiseksi haluttiin tehdä näkyväksi se, missä merkityksessä kukin käytti käsitteistöään. Tässä yhteydessä ontologioilla tarkoitetaan siis tietyn aihealueen (tai järjestelmän) ”formaalia, eksplisiittistä jaetun käsitteistön määrittelyä.” (Guarino et al. 2009, 1-3)

Termillä yläontologia (engl. upper ontology, upper-level ontology, top-level ontology, foundational ontology) puolestaan viitataan tämän käsitteistön hierarkkisen kokonaisuuden ylimpiin tasoihin. Tämä taso koostuu yleensä muutamasta laaja-alaisesta luokasta (class) tai käsitteestä (concept) ja niiden ominaisuuksia kuvaavista määritteistä, sekä koneluettavista loogisista säännöistä.<sup>1</sup> Yläontologioita on kehitetty varsinaisesta kuvailukäsitteistöä erillisinä kokonaisuuksina, mutta ne voivat olla myös erottamaton osa laajaa kuvailuontologiaa.

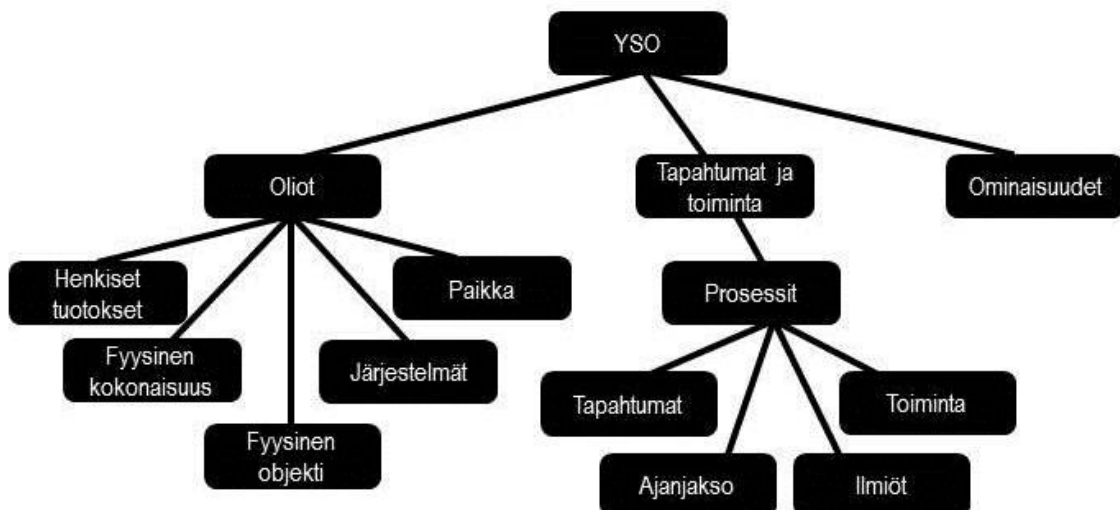
Yläontologian tarkoituksena on tehdä näkyväksi valittujen tiedonalojen yleisimmät käsiteluoikat ja täten mahdollistaa osaltaan semanttinen yhteentoimivuus eri tietoresurssien välillä. Voidaan esimerkiksi ajatella että sekä julkisten palveluiden että arkkitehtuurin asiansanastosta löytyy käsite *uimahallit*. Eri tiedonaloilla sama merkkijono voi kuitenkin tarkoittaa käsitteellisesti hyvin eri asioita: palveluiden näkökulmasta kansalaisille tarjottuja virkistystoimintoja ja arkkitehtuurin näkökulmasta tiettyyn käyttöön rakennettua rakennusta. Kun halutaan saada molemmat resurssit keskustelemaan keskenään, voidaan yläontologian avulla määrittää, mikä kunkin käsitteen sisällöllinen merkitys on. Yläontologia voisikin esimerkiksi kertoa, että julkisten palveluiden asiansanaston uimahallien yläkäsite on *julkiset palvelut* ja arkkitehtuurin puolestaan *julkiset rakennukset* ja täten käsitteet eivät vastaa

1 Ontologian ylimmille luokille määritellyt ominaisuudet periytyvät useimmissa tapauksissa kaikille luokkien alakäsitteille.

toisiaan merkkijonoyhtäläisyydestä huolimatta.<sup>2</sup> Kyse onkin pohjimmiltaan tietyn sanastoresurssin käsitteiden *ontologisen sitoumuksen* (engl. ontological commitment) (Gruber 1993, 2, Guarino 1994) ilmaisemisesta formaalissa muodossa niin, että se on ymmärrettävissä ja hyödynnettävissä myös koneellisesti. Tarkoituksena on, että tiedon jakamisen mahdollistamiseksi jokainen tekisi näkyväksi sen, missä merkityksessä käsitteitä käytetään omassa ympäristössä.

Koska yläontologioiden kehittämisessä työskennellään yleisimpien ajateltavissa olevien ryhmittelyjen parissa, on niihin usein haettu innoitusta filosofisen ontologian, eli nk. yleisen metafysiikan perinteestä, jossa maailmaa ja eri tiedonaloja on kategorisoitu läpi historian (Masolo 2003, 7-12). Sisällöllisenä erona on kuitenkin se, että informaatiotieteen ontologiat eivät yleensä pyri luokittelemaan kaikkea olevaa tai maailmaa yleensä, vaan tietyn rajatun aihealueen yhteisten käsitteistöjen sisältöjä (Smith 2003), ja näin mahdollistamaan tietojen vaihdon eri järjestelmien välillä. Lisäksi näissä ontologioissa ei oteta kantaa siihen, onko luokiteltuja asioita ”todellisuudessa” olemassa; jos luokittelu on toimiva tiedonjakamisen kannalta, se täyttää tehtävänsä.

Erillisiä ja monessa yhteydessä hyödynnettäviä yläontologioita on kehitetty 2000-luvun alusta lähtien, ja niitä on saatavilla vapaasti verkosta koneluettavassa muodossa. Tunnetuimpia näistä ovat muiden muassa BFO (Basic Formal Ontology), jota on hyödynnetty erityisesti biolääketieteen piirissä, SUMO (Suggested Upper Merged Ontology), ja DOLCE, jota on käytetty esimerkiksi WordNetin ontologisoimiseen. Suomessa Yleisen suomalaisen asiasanasto YSAn pohjalta kehitetty Yleinen suomalainen ontologia YSO sisältää yläontologian, jonka ylimmät luokat on esitetty kuvassa 1.



**Kuva 1.** YSON ylärakenne

YSON yläontologiaa ei pidä nähdä maailman sinänsä kategorisointina, vaan käytännön tarpeesta kummunneiden sanastotyökalujen käsitteiden yleisimpinä luokkina. YSON ylähierarkian avulla voidaan esimerkiksi tehdä näkyväksi se, missä merkityksessä YSAn

2 Kuvailuontologioissa yläontologian käsitteet on yleensä erotettu varsinaisista kuvailukäsitteistä, eikä niitä ole tarkoitettu käytettäväksi kuvailuun.

käsitteitä on kirjastojen bibliografisen aineiston viitetietokannoissa käytetty. Kun tiedetään, missä merkityksessä jokin käsite on esimerkiksi kansallisbibliografia Fennicassa ymmärretty, voidaan Fennican aineistot linkittää muita kuvailusanastoja käyttäviin järjestelmiin, tai vaikkapa Wikipedian kaltaisiin tietoresursseihin. YSON käsitehierarkia on ilmaistu semanttisen webin standardien mukaisesti ja julkaistu avoimena datana, jotta se olisi mahdollisimman suuren joukon hyödynnettävissä (Lappalainen 2014, 6-7).

YSOa kehittävän ONKI-projektin ja sen edeltäjän FinnONTO-hankeen visiossa YSON ylähierarkiaa voitaisiin käyttää yhtenäisen sanasto- tai ontologiapilven ”liimana”, jonka avulla eri muistiorganisaatioiden ja julkisen sektorin toimijoiden tietämysjärjestelmät saataisiin keskustelemaan keskenään. YSON ylähierarkia olisi siis se *tapa*, jolla sanastojen ontologinen sitoumus tehtäisiin näkyväksi. Tämä olisi tietoa hakevan kansalaisen kannalta erittäin tavoiteltavaa. Nykytilanteessa tieto ei ole yhteismitallista välttämättä edes yhden organisaation eri osien sisällä, ja kansalaisen voi olla hyvin vaikea löytää tietoa tietystä eri tahojen vastuualueisiin kuuluvasta asiasta.

## Lähteet

Borgo, Stefano; Masolo, Claudio 2009: “Foundational Choices in DOLCE”, teoksessa *Handbook on Ontologies, Second Edition*.

Gruber, Tom 1993: “Toward Principles for the Design of Ontologies Used for Knowledge Sharing.”

Guarino, Nicola; Carrara, Massimiliano; Giarretta, Pierdaniele 1994: “Formalizing Ontological Commitments.” <http://www.aaai.org/Papers/AAAI/1994/AAAI94-085.pdf>

Guarino, Nicola; Oberle, Daniel; Staab, Steffen 2009: “What is an Ontology?”, teoksessa Staab & Studer (ed.) *Handbook on Ontologies, Second Edition*.

Lappalainen, Mikko; Frosterus, Matias; Nykyri, Susanna 2014: “Reuse of library thesaurus data as ontologies for the public sector.” <http://library.ifla.org/819/1/086-lappalainen-en.pdf>

Masolo, Claudio et al 2003: “WonderWeb Deliverable D18: Ontology Library (final).” <http://www.loa-cnr.it/Papers/D18.pdf>

Smith, Barry 2003: ”Ontology”, teoksessa *Blackwell Guide to the Philosophy of Computing and Information*.