



2.11.2016

Yhteinen tiedon hallinta -hanke

Taustatietoa *Sanaston metatietomallin määrittely* -työpajan keskusteluun

Sanastotyötä tehdään paljon. Tuotettavan sanaston käyttötarve ja -kohde vaihtelevat ja siksi vaihtelevat myös sanastojen muoto, työ- ja esitystavat.

Sanastojen ja sanastotyön merkitys on kasvanut jatkuvasti tiedonhallinnan ja palveluiden ja palveluprosessien sähköistymisen myötä. Edelleen sanastoja tarvitaan myös perinteiseen käyttötarpeeseen, ihmisen väliseen viestintään. Pääsääntöisesti voidaan erottaa kaksi tärkeää sanastotyyppiä: *asiasanastot* ja *määrittelevät (terminologiset) sanastot*.

Asiasanastot

Yleisesti asiasanastolla viitataan systemaattiseen ja kontrolloituun hakusanaluetteloon, jota käytetään erityisesti aiheenmukaiseen tiedonhakuun ja indeksointiin; näin voidaan parantaa esimerkiksi jonkin resurssin löydettävyyttä. Asiasanastossa ilmaistaan tyypillisesti käsitteiden ylä-alakäsitesuhteet (LT-laajempi termi, ST- suppeampi termi), rinnakkaisuudet (RT - rinnakkaistermi, vrt. assosiatiiviset suhteet) sekä suositeltava termi ja mahdollisesti ohjaustermejä. Varsinaista käsitteen määritelmää ei yleensä ole esitetty, mutta estettä sen lisäämiselle ei sinänsä ole ja se auttaa käyttämään käsitteitä oikealla tavalla kuvailussa ja tiedonhaussa. Käsitevalikoima ja merkityt suhteet perustuvat kuvailun ja tiedonhaun tarpeisiin.

Perinteisesti asiasanastoja on julkaistu painettuina ja sähköisinä dokumentteina (esimerkiksi doc, odt, pdf, html). Suomessa on tehty yli kymmenen vuoden ajan myös asiasanastopohjaisia ontologioita, asiasanastojen kone-luettavia muotoja, joihin mielletään kuuluvaksi paitsi SKOS-formaatti, URIt niin myös kokonaishierarkia. Nämä sanastot voidaan julkaista ja niitä voidaan hyödyntää sekä käyttöliittymän että rajapintojen kautta Kansalliskirjaston ylläpitämästä Finto-palvelusta.

Asiasanastopohjaisia ontologioita tuotetaan ja ylläpidetään ontologiaeditoreilla, joita on saatavilla sekä kaupallisina tuotteita (esimerkiksi TopBraid Composer) että avoimen lähdekoodin tuotteina (esimerkiksi VocBench).

Terminologinen sanasto

On sanastotyyppi, joka sisältää terminologista tietoa eli erikoisalojen käsitteisiin ja niiden nimityksiin liittyvää tietoa. Terminologinen sanasto perustuu käsitteanalyysiin, jonka tuloksena käsitteelle annetaan nimi (termi), sen sisältö kuvataan (määritelmä ja mahdollisesti huomautus) ja sen suhteet muihin käsittehierarkian käsitteisiin kuvataan; määritelmä ja huomautus sisältävät tiedon käsitteen ymmärtämisen kannalta keskeisistä käsitesuhteista.

Terminologista sanastoa käytetään perinteisesti erityiskielen (esim. jonkin toimialan keskeiset käsitteet) eksaktiin kuvaamiseen, viestinnän ja käännoistyön

apuvälineenä. Lisäksi terminologista sanastotyötä on hyödynnetty tietojärjestelmäkehityksen tukena. Käsitevalikoima perustuu kohdealan keskeisiin käsitteisiin ja tavoitteena on antaa tietoa käsitteiden sisällöstä ja suhteista toisiin käsitteisiin.

Terminologisen sanastotyön menetelmät ovat ISO-standardoituja (ISO/TC 37). Terminologisia sanastoja tuotetaan ja ylläpidetään tehtävään räätälöidyillä sanastoeditoreilla. Terminologisten sanastojen kuvaamiseen on kehitetty erityinen xml-formaatti. Ontologisoituja terminologisia sanastoja ei ole julkaistu. Terminologisten sanastojen sisältöä on saatettu liittää osaksi asiasanastopohjaisia ontologioita.

Ontologia

Ontologia-käsitteenä on monimerkityksellinen; tietojenkäsittelytieteissä sillä tarkoitetaan kuvausta tarkasteltavan sovellusalueen käsitteistä ja käsitteiden välisistä suhteista koneluettavassa muodossa.

SKOS (Simple Knowledge Organization System) on W3C:n tuottama standardi sanastoresurssin tekniselle esitystavalle. Tarkemmin kyse on linkitetyn datan formaatista. Sitä on hyödynnetty lähinnä asiasanastojen ontologisoinnissa; SKOS-muoto tuottaa asiasanastolle perinteistä muotoa täsmällisemmän rakenteen. Standardiin lisätyt ns. laajennokset (*SKOS-Extended labels*, SKOS-XL) mahdollistavat myös monimutkaisempien sanastoresurssien esittämisen.

Ontologia itsessään ei ole minkään tietäntyyppisen sanastoresurssin synonyymi vaan ontologiamuotoisena voidaan ilmaista erilaisia aineistoja linkitetyn datan periaatteiden mukaisesti. SKOS on tällä hetkellä kansainvälinen standardi ko. tarkoitukseen.

Eri sanastoja

Asiasanastot ja terminologiset sanastot sisältävät siis erilaisia rakenteita ja käsitekokonaisuuksia eri tarkoituksiin. Resurssien yhdistäminen ”parhaita paloja poimien” ei ole mielekäästä.

Asiasanastopohjaisen ontologian käsitteistö on rajattu kuvailun ja tiedonhaun tarpeiden mukaan sopivalle tarkkuustasolle. Siksi asiasanastopohjaisen ontologian kokonaishierarkian yhdistäminen määrittelevään sanastoon, jota ei käytetä kuvailuun, ei ole hyödyllistä.

Jos määrittelevän sanaston rakenne ja käsitteistö eivät vastaa kuvailun ja tiedonhaun tarpeita, määrittelevän sanaston hierarkkista rakennetta ei aina ole kaikilta osin tarkoituksenmukaista tuoda ontologiaan, eikä kuvailussa välttämättä edes tarvita kaikkia määrittelevän sanaston käsitteitä.

Yhteentoimivuustyö

Semanttisen yhteentoimivuuden toteuttaminen tietojärjestelmien välisessä tiedonvaihdossa edellyttää määrittelevän sanaston olemassa oloa. Yhteentoimivuusmenetelmän mukaisesti tietomallinnuksen tarvitsemat käsitteiden merkitykset haetaan määrittelevästä sanastosta. Tavoitteena on vähentää jatkuvaa uudelleenmäärittelyn tarvetta mallinnustyössä ja samalla tehostaa sanastojen käyttöä, joita lähtökohtaisesti tuotetaan muuhunkin toiminnan tarvitsemaan työhön, esimerkiksi viestintään.

Terminologisen sanaston tarkkuustason nähdään soveltuvan sen hyödyntämiseen myös tietojärjestelmien semanttisen yhteentoimivuuden toteuttamisessa.

"Termieditori"
 Yhteentoimivuusmenetelmän mukaisen työn tukemiseksi on käynnissä järjestelmäkehitystyö, jonka yhtenä työkaluosana kehitetään ns. "Termieditoria". Tavoitteena on, että editorilla voidaan jatkossa tuottaa koneluettavia määritteleviä sanastoja jopa terminologisen sanaston tarkkuudella. Editorin tekniseksi formaatiksi on suunniteltu SKOS-standardia tarvittavilla laajennoksilla. Editorilla tuotettavien sanastojen julkaisupalveluksi on suunniteltu Finto-palvelua. Sama tekninen formaatti mahdollistaa palveluiden hyvän teknisen yhteentoimivuuden.

Miksi JHSMeta ei riitä?
 JHSMetan taustalla oleva metatietomalli ei noudata tunnustettua standardia. SKOS-esitystapa mahdollistaa formaalin esitystavan teknisten tietomäärittysten kuvaamiseen. Tulevaisuutta ajatellen SKOS-muoto mahdollistaa resurssien linkittämisen; voidaan esimerkiksi tuoda käytettyihin käsitteisiin perustuen tietomallien yhteyteen tieto siitä missä järjestelmässä/rekistereissä kyseisellä käsitteellä kuvattuja tietoja käsitellään. Toinen tulevaisuuden käyttökohde koneluettaville terminologisille sanastoille on hyvin todennäköisesti konekäyttäminen.

Työpajassa käsiteltävä aineisto

Ehdotus
 Työpajatyöskentelyn pohjaksi on luotu alustava soveltamisprofiili määrittelevälle sanastolle, se on saatavilla osoitteessa: <http://iow.csc.fi/model/st/>

Tätä ehdotusta voidaan muokata/supistaa/laajentaa tunnistettujen käyttötarkpeiden pohjalta.

Profiilin muodostamisessa on käytetty seuraavia lähteitä:

- SKOS spesifikaatio koneluettavien käsitejärjestelmien muodostamiseen: <https://www.w3.org/TR/skos-reference/>
- Soveltuvien osien JHS 175 Liite 2. (http://docs.jhs-suositukset.fi/jhs-suositukset/JHS175_liite2/JHS175_liite2.pdf)
- Terminologian sanasto: <http://www.tsk.fi/fi/info/TerminologianSanasto.pdf>
- Nordic Terminological Record Format (NTRF, taustalla mm. ISO-standardit): <http://www.nordterm.net/filer/NTRF/Ntrf.pdf>

Mahdollisuuksien mukaan voi tutustua myös: ISO 10241-1:2011: Terminological entries in standards -- Part 1: General requirements and examples of presentation:

http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=40362

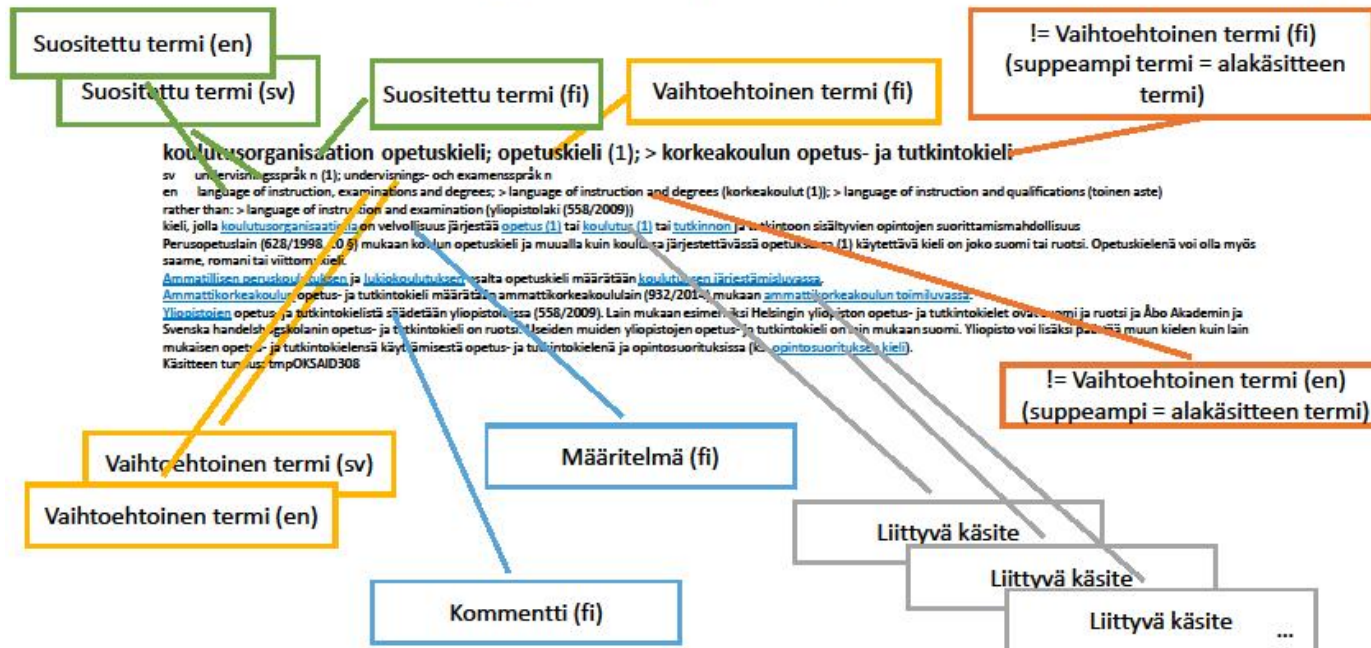
Tavoite
 Keskustelun tavoitteena on yhteisesti sopimalla muodostaa "Määrittelevän sanaston soveltamisprofiili", jossa määritellään mitä tietoja sanastossa tarvitsee kuvata. Samalla keskustellaan käyttötapausten pohjalta perustelut esitetyille tarpeille.

Hyödyntäminen
 Keskustelun tuloksena syntyvää profiilia hyödynnetään Yhteentoimivuuden välineistön kehitystyössä, osana *Yhteinen tiedon hallinta* -hanketta. Ensisijaisesti profiili otetaan käyttöön "Termieditorin" kehitystyössä ja profiilin toimivuutta myös testataan todellisella aineistolla. Lisäksi määrittelevän sanaston profiilin käyttöönoton mahdollisesti vaatimista muutoksista Finto-palveluun keskustellaan hankkeeseen sisältyvän Finto-tutkimusprojektin kanssa ja mahdolliset kehitystoimenpiteet ohjataan projektin normaalien ohjausprosessien kautta.

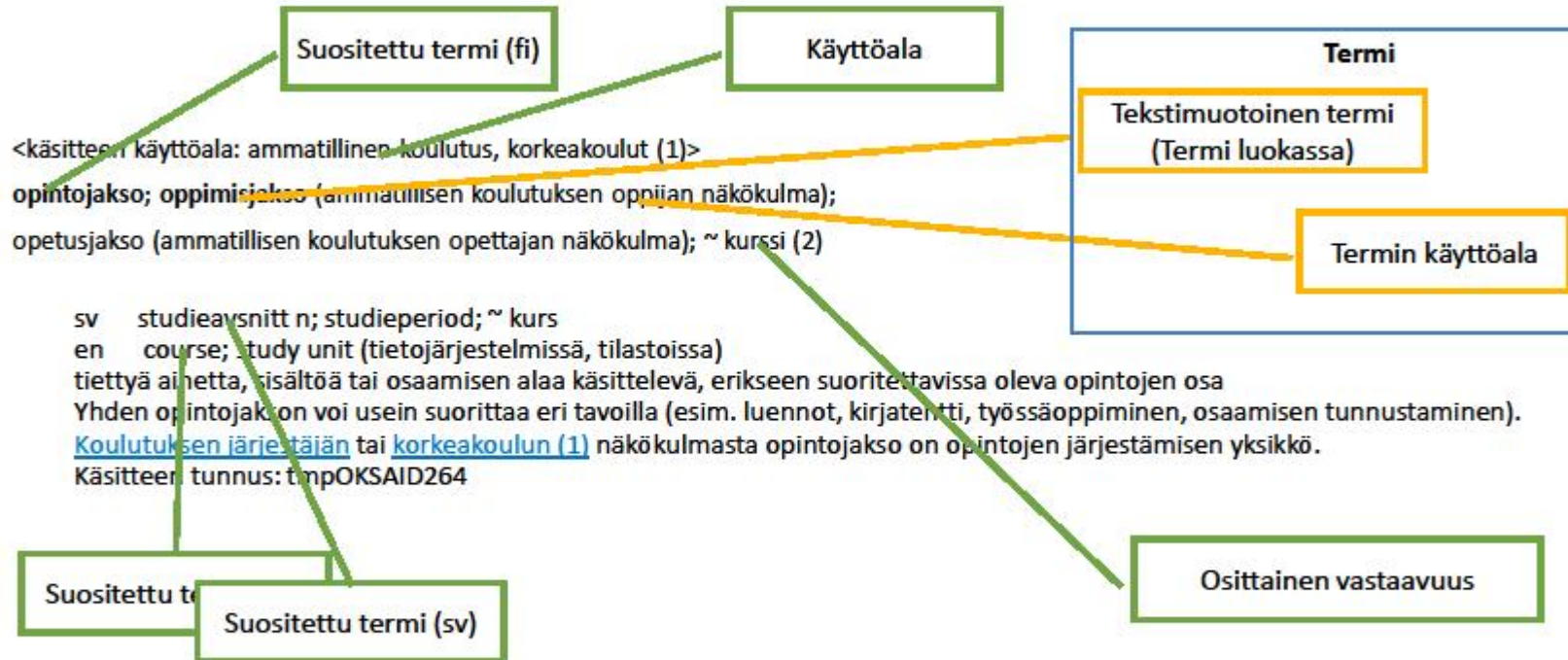


Yhteinen tiedon hallinta -hanke

Koulutusorganisaation opetuskieli (OKSA / Terminologinen sanasto)



Opintojakso (OKSA / Terminologinen sanasto)



Organisaatio (JHSMeta.fi)

