

YSAsta YSOon: retrospektiivi ja mitä seuraavaksi?

Jarmo Saarikko, Kansalliskirjasto
Kuvailun tiedotuspäivä 22.10.2019

Kirjastoverkon siirtymä

- Ontologiaan siirtymisestä puhuttu yli 10 vuotta eri yhteyksissä
- Koskettaa koko suomalaista sisällönkuvailijakenttää
 - yhteisluetteloinnissa mukana olevat
 - Kansalliskirjaston ylläpitämiä yleisiä sanastoja käyttävät organisaatiot
 - Käyttäjinä muitakin kuin kirjastoja ja kirjastojärjestelmiä: museot, arkistot, julkinen sektori, yritykset
- **YSAsta ja Allärsista YSOon siirtymä** –projekti 2018-2020
 - Käytännössä paljon monimutkaisempi prosessi

Sisällönkuvailun muutos ja evoluutio

- Yksikielisistä → monikielisiin sanastoihin
- Ketjuista → yksittäisiin käsitteisiin
- Tesauruksesta → ontologiaan
- Uusia mahdollisuuksia
 - Uusia MARC21-kenttiä aktiivikäyttöön
 - Linkitykset muihin sanastoihin
 - Automaattiset käännökset
 - Linkitetty data
 - Käsitteet ja termit selattavissa finto.fi-palvelussa

Yksikielisistä sanastoista...

- Ylläpito päättynyt ja poistuu käytöstä:
 - **YSA** – yleinen suomalainen asiasanasto
 - **Allärs** – Allmän tesaurus på svenska
 - **Musa** – Musiikin asiasanasto
 - **Cilla** – Specialtesaurus för musik

...monikielisiin sanastoihin

- Tilalle uudet sanastot:
 - **YSO** – yleinen suomalainen ontologia
 - *ALLFO – allmän finländsk ontologi*
 - **YSO-paikat**
 - *ALLFO – orter*
 - **SLM** – suomalainen laji- ja muototyyppisanasto
 - *FGF – finländsk genre- och formlista*
 - **Seko** – suomalainen esityskokoonpanosanasto
 - *(Besättning för framförande av musikaliskt verk)*
 - **MTS** – metatietosanasto (osittain)
 - *Metadataordlista*

Kuvailu yksittäisillä käsitteillä

- Kullakin käsitteellä kielestä riippumaton tunniste
- Kuvailussa kukin käsite omaan MARC-kenttäänsä
 - Poikkeuksena kenttä **382 Esityskokoonpano**
- Uusi ja kattava hierarkia
 - Ontologiassa alakäsitteellä on yläkäsitteen ominaisuudet
 - Seurauksena jonkin verran uusia käsitteitä
 - YSOssa on 3 juurikäsitettä
 - ”Oliot”, ”Ominaisuudet” ja ”Tapahtumat ja toiminta”

Erilaisia käsitetyyppejä

- Kuvailussa käytetään YSO:n yleiskäsitteitä ja yksilökäsitteitä
- YSO ontologiassa on myös ryhmitteleviä käsitteitä
 - **Näitä ei käytetä kuvailussa**
 - Hierarkisoiva käsite – esim. ”*orgaaniset objektit*”
 - Kokoelmat – esim. ”*aine käytön mukaan*”
 - Ryhmät – esim. ”*52 Historia*” (vrt. MARC: 072 \$2kkaa)
 - Eivät ole termeinä Asteri auktoriteettitietokannassa

Monikielisyys

- Tuotettu uusia termejä (sulikutarkenteilla), jotta erikieliset käsitteet vastaisivat tarkemmin toisiaan
 - YSO/ALLFO SLM/FGF
 - Käsitteiden muodostuskieli suomi ja ruotsi
 - MARC \$2 osakentässä käytetään sanastotunnuksen kielitarkennetta termin kielen ilmaisuun
 - yso/fin, yso/swe, slm/fin, slm/swe
 - englanninkielinen termi on täydentävä käännös, jota ei käytetä Melinda-kuvailussa
- Termien automaattiset käännökset
 - Finna hyödyntää jo nyt
 - Tulossa myös Melindaan (suomi→ruotsi→suomi)

Esimerkkejä sulkutarkenteista

suomi

ohjaus (hallinta)

ohjaus (neuvonta ja opastus)

ohjaus (taiteet ja media)

putoaminen

kaatuminen

osoitteet

adressit (vetoomukset)

suruadressit

ruotsi

styrning

handledning

regi

fall (ned)

fall (omkull)

adresser (kommunikation)

adresser (petitioner)

kondoleansadresser

Linkitykset YSOon / YSOsta

- YSO-pohjaisista erikoissanastoista (KOKO)
- Kansainvälisistä sanastoista ja palveluista, joissa on käännöksiä ja lisää tietoa käsitteistä
 - LOC Subject Headings
 - Wikidata
 - Käytössä jo YSO-paikat ontologiassa
 - Tulossa: YSO:n käytetyimmät käsitteet
 - Useimmiten löytyy lähes vastaava käsite
- MML:n Paikannimirekisteri (YSO-paikat)
- YSO on kuvailukäytössä monissa tietojärjestelmissä myös KAM-sektorin ulkopuolella

YSO-konversiot

- Konversiosäännöt
- Sanastojen konversiot
 - SKOS → MARCXML
- BIB-tietueiden konversiot
 - Melinda – kirjastojen yhteistietokannan konversio
 - Paikallistietokantojen konversiot
 - Muut konversiot
- Muiden YSAa käyttäneiden tietojärjestelmien konversiot
 - esim. KAVI

Konversiosäännöt

- Valmisteluryhmissä mukana kirjastojen ja kuvailuryhmien edustajat
- Auktoriteettitietueiden valmistelu
 - Linkitettyjen termien näyttäminen tietueessa
 - Kieliversiointi ja sanastotunnukset
- BIB-tietueiden konversio
 - Vertailu nykyisiin ja entisiin kuvailusääntöihin ja käytäntöihin
 - Käsiteltävien kenttien valinta
 - Osakenttien käsittely
 - Korvaavien termien valinta sekä kenttien järjestys ym.
 - Eri kuvailutavoista sekä virheellisistä kuvailuista ”selviäminen” oli erittäin haastavaa koodata sääntöihin ja ohjelmaan

Musiikki- ja elokuva-aineistojen konversio

- Konversioryhmän työskennellessä todettiin, että näiden aineistojen kuvailu tapahtunut huomattavasti toisenlaisella tavalla
 - Vaikka Musa ja Cilla oli yhdistetty YSAan ja Allärsiin, BIB-tietueissa oli vielä paljon näillä kuvailtuja tietueita
- Musiikkiaineistojen erillisestä konversiosta luovuttiin aikataulusyistä
 - Keväällä 2019 perustettu uusi ryhmä joutui valmistelemaan erittäin nopeasti täydentävät konversiosäännöt, jotka yhdistettiin alkuperäiseen konversiosääntöön
 - Erilaisuus huomattavaa erityisesti SLM-termien käytössä
 - Konversioon lisättiin mm. 382-kentän ja Seko-sanaston käyttö
- Mm. näistä syistä testaaminen oikealla aineistolla ennen Melindan konversiota jäi vähäiseksi

Sanastojen konversio ja vienti Asteriin

- YSO, YSO-paikat ja SLM
 - Konversio SKOS-tietomuodosta MARCXML-muotoon
 - Saman käsitteen muut termit 5XX kentässä
 - Vastaavat tai lähitermit muissa sanastoissa 7XX kentässä
 - Viety Asteriin toukokuussa ja lokakuun alussa 2019
 - Uudet ja muuttuneet termit viedään tästä eteenpäin viikoittain
 - Käytöstä poistuneet käsitteet → merkitään seuraaja
 - Automaattinen päivitys – ensi vuonna
 - Sanastot ladattavissa kokonaisena MARCXML muodossa Fintosta (myös Githubissa) – lokakuu 2019
 - Sanastot käytettävissä kuvailussa myös Finton rajapinnan kautta
 - Esim. Aurora, Koha-kirjastojen Täti
 - Ei tiedossa: Alman kuvailuympäristö
 - Fintoon lisätty tuki sanaston MARC-tunnuksen välittämiseen

Kirjastojärjestelmien konversioiden aikataulu

- Kirjastojen yhteistietokannan konversio (Melinda)
 - Toteutettu heinäkuussa 2019
- Paikallistietokantojen konversiot
 - Koskee korkeakoulukirjastoja, joilla Kansalliskirjasto tekee konversion
 - Voyager-kirjastot – lokakuussa 2019
 - Koha-kirjastot – marraskuussa 2019
 - Axiell Aurora ym.
- Muiden YSAa käyttäneiden järjestelmien YSO-konversiot
 - esim. Sierra – aikataulu ei tiedossa

Konversio oli monimutkaista

- Konversion toteutus oli huomattavasti laajempi ja monimutkaisempi kuin alun perin kaavailtu “pelkkä” asiasanakenttien YSAsta YSOon konversio.
- Konversiossa käsiteltiin kaikki aineistotyypit, joissa oli käytetty em. sanastoja kuvailussa.

Melindan yso-konversion toteutus

- Melindan ysa-termejä ja kenttiä ”siivottiin” noin vuoden ajan ennen konversiota
- Kaikki **ysa**, **allars**, **musa** ja **cilla** -merkityt termit konvertoitiin
- Konversio käsitteli em. termit kentissä
 - 385, 567, 648, 650, 651 ja 655
- Konversiossa purettiin ketjut
 - Eri osakenttien termit siirrettiin niitä vastaaviin erillisiin kenttiin
 - Termejä siirrettiin myös kenttiin 257, 370, 382, 388 ja 653
- Käytettiin sanastoja **yso**, **yso-paikat**, **slm**, **seko**
 - Näihin sanastoihin kuulumattomat termit siirrettiin kenttään 653 (+ 2. indikaattori)
 - Kentässä 648 lisättiin sanastotunnus ilman \$0 osakenttää

Musiikki- ja elokuva-aineistot, pelit

- Nämä konvertoitiin eri tavalla kuin tekstiaineistot
- Vastintermit etsittiin ensisijaisesti SLM sanastosta ja vietiin 655-kenttään
 - “aiheet” -termin jälkeiset termit valittiin aina YSO:sta
 - “aiheet” ja “musiikki” -termit poistettiin
- Samaan ketjuun kuuluneiden termeille lisättiin ko. termin uuden kentän ‡8-osakenttään ketjun järjestysnumero muodossa ‡81\u, ‡82\u, jne.

Melindan yso-konversion tuloksia

- Noin 10 miljoonaa tietuetta, joita ei konvertoitu mitenkään
- Noin 4.9 miljoonaa tietuetta konvertoitiin
- Poistettiin noin 23 miljoonaa kenttää
- Tuotettiin 45 miljoonaa **YSO** ja **SLM**-termiä (n. 22 milj. suomeksi ja ruotsiksi)
- Tuotettiin noin 0.5 miljoonaa **seko**-termiä
- Tarkistettavaksi jäi noin 3,5% konvertoiduista kentistä

Melindan yso-konversion tarkistuslista

Noin **800.000** kenttää (3,5% konvertoiduista kentistä)

- Jokin kentän termi ei löytynyt käytetyistä sanastoista
 - **600.000** kenttää, 170.000 eri termiä, joista 300 esiintyi yli 100 kertaa ja 135.000 termiä vain 1-2 kertaa.
 - Nämä termit ovat nyt kentässä 653.
- Termille löytyi useampi mahdollinen vaihtoehto
 - **160.000** kenttää, 1.400 eri termiä, joista 300 vain 1-2 kertaa
- Poistetut termit
 - **21.000** kenttää (musiikki, musik, aiheet, fiktio, fiktion)
- Kentässä \neq 6 osakenttä - kyrilliset kentät
 - **3.400** kenttää

Käsitteitä, joita ei voi konvertoida suoraan

- ❖ Syynä eroihin mm. yksikielisestä monikieliseen sanastoon siirtyminen sekä muut sanastoissa tapahtuneet muutokset
 - **mallit**
 - mallit (mallintaminen)
 - mallit (ammatit)
 - käsityömallit
 - roolimallit
 - käsittemallit
 - valumallit
 - **messut**
 - messut (tapahtumat)
 - messut (sävellykset)
 - messut (jumalanpalvelus)
 - **opetusohjelmat**
 - opetusohjelmat (oppilaitokset)
 - opetusohjelmat (tietokoneohjelmat)
 - opetusohjelmat (media)
 - **ohjelmat**
 - ohjelmat (suunnitelmat)
 - ohjelmat (media)
 - tietokoneohjelmat
 - **ohjaus**
 - ohjaus (taiteet ja media)
 - ohjaus (neuvonta ja opastus)
 - ohjaus (hallinta)
- ❖ Oikea termi on valittava tutkimalla tietuetta tai kuvailun kohdetta.

Mitä seuraavaksi?

- Paikalliskantojen konversiot
- Melindan BIB-tietueiden korjaukset tarkistuslistan mukaan
- Sanastojen täydennykset ja päivitykset Asteriin
 - Runsaslukuisimpia tarkistettavia termejä voidaan lisätä sanastoihin
- Konversio-ohjelman päivitykset

Paikallistietokantojen konversiot

- Valinnaiset toiminnot
 - **Kaksikielisyys** – toisen kielen termin lisääminen
 - **†8-osakentän lisääminen** ketjuista purettuihin termeihin (musiikki)
 - **Kaikkien tietueiden tulostus** (myös konvertoimattomat)
- Melindaan kuuluvien kirjastokantojen konversiot
 - Voyager ja Koha (yliopisto- ja amk-kirjastot)
 - Totetutus Kansalliskirjastossa, loka-marraskuu 2019
 - Viivästynyt alkuperäisestä mm. samanaikaisen Koha-siirtymän vuoksi
 - Axiell Aurora ym.
- **Konversio suositellaan tehtäväksi kaikille Finnaa käyttäville organisaatioille, jotka ovat käyttäneet YSA/Allärs -sanastoja**
 - Finna hyödyntää ontologian termien tunnisteita ja täydentää asiasanojen muunkieliset termit hakuindeksiin
 - Aiheet haettavissa eri kielellä, kuin millä kuvailu on tehty

Konversio-ohjelma



- Kansalliskirjasto tarjoaa Python3-kielisen avoimen koodin ohjelman
 - Kenen tahansa käytettävissä ja muokattavissa omaan tietokantaan soveltuva versio
 - Havaitut ongelmat konversiossa ja ohjelman tuottamilla tarkistuslistoilla on korjattu ohjelmaan. Mm. aineistotyyppien tunnistamista ja termien käsittelyjärjestystä on parannettu
 - Konversio-ohjelmaan tehdyt muutokset ja versiopäivitykset on dokumentoitu Githubissa
 - <https://github.com/NatLibFi/yso-marcbib>
- Konversiota harkitessanne poimikaa aina uusimmat versiot ohjelmasta ja sanastoista
 - Korjauksia ohjelmaan tehdään, mikäli konversioissa havaitaan esimerkiksi erilaisista aineistoista johtuvia ongelmia.



Konversio-ohjelman tarkistuslistat



Konversio-ohjelma tuottaa tarkistuslistalle kaikki ne kentät, joita ei voida konvertoida suoraan. Joitakin näistä voi korjata erillisajoina, mutta suurin osa käsiteltävä yksitellen

- Termit, joita ei löytynyt sanastoista
 - Vapaan indeksoinnin termit, kirjoitusvirheet, muunkieliset termit
- Käsitteet, joille ei löytynyt yksiselitteistä vastinetta
 - Useimmissa tapauksissa käsitteelle löytyi useampia vastineita
 - Esim. termillä on sulkutarkenteellinen muoto kahdessa tai useammassa käsitteessä
 - Kaikki 653 kenttään siirretyt termit ovat mukana tarkistuslistalla

Melindan konversiossa havaittua

Havaitut ongelmat on mahdollisuuksien mukaan korjattu konversio-ohjelmaan tai sanastoihin, jotteivät ne toistuisi seuraavissa konversioissa.

- Joissakin tapauksissa sanastoista löytynyt “oikea” termi ei viittaakaan samaan käsitteeseen ja sanastossa on vain yksi vaihtoehto (joka on kyseisessä tapauksessa väärä)
 - Esimerkiksi ketjussa ollut *Viborg*, jolla tarkoitettiin tanskalaista kaupunkia, konvertoitui Suomen Viipuriksi.
 - Muita hankalia tapauksia esim. *Kemi*, *Georgia*, *Lappi*, *Puškin*
- Ruotsinkielinen soitinnimi korvautui suomenkielisellä seko-termillä
 - Koska seko on yksikielinen, ruotsinkielinen termi poistui kokonaan

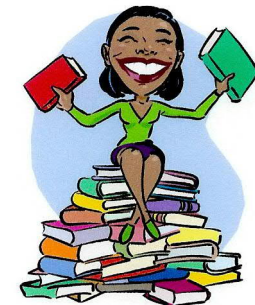
Erikoistapauksia

- ‡g-, ‡3-, ‡6-osakenttien käsittely
- Melindan konversiossa lisätyt ‡8-osakentät osoittautuivat useimmissa tapauksissa melko tarpeettomiksi, joten niitä poistetaan muiden korjausten ohessa
- Osa YSO-termejä, kun olisi pitänyt olla SLM-termi tai päinvastoin
 - Tämä johtui eri aineistotyyppien erilaisesta käsittelyjärjestyksestä vastintermien haussa
- Melindan muut kuin konvertoidut kentät
 - ysa- ja allars-termit on kartoitettu ja käsitelty manuaalisesti (eniten 600 ja 610 kentissä)

Tulevaisuus?

- Mitä sanastoja tarvitaan kuvailussa?
 - Eri aineistotyypit
 - **Fiktioaineistot**
 - **Tutkimusdata, jne.**

Fiktioaineistojen kuvailu



- Tämän hetken tilanne:
 - **Kaunokki** ja **Bella**: ylläpito on jäädytetty
 - Kaunokin, Bellan ja YSO:n käsitteet päivitetään **KAUNO**on
 - Aiheiden kuvailussa käytetään KAUNOa (MARC 650)
 - Ylläpito Helsingin kaupunginkirjastossa
 - Maantieteellinen paikka (MARC 370 ja 651)
 - Käytetään **YSO-paikat** sanastoa
 - Lajityyppi- ja muototermit (MARC 655)
 - Käytetään **SLM** sanastoa
- Mahdolliset BIB-tietueiden konversiot
 - Kaunokki + Bella → Kauno + YSO-paikat + SLM

Linkitettyjen sanastojen tulevaisuus



- Kuka tarvitsee? Missä tarvitaan?
 - Erilaiset tiedontuottajat ja aineistot sekä julkaisujärjestelmät
- Sanastot julkaistaan avoimesti
- Hyödynnetään linkityksiä tehokkaasti
 - YSO:n kaikille yhteiset käsitteet + erikoissanastot → **KOKO**
 - Mihin muihin sanastoihin ja ontologioihin pitäisi YSO linkittää?
 - Esimerkiksi terminologiset sanastot, joissa **määritelmät**?
 - Voiko linkityksiä olla liikaa?
- Ontologioiden täydentäminen uusilla ominaisuuksilla
 - Tarvitaanko käsitteiden välillä muitakin kuin hierarkiasuhteita?
 - Tietämysjärjestelmät
 - Erikoissanastojen jatkuva ylläpito ja resursointi on vaativa haaste

Kiitos!

Tämä ei ole loppu, vaan uuden alku!



Konversio-ohjelmat:

Okko Vainonen, Alex Kourijoki

Projektin vastaava: Matias Frosterus

Yhteistyöryhmä: Mikko Lappalainen

<https://www.kiwi.fi/display/ysall2yso>

#ysall2yso



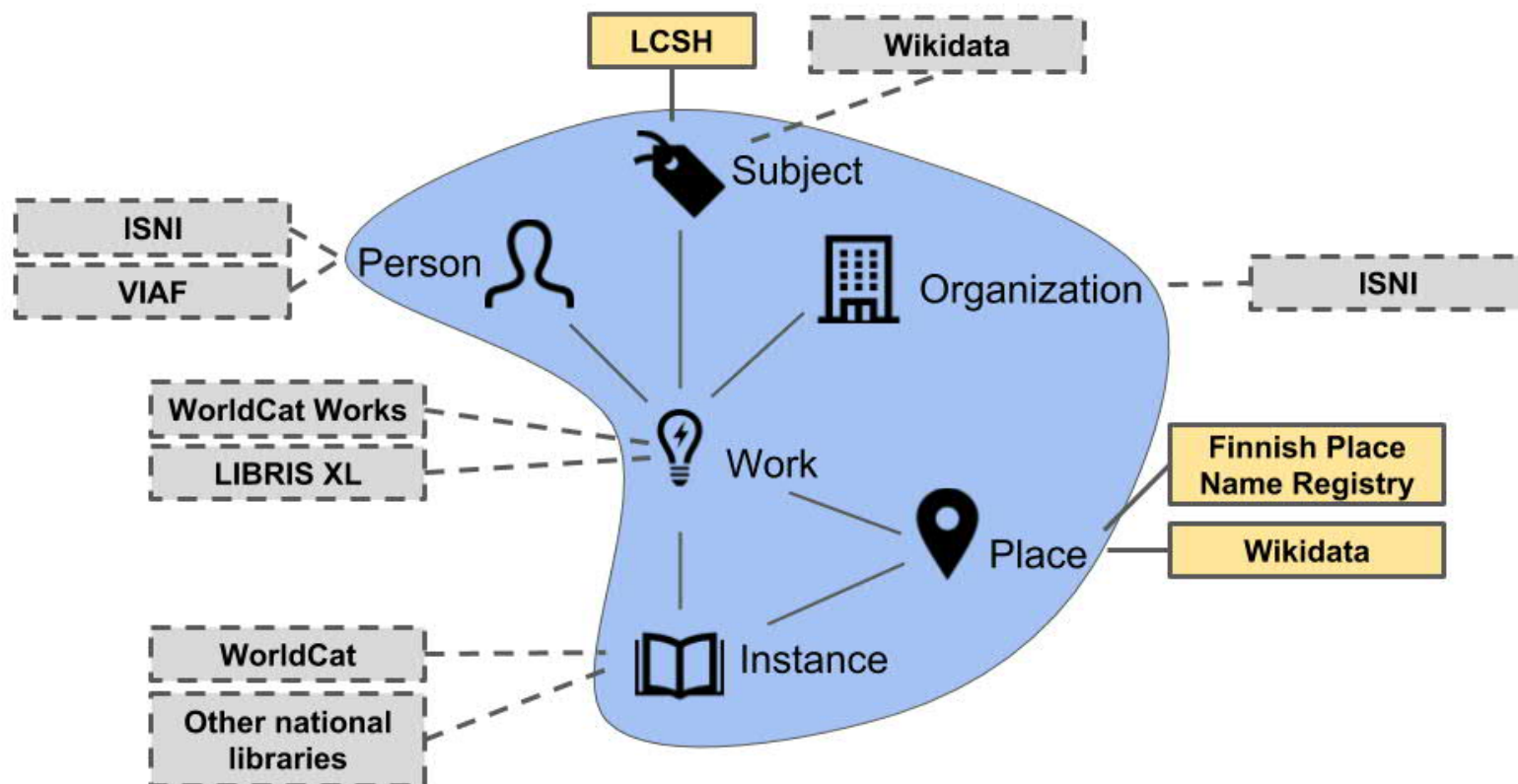
KANSALLISKIRJASTO

Esimerkkejä

- Linkitetty avoin data (LOD)
- Ontologioiden täydentäminen uudentlaisilla suhteilla
 - Esimerkkinä Agrontology
- Käsitteet ja linkitykset eri sanastoissa
 - YSO – LCSH – jne.

Linkitetty avoin data (LOD)

- Kirjastoaineistot ovat mukana maailmanlaajuisessa linkitetyn datan verkossa
 - Metadatan avoin julkaiseminen
 - Linkitetyt auktoriteetit ja sanastot
 - Kieliriippumattomat tunnisteet
- <http://data.nationallibrary.fi/>
- Esimerkki hyödyntämisestä
 - <https://github.com/COMHIS/fennica>



Current and potential linking targets for Fennica Linked Data.

<https://pro.europeana.eu/page/issue-9-swib> Osma Suominen @ SWIB18

Ontologioiden täydentäminen

- Ontologian tietomallia voidaan täydentää muillakin kuin hierarkkisilla suhteilla
 - causes *is inverse of* isCausedBy ('water erosion' <isCausedBy> 'water')
 - controls ↔ isControlledBy
 - affects ↔ isAffectedBy ('pest' <isAffectedBy> 'pest control')
 - isComponentOf ↔ hasComponent ('tree' <hasComponent> 'leaf')
 - isDerivedFrom ↔ isSourceOf ('plant' <isSourceOf> 'plant oil')
 - isMadeFrom ↔ isUsedToMake ('hops' <isUsedToMake> 'beer')
- Esimerkiksi eliöt, niiden nimet ja niiden väliset suhteet
 - hasGenus ↔ includesSpecies
 - hasSpecies ↔ includesSubspecies
 - hasSynonym ↔ isLocalNameOf
 - hasTaxonomicCode
 - developsFrom ↔ developsInto
 - hasDisease ↔ isDiseaseFor
 - pestOf ↔ hasPest ('Bactrocera dorsalis' <pestOf> 'Litchi chinensis')
 - isBeneficialFor ↔ benefitsFrom ('plant health' <benefitsFrom> 'pest control')

kissaeläimet yso:p864 Agrovocissa

| | | |
|--------------------|---|-------------------------|
| PREFERRED TERM | cats | <i>käytettävä termi</i> |
| BROADER CONCEPT | Felidae (en) | <i>yläkäsité</i> |
| NARROWER CONCEPTS | kittens (en) | <i>alakäsité</i> |
| ENTRY TERMS | Felis catus (en) | <i>ohjaustermi</i> |
| IS AFFLICTED BY | cytauzoonosis (en) | <i>vaikuttaja</i> |
| IS MEMBER OF | furbearing animals (en), monogastric animals (en) | <i>kuuluu</i> |
| IS USED AS | laboratory animals (en), pet animals (en) | <i>käytetään</i> |
| IN OTHER LANGUAGES | 猫 Chinese chat French Katze German КОШКИ Russian | |
| URI | http://aims.fao.org/aos/agrovoc/c_1390 | |

lehmät yso:p10991 Agrovocissa

Useita yläkäsitteitä

- ... > vertebrates > mammals > ruminants > cattle > cows
- ... > useful animals > domestic animals > livestock > cattle > cows
- organisms > females > female animals > cows

PREFERRED TERM

cows

BROADER CONCEPT

cattle (en), female animals (en)

NARROWER CONCEPTS

dairy cows (en), suckler cows (en)

SCOPE NOTE

Use only for cattle and zebu cattle;
for other species use <females> plus the descriptor
for the species (en)

INCLUDES

heifers (en)

IS MEMBER OF

milk yielding animals (en)

IS SOURCE OF

cow milk (en)

IN OTHER LANGUAGES

母牛 Chinese

Vache French

Kuh German

коровы Russian

URI

http://aims.fao.org/aos/agrovoc/c_1939



KANSALLISKIRJASTO

Linkitykset: poro

LCSH linkitykset

| | |
|-----------------------|-----------------------------------|
| LOC | Reindeer |
| YSO | Poro |
| Wikidata | Rangifer tarandus |
| OCLC – FAST | Reindeer |
| U.S. NAL Thesaurus | Rangifer tarandus |
| DNB | Ren |
| BnF | Renne |
| NCLibrary of Florence | Renne |

Muita sanastoja

| | |
|-----------------|------------------------------|
| Lajitietokeskus | Peura |
| Wikidata | peura |
| Agrovoc | Poro |
| Eurovoc | Poro |
| ZBW | Rentiere @de |
| GBIF | Peura |
| Libris | renar @sv |