

Ontologioiden käytön moninaisuudesta ja tulevaisuusnäkymistä: Finto-palvelun sidosryhmien ontologiapalvelulle kohdistamat kehitystoiveet

Kirjoittajat: Susanna Nykyri ja Tuomas Palonen

11. helmikuuta 2014

Kansalliskirjasto / Tutkimuskirjasto



Sisällys

Tiivistelmä	3
1 Johdanto	3
2. Haastateltavien esittely.....	5
3. Sisällönkuvailun käytännöistä ja tavoitteista.....	7
4. Järjestelmistä	14
4.1. Ontologioiden hyödyntämisestä järjestelmissä	14
4.2. Ontologioita hyödyntämättömistä järjestelmistä.....	16
4.3. Teknisistä ja organisatorisista haasteista	18
5. Erikoisontologioista ja -sanastostoista sekä niiden ylläpidosta	21
5.1. Erikoisontologioiden ja YSON suhteesta	22
5.2. Erikoisontologian ylläpidon haasteista.....	24
6. Ontologioiden hyödyistä, haasteista ja kehitystoiveista.....	26
6.1. Näkemyksiä ontologioiden hyödyistä	26
6.2. Ontologioiden haasteista	28
6.3. Kehitystoiveista.....	30
6.4. Instanssiontologioista	33
6.5. Ontologioiden kehitystyön mallista	34
7. Johtopäätökset	36
7.1. Sidosryhmien käytännöistä ja toiveista ja niihin vastaamisesta	36
7.2. YSON ja Finto-palvelun kehitysmahdollisuuksia.....	39
7.3. Suosituksia ja kehitysideoita sidosryhmille	41
Lähteet	44
Liite 1: Haastattelurunko.....	45

Tiivistelmä

Projektiryhmän toteuttama sidosryhmähaastatteluraportti perustuu haastatteluihin, jotka on tehty ONKI-projektin alkuvaiheessa (19.6.2013–7.11.2013). Haastatteluja on tehty niin laajasti, että raportti toimii myös läpileikkauksena Suomen ontologioiden käyttöön ja ontologiapalveluun kohdistuviin toiveisiin; siinä on esitetty tilanne vuonna 2013 ja tulevaan kehitystyöhön kohdistuvat toiveet ja visiot. Haastateltavat organisaatiot ja sidosryhmät edustavat laajasti kirjasto- ja muistialan organisaatioita sekä myös hallintoa ja mediaa. Kaikkiaan kahtatoista eri yhteistyötahoa kuultiin.

Raportissa käsitellään ontologioihin liittyviä toiveita ja käyttökokemuksia. Keskeisimpiä huomioita ovat selkeä kysyntä ja tarve ontologiapalvelulle, YSOon kohdistuvat kehitys- ja laajennustoiveet, käyttäjäkunnan huomattava heterogeisuus, ontologioiden käyttönoton haastavuus ja uudenlaiset käyttötarpeet.

1 Johdanto

Kansalliskirjastossa käynnistyi maaliskuussa 2013 kansallista ontologiapalvelua kehittävä ONKI-projekti. Palvelu sai syksyllä 2013 uudeksi nimekseen Finto (ks. palvelusta projektin wiki-sivusto¹). Projektin ensimmäisen vaiheen keskeisiksi tavoitteiksi oli määritelty vakiinnuttaa ja ylläpitää kansallista ontologiapalvelua sekä varmistaa palvelun jatkuminen; siirtää palvelun tekninen infrastruktuuri tutkimushankkeelta Kansalliskirjastoon; parantaa infrastruktuuria ja toimintaprosesseja; kehittää Yleistä suomalaista ontologiaa ja Allmän svensk ontologita; perustaa palvelua ohjaamaan konsortio; ja laatia palvelulle jatkosuunnitelma, toimintaperiaatteet ja pysyvä toimintamalli.

Meneillään oleva ONKI-projekti pohjautuu Aalto-yliopistossa vuosina 2003–2012 toimineen FinnONTO-projektin työlle. Tästä taustasta voi lukea projektista vastanneen Semanttisen laskennan tutkimusryhmän sivuilta². Vuonna 2013 ontologioiden ja ontologiapalvelun kehitystyö siirtyi Kansalliskirjaston ONKI-projektin vastuulle ja palvelun nimi vaihtui ONKIsta Fintoksi. Siirtymän yhteydessä Aalto-yliopiston aikainen fokus ja tavoitteet ovat osin muuttuneet, sillä kyse ei ole tutkimushankkeesta vaan käytännön palvelun kehittämisestä. Kansalliskirjaston ONKI-projekti on teettänyt Tekniikan Sanastokeskuksella FinnONTO:n aikaisista ontologisointityön toimintaperiaatteista raportin, joka julkaistaan alkuvuonna 2014. Kansalliskirjaston projektin teoreettisemmista haasteista ks. Lappalainen, Nykyri ja Palonen 2013.

ONKI-projektissa YSOa eli Yleistä suomalaista ontologiaa kehitetään lähtökohtana YSO:n keskeiset käyttötarpeet. YSO puolestaan perustuu käsitteistöltään Yleiseen suomalaiseen asiasanastoon YSAan, jolla on suomalaisessa sisällönkuvailussa monikymmenvuotiset perinteet. Tämä raportti on kooste keskeisiltä käyttäjäryhmiltä kerätystä palautteesta, käytännöistä ja ideoista. Tulevia ja nykyisiä käyttäjiä haastateltiin, jotta voidaan suunnitella ontologiaa ja palvelua nykyisten ja potentiaalisten käyttäjien näkemykset, kokemukset ja toiveet huomioon ottaen. Kansallisen Finto-ontologiapalvelun kehittäjillä on tukenaan myös Kansalliskirjaston sisäistä asiantuntemusta mm.

¹ Projektin wiki-sivusto, URL: <https://wiki.helsinki.fi/display/ONKI/>

² URL: <http://www.seco.tkk.fi/>

sisällönkuvailijoiden ollessa osa ONKI-tiimiä. Koska palvelua ja ontologiaa kehitetään laajemmalle käyttäjäryhmälle ja myös muille kuin perinteisille kirjastoalan organisaatioille, on heidän kuulemisensa erityisen tärkeää.

Sidosryhmien tarpeet ovat heterogeenisia, ja ne eivät ole kaikilta osin identtisiä kirjastoalan tarpeiden kanssa; kirjastoala onkin edelleen palvelun tärkeä ja varsin keskeinen asiakasryhmä, mutta mukana on myös muita käyttäjäryhmiä, joilla on osin eriävät tarpeet ja odotukset Kansalliskirjaston ontologiapalveluille. Haastatteluisissa ja tapaamisissa ovat painottuneet keskeisen kirjastoalan lisäksi erityisesti media-ala, ja myös muut, ei-perinteiset metadatala hyödyntävät alat, kuten verkkosivustojen kehittäjät. Olemme pyrkineet kuulemaan käyttäjiä laajasti, ja tarkoituksemme on jatkaa toiveiden ja käytäntöjen kuulemista koko projektin ajan.

Nyt käsillä oleva raportti perustuu haastatteluihin, jotka on tehty ONKI-projektin alkuvaiheessa (19.6.2013–7.11.2013) eli ennen nykyisen Finto-palvelun julkistamista. Haastatteluja on tehty niin laajasti, että raportti toimii myös läpileikkauksena Suomen ontologioiden käyttöön ja ontologiapalvelun kehitysohjelmaan kohdistuviin toiveisiin; siinä on esitetty tilanne vuonna 2013 ja tulevaan kehitysohjelmaan kohdistuvat toiveet ja visiot.

Haastattelurunko on esitetty liitteenä 1. Kaikissa haastatteluisissa haastattelijoina toimivat projektin tietoasiantuntijat Tuomas Palonen ja Susanna Nykyri, sekä osassa myös Mikko Lappalainen. Haastattelut kestivät 1–2,5 tuntia, ne nauhoitettiin haastateltavien luvalla, ja kirjoitettiin puhtaaksi. Nauhoitettua haastatteluaineistoa kertyi yhteensä yhdeksäntoista tuntia. Haastattelut tapahtuivat pääosin haastateltavien organisaatioiden tiloissa.

Raportti etenee taustoittavista luvuista (luvut 1–3) kehityskohdekohtaisiin lukuihin (4–6) ja lopuksi (luku 7) on esitetty sidosryhmähaastatteluista heränneet konkreettiset toimenpide-ehdotukset ja ratkaistavat kysymykset. Tekijät haluavat korostaa, että kehitysohjelma on jatkuvaa, joten paitsi että uusia teemoja nousee, on myös ns. kestoaiheista (kuten järjestelmien ja sanaston ylläpito) syytä olla kehitysohjelman eri vaiheissa jatkuvassa ajatustenvaihdossa laajasti erilaisten sidosryhmien kanssa.

2. Haastateltavien esittely

Haastateltavat organisaatiot ja sidosryhmät edustavat laajasti kirjasto- ja muistialan organisaatioita sekä myös hallintoa ja mediaa. Kaikkiaan kahtatoista eri yhteistyötahoa kuultiin. Haastattelut olivat kahta tilaisuutta lukuun ottamatta ryhmähaastatteluita, joissa haastateltavia oli läsnä kahdesta kuuteen henkilöä. Haastateltavat ovat erilaisia ontologiatyön kannalta keskeisiä asiantuntijoita ja yhteistyötahoja ollen sisällönkuvailijoita, erikoisontologian ylläpitäjiä, tietopalveluasiantuntijoita ja/tai teknisiä asiantuntijoita.

Haastatellut organisaatiot ja sidosryhmät ovat: Svenska YLE, YLEn Elävä arkisto, YLEn kirjastotietopalvelu, Brages Pressarkiv, Kirjasampo, Terveiden ja hyvinvoinnin laitos (THL), Julkishallinnon ontologian JUHOn laatijoita, Suomalaisen Kirjallisuuden Seura (SKS), Viikin tiedekirjasto, Helsingin kaupungin verkkopalvelu (julkisen hallinnon palveluontologia JUPOn käyttäjiä), Yritys-Suomi ja Museo 2015. Lisäksi myös Kansalliskirjaston sisällä on tehty tiivistä yhteistyötä etenkin sanastotyön asiantuntija Mirja Anttilan ja sisällönkuvailija Liisa Virtasen kanssa. Sanastokeskus TSK:n edustajaa Katri Seppälää on kuultu YSON päivitysprosesseissa ja myös FinnONTO:n aikaisista ratkaisuksista ja periaatteista.³

Haastatellut tahot edustavat varsin heterogeenisiä käyttötärpeita, painottuen mediaan, hallintoon, kulttuurialaan ja erikoiskirjastoihin. Media ei ole ainoa uudenlainen metadatan sisällöntuottajataho, vaan uudenlaisia toimijoita ovat myös esimerkiksi Museo 2015, Kirjasampo ja Yritys-Suomi. Museo 2015 on iso metadatan hyödyntävä järjestelmä, ja Kirjasampo ja Yritys-Suomi verkkoportaalaja. Sisällönkuvailukentällä ovat uutena vahvana trendinä paitsi uudenlaiset järjestelmät, myös se, että sisällönkuvailijat ovat taustaltaan aiempaa heterogeenisempiä ja sisällönkuvailua tuottavat usein kuvailtavan substanssin osajat, kuten tutkijat, virkamiehet, toimittajat ja lisäksi myös kesätyöntekijät. Sen vuoksi myös he, joilla sisällönkuvailu on vain yksi osa muita tehtäviään, ovat haastateltavina runsaasti edustettuina.

Noin puolella haastateltavista on kokemusta erikoisontologian laadinnasta, ja noin puolella haastateltavista on muuten työnsä ja toimenkuvansa kautta kokemusta ontologian käytöstä ja käyttöön kohdistuvista tarpeista. Valtaosalla haastateltavista ei ole ontologiaa ns. sisäänrakennettuna organisaationsa sisällönkuvailun ja/tai tiedonhaun järjestelmään, mutta melkein kaikissa organisaatioissa sellainen on työn alla, suunnitteilla tai vähintään toiveissa.

Haastateltavat suhtautuivat ontologioihin pääosin myönteisesti ja osa määritteli itsensä ontologiamyönteiseksi. He toivat kuitenkin usein esille olevansa tässä asiassa edelläkävijöitä, ja että kirjastoalalla perinteinen ns. tesaurosajattelu vielä dominoi asenteita ja käytäntöjä. Jotkut haastateltavat myös kokivat olleensa toimijoina ja kehittäjinä melko yksinäisiä, ja kaikki kertoivat kaipaavansa selkeää keskitettyä koordinoijaa.

³ FinnONTO-projektin tuottama dokumentointi on ollut lähinnä teknispainotteista (ks. Semanttisen tutkimusryhmän kotisivut, URL:<http://www.seco.tkk.fi/>). Sisällöllisistä ontologisointiratkaisuksista ja periaatteista ei ole ollut saatavilla kattavaa julkista dokumentaatiota, joten Kansalliskirjaston ONKI-projekti on teettänyt FinnONTO:n aikaisesta toiminnasta raportin, joka julkaistaan alkuvuodesta 2014.

“Tarvittaisiin paljon enemmän sitä, että eri projektien kehittäjät istuisivat samaan pöytään, workshoppaisivat. Tämä on kuitenkin niin valtava muutos. Olisi hyvä, jos tämä keskustelu tapahtuisi FINNAn puitteissa, että siellä lähdettäisiin testaamaan näitä ontologioita. Mutta ajatusmalli on monella se, että ensin kehitetään omia isoja järjestelmiä ja sitten katsotaan. Pitäisi lähteä ongelmalähtöisesti ratkaisemaan yhdessä: mitä tarvitaan, kuka tekee mitäkin, miten tämä tehdään fiksusti. Arkistojen ja kirjastojen pomojen pitäisi miettiä, miten tilanteen voisi ratkaista. Päättäjiä pitäisi saada sitoutumaan henkisesti. Tarvittaisiin joku maaginen konsultti.”

3. Sisällönkuvailun käytännöistä ja tavoitteista

Haastateltavat pitivät sisällönkuvailua tärkeänä työnä ja erityisosaamista vaativana. He kuitenkin usein kokivat, että sen hyödyt, hyödynnettävyys ja funktio eivät aina avaudu muille kuin sisällönkuvailun ammattilaisille, ja että ei-ammattilaisten tekemä sisällönkuvailu ei ole laadultaan ammattilaisten tasoa. ”Kaivattaisiin sanastoalan ammattilaisia kertomaan, että kuvailu on relevanttia muillekin kuin erikoisasantuntijalle.”

Sisällönkuvailun tavoitteena on haastatelluissa organisaatioissa sisällönkuvaillun aineiston löydettävyys. Osassa organisaatioista tuotetaan itse myös kuvailtava aineisto, ja primaarilähteen tuottaja on myös sen metadatan laatija – näin on esimerkiksi Svenska YLEllä, jossa toimittajat sisällönkuvailevat itse omat juttunsa.

Haastateltavat erottelivat usein sisällönkuvailun tarpeet ja käytännöt ja analysoivat niitä sen mukaan, onko kyse ammattilaisista ja/tai oman organisaation sisäisistä käyttäjistä vai oman organisaation ulkoisista käyttäjistä. ”Miten kansalaiset käyttävät Ylen sisältöä on ihan eri kysymys verrattuna siihen, miten [sitä] talon sisällä käytetään.” Hakujen on ajateltu kuitenkin yleensä kohdistuvan kaikkiin sanoihin, mukaan lukien asiasanat. Asiasanastojen laadinnassa ja käytössä ajatellaan usein aiheenmukaista löydettävyyttä, mutta myös erityisten instanssien löytäminen on koettu tärkeäksi.

Asiasanastossa – eli tässä tapauksessa ontologiassa – tarvitaan siis myös paitsi temaattisia asiasanoja (eli jotain aihetta kuvaavaa) ja yleisnimityksiä, niin myös mahdollisimman tarkkoja käsitteitä sekä monenlaisia erisnimiä kuten esimerkiksi henkilöä, paikkaa ja yritystä ilmaisevia. Erikoissanastoista kipeimmin haastateltavat kaipasivat paikkaontologiaa eli mahdollisuutta käyttää paikannimiä sisällönkuvailussa ja ajatella niitä hyödynnettävän myös tiedonhakukyselyiden laajentamisissa ja supistamisissa. Ontologian keskeisiä hyötyjä katsottiin olevan, että ontologia tarjoaa runsaasti vaihtoehtoisia ilmaisutapoja samalle käsitteelle, sillä ”tavallinen tallaja ei tiedä eksakteja asiasanaston termejä” ja myös ”avoimen datan entusiasteille asiasanoitus on vierasta; sitä käyttävät etupäässä alan ammattilaiset”.

Organisaatioissa, jotka edustavat perinteisempää kirjastosektoria tai ovat muutoin siihen rinnastettavissa, ajatellaan useammin sisällönkuvailun periaatteita ja etenkin sen yhdenmukaisuutta. Motiivina periaatteille ja yhdenmukaisille käytännöille koetaan tiedon parempi löydettävyys niin saannin kuin tarkkuuden kannalta. Sisällönkuvailuperiaatteista sopiminen ei ole kuitenkaan itsestäänselvyys myöskään perinteisemmällä kirjastosektorilla, vaan käytännöt ovat eläneet ajassa ja toimijoiden mukaan, ja käytetyt periaatteet eivät ole aina välittyneet uusille toimijoille. ”En tiedä, onko meillä koskaan määritelty sisällönkuvailun tavoitteita.” Paljon osaamista on myös ns. hiljaisena tietona, jota ei siis ole julkilausuttu tai raportoitu. Kuvailuohjeiden laatimisen tärkeyteen on kuitenkin moniaalla herätty.

”Mä tuotan luettelointiohjeet kaikille museoammattilaisille. Emme erikseen opeta, miten asiasanoitetaan. Lähtökohtana on tarkin mahdollinen taso. Tuodaan erilaisia näkökulmia esiin kuvailtavasta objektista: konteksti, ulkonäkö jne.”

”Meillä on ollut oma ohjeistus sisällönkuvailuun. Aluksi se oli hiljaista tietoa. Nyt kun kaikissa organisaatiossa ihmiset jäävät eläkkeelle niin niitä on kirjattu ylös, kun uusia ihmisiä on tullut töihin. Nyt kun on tullut tämä valtakunnallinen sisällönkuvailun asiantuntijaryhmä, jossa ollaan jäsenenä, niin ei kehtaa mennä sinne, jos meillä ei noudateta niitä ohjeita. Tässäkin asiassa ollaan siirtymässä toisenlaiseen ajatteluun. Tiettyä logiikkaa täytyy olla erikoisaloilla, että löytää sen aineiston, jota sisällönkuvaillaan. Omiakin ohjeita täytyy olla yleisten ohjeiden lisäksi. Tehdään talon sisällä samalla tavalla, että löydetään oma aineistomme.”

Haastatteluissa esitettiin keskenään ristiriitaisiakin toiveita, jotka heijastuvat käytännön tarpeista ja joiden välillä on vaikea tasapainoilla. - Ei ollut tavatonta, että haastateltava saattoi kaivata eri organisaatioiden välistä yhdenmukaisuutta yhteisen sanaston ja kuvailuperiaatteiden muodossa, mutta eri yhteydessä peräänkuulutti etenkin organisaatiokohtaisia ratkaisuja. Yhteneväisyyden voidaankin ajatella olevan toivottua ja toimivan – mutta vain tiettyyn ja joskus kovin yksilölliseen rajaan asti.

Erisnimien lisäksi sisällönkuvailun prosessista keskusteltaessa esille noussut keskeinen erityistarve on monikielisyys. Kansainvälistyvä asiakaskunta tarvitsisi nykyistä kehittyneemmät sanastovälineet.

”Meillä on paljon englanninkielisiä opettajia, opiskelijoita ja tutkijoita. Suomen kielellähän me indeksoidaan. Toisaalta suurimmalla osalla termeistähän on englanninkielinen käännös Agrovocin tai jonkun muun kautta, mutta nehan eivät tule automaattisesti käyttöön järjestelmässä. Käyttäjän pitäisi etsiä englanninkielinen sana ja hakea vastaavalla suomenkielisellä sanalla, mutta veikkaan että kukaan ei tee näin. Se on liian työlästä. Mutta se on nyt pohjana kehittyneemmälle välineelle.”

Asiasanoittamisen työläys on haastatteluissa yhteisesti tunnustettu asia. Silti sitä ei pidetty turhana työnä, päinvastoin, ja kritiikki koski enemmän sitä kuka asiasanoittaa ja miten. ”--- meillä asiasanoittaminen on hyvin monentasoista. Meillä on paljon harjoittelijoita. --- Toivoisimme, että asiasanoitus tehtäisiin muualla.” Osa mietti vaihtoehtoisia sisällönkuvailutapoja, ja etenkin asiasanojen ketjutusta. Tässäkin asiassa koettiin puutteita itse hakujärjestelmissä, ja järjestelmien puutteiden katsottiin lisäävän kohtuuttomasti intellektuaalista ja manuaalista työpanosta. Käytäntöjen todetaan olevan kuitenkin edelleen hajanaiset, vaikka esimerkiksi ketjutuksesta on yleisesti jo luovuttu.

Osa haastateltavista nosti esille sisällönkuvailun joukkoistamisen. Kulttuuripuolella asia nostettiin esille myönteisimmässä hengessä, mutta siellä silti useammin osana metadatan täydentämistä.

”Meidän pikkutarkka luettelointitapa on jo ehkä vähän väistyvä. Asiasanoituksesta varmasti laistetaan tulevaisuudessa. Jos on iso aineisto niin kuvaillaan vähäisesti ja asiakas tulee sitten paikalle ja tutkii aineistoa itse. Briteissä on hienoja projekteja joukkoistamisessa sisällönkuvailussa.”

Museopuolella käyttäjät voivat täydentää sisällönkuvailutietoja usein esimerkiksi lisäämällä tietoja kuvissa esiintyvistä henkilöistä. Ongelmana on kuitenkin tiedon luotettavuus ja että järjestelmissä ei ole mahdollista tallentaa metatietoon liittyvää informaatiota kylliksi – kuka antoi tiedon, ja milloin.

Kirjastosektorin puolella sisällönkuvailun yhteydessä haastateltavat ottivat usein esille myös tiedonhakijat ja heidän tiedonhakukäyttäytymisensä ja opastuksen.

”Meillä on hyvin monenlaisia tiedonhakijoita. --- Asiakaskunta on hyvin laaja. --- Jos ajatellaan sisällönkuvailua niin tavoitteena on ottaa mahdollisimman monta näkökulmaa huomioon. Hakijoita voi olla hyvin monenlaisia. --- Sehän on hirveän vaihtelevaa tällaisessa asiakaskunnassa, miten osataan käyttää hakuvälineitä. Tiedonhakua opastetaan --- Järjestämme koulutusta ulkopuolisille jonkin verran. ---”

Uutisten sisällönkuvailussa ei ajatella tiettyjä kohderyhmiä. Median puolella arveltiin kuitenkin etenkin nuorempien käyttäjien selailevan sisältöjä myös käsitteiden avulla. Sisällönkuvailtavat jutut pyritään kuitenkin yhdistämään asiasanoihin vastaaviin aiempiin artikkeleihin. Lisäksi osa toimittajista lisää myös suoria linkkejä muihin artikkeleihin, sillä automaation toimivuuteen ei vielä luoteta. Sisällönkuvailukäytäntö onkin varsin toimittajakohtaista. YLEllä on kuitenkin menossa hanke ”Metatieto kuntoon”, ja voidaan olettaa, että hankkeen edetessä ja muutenkin ontologia- ja sisällönkuvailutietouden lisääntyessä yhdenmukaisuus tulee kasvamaan. Median puolella sisällönkuvailun tarkkuuden rinnalla on voimakas pyrkimys muiden, kiinnostavuudeltaan ja aihepiiriltään todennäköisesti relevanttien aineistojen löytyvyys. ”*Ainahan on parempi, että ihmiset lukevat useampia tekstejä.*”

Median kirjastopalvelupuolella sisällönkuvailua tehdään etenkin toimittajia varten, joiden itsenäistä tiedonhakua kannustetaan. Toimittajat tekevät tiedonhakua juttujen taustoitukseen ja faktojen tarkistamiseen.

Kirjasamossa keskitytään kaunokirjallisuuden sisällönkuvailun tarpeisiin. Tavoitteena on teosten löytyminen ja myös teosten kytkeminen toisiinsa eri kriteerein. Ontologia on keskeinen sanastotyökalu ja väline, mutta sisällönkuvailua pyritään täydentämään myös lisäämällä ns. hiljaista tietoa tarjoamalla aktiivisesti kirjastohenkilökunnan laatimia teossuosituksia.

Hallinnon puolella aineisto ja järjestelmät ovat nykyisellään varsin heterogeenisiä, joskin aktiivinen suunta ja tavoite ovat selkeyttäminen ja yhtenäistäminen.

”Sisällönkuvailun tavoitteet riippuu aineistosta ja järjestelmästä. Jos nyt ajatellaan, että se on ensisijaisesti tiedonhakua varten niin tietenkin palvellaan järjestelmän käyttäjiä. Jos nyt ajattelee vaikka meidän asianhallintajärjestelmää, jota käyttää meidän koko henkilöstö mutta ei ulkopuoliset, niin meidän käyttäjät ovat sillä lailla eriytyneet, että melkein jokaisella on oma spesifi asiantuntemusalueensa ja osastoittain ei kovin tarkkaan seurata toisten tekemisiä, mikä tarkoittaa sitä, että terminologia ei ole välttämättä tuttua. Siinä kuvailu auttaa pikkuisen systematisoimaan käsitteistöä niille henkilöille, jotka eivät tunne aihetta tarkkaan.”

Hallinnossa tiedonhakijoina ajatellaan olevan ministeriöiden virkamiehet. Heillä ei ole sisällönkuvailun ammattilaisen näkökulmaa asiaan, eikä sitä myöskään odoteta olevan. Kuten median puolella, myös hallinnossa osassa organisaatioista sisällönkuvailun tuottavat sisällönkuvailun ammattilaiset, osassa primäärilähteiden tekijät. Sisällönkuvailu kärsii paitsi median, myös hallinnon puolella usein motivaatio-ongelmista.

Haastatellut toimittajat kokivat sisällönkuvailuprosessissa tekniikan dominoivan melko paljon. Käytäntöjen todettiin olevan hajanaisia, ja toimittajakohtaiseenkin systematiikkaan vaikuttivat monet muutkin asiat kuin omaksutut sisällönkuvailuperiaatteet. Mediapuolella sisällönkuvailun laatua heikentäviksi todettiin automatiikan puutteet, manuaalisen ja intellektuaalisen työn vähyys, aikapula, sisällönkuvailua tekevien kirjava taso ja jo aiemmin mainitut motivaatio-ongelmat.

”Artikkelia kirjoitettaessa järjestelmä tarjoaa automaattisesti KOKO-käsitteitä tekstin pohjalta. Nämä hyväksytään tai jätetään hyväksymättä. Toimittaja voi myös lisätä omia käsitteitä. Tämä on hyvin toimittajakohtaista: monet käyttävät pelkkää automatiikkaa puhdistamatta lainkaan. Sisällönkuvailu on silloin aika heikkolaatuista, mutta jotkut tekevät tosi laadukastakin sisällönkuvailua. Enemmän kuvailua tekevät kuluttavat siihen artikkelikohtaisesti ehkä minuutin, jos sitäkään.”

”Järjestelmässä on se heikkous, että se ei ehdota kontekstuaalisia termejä eli ei osaa päätellä sellaisia sanoja, joita tekstissä ei suoraan lue. Järjestelmä tarjoaa myös uutisartikkeliin liittyviä muita artikkeleja (relaterat material), joita voi myös selata ja tarkistaa, mitä käsitteitä niissä on käytetty. Relaterat material -toiminto on vielä aika karkea, mutta toimii jo nytkin hyvin.”

Järjestelmän oleminen suorassa yhteydessä ONKIin koettiin suureksi avuksi autocomplete-toiminnon ehdottaessa käsitteitä. Kaikki tarvittava kehitystyö ei siis tapahdu käsitteiden valinnassa ja ilmaisussa, vaan järjestelmien katsottiin olevan elimellinen osa sisällönkuvailuprosessia ja laatua. Lisäksi silloin, kun aineistolle on valmiit asiasanaehdotukset, käydään ne vielä manuaalisesti läpi, jotta oman organisaation tarpeet ja käytännöt tulevat paremmin otetuiksi huomioon.

”Jos tulee uusia kirjoja niin sieltä valitaan BTJ:n tiedot ja vähän karsitaan ja lisätään asiasanoitusta. Tavoitteena on, että toimittajat löytäisivät paremmin.”

Se, että ONKI-palvelun vuoden 2013 versiossa ei ole erisnimiä systemaattisesti saatavilla, koettiin puutteeksi ontologian sisällössä, joka vaikeuttaa ja hidastaa sisällönkuvailuprosessia. Lisäksi ongelmia aiheuttaa sisällönkuvailuperiaatteiden, -käytäntöjen ja järjestelmien muuttaminen, monen eri version jäädessä elämään rinnakkain.

”Kuvailuprosessi tuntuu pikkuisen liikaa aikaavievältä. Joudutaan pähkäilemään paljon sellaista, mitä ei ONKIsta löydy: tapahtumat, organisaatiot, henkilöt, paikat.”

”Muuttuvia paikannimiä on yritetty ratkaista erilaisilla luovilla tavoilla. Asiakkaat eivät näitä merkkaukseta tunne.”

Haastatteluissa esille nousseet sisällönkuvailuun liittyvät haasteet ovat moniaalla. Yksi keskeisiä haasteita ja käytön esteitä ja/tai laatuun vaikuttavia puutteita on se, että käytössä on ollut FinnONTO-projektin aikainen versio, jota ei ole Aalto-yliopiston palvelimella ja tutkimusprojektissa päivitetty vastaamaan YSAn nykytilannetta. Kaikki haastateltavat ottivat esille Aalto-yliopiston tutkimusprojektin aikaiset palvelinongelmat. (Järjestelmistä ks. luku 4.) Ongelma on myös se, että järjestelmä ei useimmiten tue ontologian periaatteessa mahdollistamia kyselyjen supistamisia ja laajentamisia, joten sisällönkuvailija joutuu toimimaan kankeiden ja ominaisuuksiltaan turhan yksinkertaisten järjestelmien ehdoilla. Sisällönkuvailijaa kuormittaa, kun kuvailu on ”*pähkäämistä yleisemmän ja tarkemman tason välillä*” – ”*jos ontologiat saataisiin järjestelmiin sisälle niin että ne toimisivat niin että voisi supistaa ja laajentaa hakua niin sehän olisi mielettömän upeaa*”.

”On monia pähkäilyjä. Käytetäänkö kuvailussa sekä ’koiraa’ että ’eläimiä’? Sekä ’taidehistorioitsijoita’ että ’taidehistoriaa’? Paljonko pitää hokea samaa ja minkä asian kautta tuon kertoo, ammattinimikkeen vai aiheen kautta? Yleensä asiasanoitetaan mielialan mukaan, kun ei näitä ole sovittu yleisesti. Mietittiin vihta-aiheista artikkelia, jota kuvailtiin ’vihdat’-käsitteellä. Vähän mietittiin jotain käden taitoja, mutta ne voisivat linkittyä johonkin ihan eri asiaan ja tuoda koomisia tuloksia.”

”Käytämme samassa hierarkiassa olevia asiasanoja päällekkäin esimerkiksi, kun kyse on tietystä hyönteislajista niin kyllä siinä mielellään laittaa ’tukkimiehen täi’, ’kovakuoriaiset’ ja ’hyönteiset’. Jos joku haluaa esim. tietoa jonkun torjunta-aineen vaikutuksesta hyönteisiin niin silloin se hyönteiset on siellä oltava. Eihän hakija voi luetella niitä miljoonia lajeja, jotka voisi tulla kyseeseen. Tällaisissakin tapauksissa olisi olennaista saada se ontologia. Onhan se tavallaan tyhmää syöttää sisällönkuvailussa kaikki nämä käsitteet sinne, eikä sitä varmaan kaikki teekään. Mutta kyllähän se tiedonhakua auttaa.”

Median puolella haaste on myös sisällönkuvailun merkittävyyden hahmottamisessa ja siinä, kenen sitä kuuluisi tehdä ja miksi, ja että vaikka sanastollinen apuväline kaipaakin vielä kehittämistä, niin paremman odottaminen ei saisi nousta kynnyskysymykseksi sen nykyiselle käytölle.

”Itse kuvailulla ei sinänsä ole huono maine, mutta se ei ole ydintoimintaa. TV ja radio ovat talon sisällä prioriteetteja. Verkkosisältö on ekstraa ja sisällönkuvailu on siinä vielä lisähomman lisähomma. Joillakin saattaa olla asenneongelma, kun ei ole henkilönnimiä tarjolla eikä serverikään toimi.”

Haaste on myös sisällönkuvailuvälineen näkökulma ja hajanaiset periaatteet. Semanttisessa webissä tulisi osin paradoksaalisesti sallia kaikenkirjavuus, mutta kuitenkin taustalla pitäisi olla yhteneväisyys, jotta eri toimijoiden väliset eroavaisuudet olisivat käytettävyyden nimissä ylitettävissä. Ilmiö on esillä niin median tarpeista kuin erikoiskirjastojen omista näkökulmista puhuttaessa. Vaikeuksia tuottaa tyypillisesti esimerkiksi se, että miten kuvailla jotain asiaa

negaation kautta (kun jokin ominaisuus puuttuu; entiset ministerit) tai silloin kun tuttu ilmiö esiintyy poikkeavasti (perheväkivalta, jossa mies uhrina).

”Ongelmana on liiallinen spesifiointi. Ihmiset unohtavat sen kokonaisuuden, josta se on osa. Se ei välttämättä löydy, kun se on indeksoitu niin spesifisti, että kukaan ei vielä tiedä, että sellaisia asioita tutkitaan.”

”Meillä on omia kuvailusääntöjä. Nyt tuntuu, että verkkomaailmassa kaiken pitäisi olla yhteensopivaa. Nyt siitä vähän kärsitään, että aiemmin on tehty vain omiin tarpeisiin. Meillä on avoimempia toimijoita kuten kirjastopalvelu, jotka ovat enemmän yhteismitallisia ulkomaailman kanssa ja sitten on yhtiön omiin kokoelmiin suuntautuneita toimintoja.”

Haastateltavat kokivat asiasanoituksen funktiona olevan tiedon myöhempi löydettävyyys. Sen toteutuksen tapa ei kuitenkaan ole haastateltavien mukaan aina selvä asiasanoja valitseville ammattilaisillekaan. Mikäli järjestelmä olisi kyllin kehittynyt, auttaisi se myös sisällönkuvailijaa arvioimaan omia valintojaan. Yksilölliset erot sisällönkuvailussa olivat yleisesti tiedossa – ja myös sisällönkuvailijan sisäinen epäjohtonmukaisuus. Näihin ongelmiin koettiin avuksi sanastollinen kontrolli ja kehittyneempi tallennusjärjestelmä.

”Monille tuntuu olevan tärkeää käyttää itse keksimiä asiasanoja. Kirjoittamisessa tallentaa aineistoa järjestelmään. Siinä vaiheessa ei ehkä osaa ajatella kaikkia muita käyttäjiä ja mitä termejä ne tarvitsevat. Siinä olisi mahdollinen integraation paikka, että asiasanoittaessasi pääsisit tarkistamaan ONKI-palvelusta sen.”

”Sisällönkuvailu on persoonalla tehtävää työtä. Siihen vaikuttaa kokemus ja kaikki muukin. Ohjeet on lempeä tapa ohjata sisällönkuvailua. Ollaan kyllä menossa järjestelmässä admin-tasolla tiukempaan tapaan toimia: keskitettyä järjestelmää haetaan ja sieltä tulisi esimerkiksi materiaalienttien nimet.”

Kaunokirjallisuuden puolella ongelmia aiheuttaa tasapainoilu siinä, miten informatiivista sisällönkuvailu voi olla – kattava ei olekaan enää välttämättä tiedonhakijan etu, jos sisällönkuvailutiedot vaikkapa paljastavat romaanin juonellisia seikkoja, kuten murhaajan. Toisaalta pitäisi palvella esimerkiksi heitä, jotka puolestaan haluavat hakea veljesmurhia käsittelevää kaunokirjallisuutta. Tai entä miten tuottaa sisällönkuvailua, joka tukee esimerkiksi konservaattorin tiedontarpeita ja työtä, ja samalla vaikkapa tietynaiheisen kirjallisuuden hakijaa. Useat haastateltavat ottivat puheeksi tässäkin yhteydessä sisällönkuvailuprosessiin merkittävästi vaikuttavat käyttöliittymän ja haku- sekä tallennusjärjestelmän. Niihin tulisi panostaa nykyistä enemmän, ja esimerkiksi konservaattorin näkymä voisi olla erilainen ja hänellä olla osin erilaiset kentät käytössä kuin ns. tavallisella tiedonhakijalla. Myös mahdollisuutta voida räätälöidä näkymiä paitsi käyttäjäryhmä- niin myös yksilökohtaisesti toivottiin – kaikki kentät eivät ole kaikille tarpeellisia.

Sanastovälineiden, kuten YSO-ontologian, kehittämistä auttaisi se, jos kussakin organisaatiossa seurattaisiin sanastoonsa kohdistuvia tosiasiallisia tarpeita ja käytäntöjä. Moni sidosryhmätaho nosti aiheen esille, mutta totesi myös, että se on varsin hankalaa, jos ei peräti mahdotonta, aikapulasta ja/tai käytössä olevasta järjestelmästä johtuen. Ongelmia aiheuttavat myös organisaation asettamat paineet ja tavoitteet sisällönkuvailulle, ja niiden koetaan joskus olevan ristiriidassa käytettävyyden ja laadun tavoittelun kanssa. Lisäksi esille nousi useassa yhteydessä järjestelmien suunnittelussa esiintyvä kuilu sisällönkuvailuammattilaisten ja teknisten asiantuntijoiden välillä – ensin mainitut eivät aina kokeneet tulleen kylliksi tai pahimmassa tapauksessa lainkaan kuulluiksi järjestelmien suunnitteluprosesseissa.

”Hakusanastoksi se tehtiin, mutta sinne tuli paljon sellaisiakin sanoja, joita ei koskaan käytetä indeksoinnissa. Niitä tarkasteltiin sillä lailla, että meillä on tämäntyyppistä toimintaa, vaikka siitä ei ikinä julkaistaisi mitään. Siellä on varmaan paljon tämmöistä.”

”Olisi mielenkiintoista tutkia sanaston käyttöfrekvenssiä. Sitä ei tuosta systeemistä ilmeisesti tällä hetkellä saa ulos.”

”--- nykyään on tajuton pyrkimys yksinkertaistaa asioita, mikä on sinänsä sääli.”

”Haavetilanne olisi, että sellaiset ihmiset, jotka ymmärtäisivät jotain metatiedosta myös suunnittelisivat näitä järjestelmiä tai ainakin olisivat prosessissa mukana.”

Huolimatta sisällönkuvailun ja sen välineiden – kuten ontologioiden – haasteista, asiasanoittamisen ja ontologioiden kehittämisen koettiin kuitenkin olevan elimellisen tärkeää. Erimielisyyttä oli siinä, kenelle kuuluu käytännön vastuu asiasanoittamisesta - vastuuta niistä pyritään siirtämään toiselle organisaatiolle tai toimijataholla. Sisällönkuvailun katsottiin edellyttävän kuitenkin nimenomaan ammattitaitoa.

”Tietopalvelu ja asiakaspalvelu on meidän tärkein työtehtävä, mutta sen taustaksi tarvitaan sisällönkuvailua, johon menee suurin osa työajasta.”

4. Järjestelmistä

Haastateltavien tahojen järjestelmät palvelevat monia erilaisia toimialoja ja tarkoitusperiä ja muodostavat täten varsin kirjavan kokonaisuuden. Nimitämme tässä järjestelmiksi niin sisällönkuvailutyökaluja ja -ympäristöjä kuin loppukäyttäjälle tarkoitettuja aineiston selailu- ja hakupalveluitakin. Osa organisaatioista hyödyntää ontologioita jo nyt järjestelmissään, osa suunnittelee tai on kiinnostunut niiden hyödyntämisestä. Osassa järjestelmistä on käytössä asiasanasto ontologian sijaan. Täysimittainen ontologioiden hyödyntäminen edellyttää, että järjestelmä käyttää ontologiakäsitteiden yksilöityjä tunnisteita ja on yhteydessä ONKI-palveluun, josta sisällönkuvailuun tarvittavat käsitteet haetaan. Ontologiaa voidaan käyttää myös referenssinä vapaasanaisessa sisällönkuvailussa, mutta tällöin ontologiateknologiasta ei koidu järjestelmälle tai käyttäjälle mitään konkreettista hyötyä.

Myös käytössä olevat ontologiat ja asiasanastot vaihtelevat. Osa järjestelmistä käyttää KOKO-ontologiaa, osa YSOa. Osassa järjestelmistä on käytössä organisaation oma ontologia: nämä erikoisontologiat on ripustettu YSOon, jolloin yhdistelmä sisältää erikoisontologian lisäksi YSO:n kokonaisuudessaan. Toisaalta on myös useampi taho, joiden asiasanasto on ontologisoitu, mutta jotka tästä huolimatta käyttävät yhä järjestelmässään asiasanastoa ontologian sijaan.

Pääsääntöisesti voidaan sanoa, että ontologioiden hyödyntäminen on pisimmällä niissä organisaatioissa, joilla on joko a) omasta takaa vahvat tekniset resurssit ja ontologiaosaamista tai b) ovat olleet tiiviissä yhteistyössä FinnONTO-projektin kanssa saaden teknistä tukea sieltä.

4.1. Ontologioiden hyödyntämisestä järjestelmissä

Haastatteluissa tuli ilmi monia erilaisia tapoja, joilla järjestelmät hyödyntävät ontologioita. Koska järjestelmät on kehitetty yksilöllisesti organisaatioiden omiin tarpeisiin, kukin järjestelmä soveltaa ontologioita omalla tavallaan. Soveltaminen voidaan jakaa kolmeen kohdealaan: sisällönkuvailuun, tiedonhakuun ja aineiston selailuun. Erillishuomautuksena voidaan mainita, että ontologia tarjoaa tyypilliset kontrolloidun sanaston tuomat edut (sisällönkuvailun yhdenmukaistaminen) niihin järjestelmiin ja kuvailukohteisiin, joissa sellaista ei ole aiemmin ollut käytössä.

“Datan loppukäsittelijän roolista katsottuna kokoelmahallintajärjestelmässä olisi hyvä sitoa sisällönkuvailijoiden kädet johonkin asiasanastoon, auktoriteettilistaan niin että olisi vain ihan muutama kenttä, johon saisi vapaasti kirjoittaa. Ihmisillä on kovin erilaisia mielipiteitä, mitä mihinkin kenttään kuuluu laittaa. Se hankaloittaa datan eteenpäin laittamista.”

Ontologian integroiminen sisällönkuvailujärjestelmään mahdollistaa ns. autocomplete-toiminnon: se ehdottaa automaattisesti ontologian käsitteitä kenttään syötettyjen kirjainten perusteella. Vastaava toiminto on käytössä mm. Google-hakukoneessa. Sisällönkuvailijat ovat kokeneet tämän auttavan työtä. Toiminnon avulla voi esim. selata tietynalkuisia sanoja: *“Järjestelmässä on autocomplete. Voi yrittää kirjoittaa ja katsoa löytyykö esim. ekonomi-alkuisia sanoja. Se on iso*

apu.” Valmiina tarjottujen käsitteiden käyttö myös yhtenäistää sisällönkuvailua, hävittää väärinkirjoitusongelman ja poistaa vaivalloisen työvaiheen, jossa käsitteet kopioidaan yhtäältä ja liitetään toisaalle. Svenska Ylellä autocomplete huomioi myös ohjaustermit.

”Se on hyvä ominaisuus, että autocomplete tunnistaa synonyymit. Se käynnistää ihmisessä uuden assosiaatioketjun. Voi löytää vielä parempia termejä.”

Ontologian hierarkiarakennetta on myös hyödynnetty sisällönkuvailuvaiheessa: järjestelmä näyttää käsitteen lähihierarkian, jolloin käsitteen merkitys täsmentyy ja esim. homonymiatapaukset erottuvat toisistaan. Vastaavasti osassa järjestelmiä ontologiakäsitteen klikkaaminen avaa linkkinä kyseisen käsitteen hierarkianäkymän ONKI3-palvelussa. Vieruskäsitteitä ja assosiativisia suhteita ei missään järjestelmässä tällä hetkellä näytetä sisällönkuvailijalle, mutta haastatteluissa todettiin, että niidenkin visualisoinnista saattaisi olla apua.

Sisällönkuvailussa on sovellettu myös Aalto-yliopistossa kehitettyä ARPA-työkalua, joka tekee automaattista sisällönkuvailua peilaten tekstissä esiintyviä sanoja ontologiakäsitteisiin. ARPAn antamat tulokset on kuitenkin koettu melko epätarkoiksi ja osin käyttökelvottomiksi. ARPA ei siis tuota tällä hetkellä kovin laadukasta sisällönkuvailua, mutta se on silti koettu hyödylliseksi käsin tehtävän sisällönkuvailun apuvälineenä.

”Käytän automatiikkaa melko harvoin. Teen enemmän manuaalisti. Jos automatiikka antaa esim. seitsemän käsitettä niin jätän ehkä yksi tai kaksi.”

Tiedonhaussa ontologiaa hyödynnetään siten, että ontologiakäsitteillä voi tehdä hakuja aineistoon. Autocomplete-toiminto helpottaa tässä tiedonhakijaa, jonka ei tarvitse tuntea ontologiaa ennalta. Toisaalta hakua voi tehdä myös ilman autocomplete-toimintoa, jolloin haku kohdistuu vapaasanaisesti dokumenttien metadataan ja juoksevaan tekstiin. Joidenkin järjestelmien haussa hyödynnetään myös ontologian ohjaustermejä, eli ohjaustermillä haku tuottaa tulokseksi suositeltavalla termillä kuvaillut dokumentit. Autocomplete-toiminnon hyödyllisyyden ohella on myös pohdittu, antaako YSO:n kokoinen ontologia hakijalle liikaa vaihtoehtoja. Yksi ratkaisu tähän runsaudenpulaan on se, että autocomplete huomioi vain ne käsitteet, joita on todellisuudessa käytetty haun kohteena olevan aineiston kuvailussa. Tällaista mallia onkin jo sovellettu Yritys-Suomi-portaalissa hyvin tuloksin.

”Hakukenttään tulee lista niistä asiasanoista, joilla on jotain sisältöä järjestelmässä. Edellisessä versiossa sieltä tuli kaikki YSO-käsitteet ’eläimiin sekaantumisesta’ lähtien. Se ei ollut kovin järkevää. Nyt täältä voi joko valita asiasanan, josta tietää saavansa sisältöä tai hakemaan vapaasanahauulla, jonka pitäisi kohdistua sekä asiasanaan että tekstisisältöön.”

Monet organisaatiot ovat kiinnostuneita haunlaajennuksesta, jossa yksittäisen käsitteen lisäksi haku huomioi myös haettavan käsitteen hierarkkiset alakäsitteet: “Ontologialla on paljon mahdollisuuksia hakutoiminnon kautta: jos hakee tuolia niin voi saada muitakin huonekaluja.” Haastateltujen tahojen

järjestelmissä ei haunlaajennusta ole vielä käytössä. Monet pitävät kuitenkin sitä ontologian keskeisenä etuna.

Selailussa ontologiaa käytetään tuottamaan suosituksia (eli nostoja tai ehdotuksia), joita järjestelmä tarjoaa ontologiakäsitteiden perusteella: ehdotetaan dokumentteja, joiden kuvailussa on käytetty samoja käsitteitä. Dokumenttien ehdotusprosessin taustalla on olemassa erilaisia relevanssialgoritmeja. Tässä suhteessa pisimmälle kehitetty järjestelmä on Yritys-Suomi-palvelu, jonka suositukset huomioivat myös ontologian hierarkian eli suosituksiin nousee myös ylä- ja alakäsitteillä kuvailtuja dokumentteja – nämä tosin saavat pienemmän arvon suosituksissa kuin täsmälleen samat käsitteet. Myös assosiatiivisia suhteita on kokeiltu suosittelussa. Muiden järjestelmien suositukset eivät tällä hetkellä hyödynnä ontologian hierarkiaa tai assosiatiivisia suhteita, mutta kiinnostusta tämänsuuntaiseen kehitykseen on.

Svenska Ylellä ontologiaa on hyödynnetty siten, että järjestelmä näyttää sisällönkuvailijalle haettujen käsitteiden hierarkian. Lisäksi loppukäyttäjälle tarkoitettussa järjestelmässä kullakin kuvailussa käytetyllä käsitteellä on oma avoin verkkosivunsa, joka kokoaa automaattisesti yhteen kaiken sisällön, jonka kuvailussa kyseistä käsitettä on käytetty. Nämä koostesivut ovat osoittautuneet myös varsin suosituiksi:

”Kunnallisvaalien aikaan ‘kunnallisvaali’-asiasanan sivu sai korkeamman hakutuloksen kuin meidän oikea vaalisivusto. Piti rakentaa asiasanasivulle linkitys meidän vaalisivustoon.”

4.2. Ontologioita hyödyntämättömistä järjestelmistä

Ne järjestelmät, jotka eivät tällä hetkellä hyödynnä ontologioita, käyttävät sisällönkuvailussa pääsääntöisesti joko YSAa, organisaation omaa asiasanastoa tai molempia. Asiasanasto on joko integroitu järjestelmään tai sitä käytetään hieman työläämmällä kopioi-liitä-menetelmällä nettisivulta. Sanastoja käytetään sisällönkuvailun lisäksi myös tiedonhaussa hakemalla sanastosta haussa käytettäviä termejä.

Järjestelmän takana olevan tekniikan ja kuvailukielen ohella järjestelmän toimivuuteen ja aineiston löytyvyyteen vaikuttaa myös sisällönkuvailun laatu. Huonolaatuisen kuvailun riski kasvaa, jos se ei perustu yhteneväiseen ja kontrolloituun asiasanastoon, vaan ennemminkin sen ihmismuistinvaraiseen versioon, ja jos siitä vastaavat ei-ammattilaiset, jotka eivät tunne kuvailukäytäntöjä. Lopputuloksena saattaa olla kirjoitusvirheitä ja erilaisia kirjoitusasuja samasta termistä, tai sama käsite on sanallistettu eri tavoilla. Tällainen kuvailu on ongelmallista loppukäyttäjän kannalta, sillä vapaasanaiseen kuvailuun nojaava järjestelmä ei osaa yhdistää erilaisia kirjoitusasuja.

”Se on ollut käsinnaputtelua ja on tullut variaatiota, väärinkirjoittelua, monikkoja ja yksikköjä sekaisin, aina kaikki eivät tietenkään niitä tarkasta. - - Nyt järjestelmä antaa

niitä sanoja, joita siellä jo esiintyy. Jostakin sanasta voi olla jopa seitsemän eri kirjoitusasua. Niitä joskus korjailaan.”

Useimmat niistä haastateltavista tahoista, joiden asiasanasto on ontologisoitu, käyttävät ontologian sijaan yhä omaa asiasanastoa järjestelmässään. Tilannetta voidaan selittää kuudella syyllä:

- 1) Ontologisointi on tapahtunut tyypillisesti projektiluontoisesti FinnONTO-projektin tukemana ja rahoittamana. Usein itse työ on tehty pääosin FinnONTO:n puolella erikoisontologiaa hallinnoivan organisaation toimiessa lähinnä avustavan asiantuntijan roolissa. Kun ontologisointi on päättynyt ja sanastoyhteistyö FinnONTO:n kanssa vähentynyt, organisaatio ei ole välttämättä itse kyennyt osoittamaan ontologialle ylläpitoresurssia. Päivitys on vaatinut teknisiä työkaluja ja konsultointia. Toisaalta oman asiasanaston ylläpito on voitu hoitaa itsenäisemmin, tutuilla työkaluilla;
- 2) Asiaan on vaikuttanut myös YSON aktiivisen päivittyvyyden luoma synkronointipaine: mikäli erikoisontologiaa ei päivitetä säännöllisesti YSOssa tapahtuvia muutoksia vasten, siitä tulee nopeasti vanhanaikainen;
- 3) Ontologisoinnin lopputulokseen ei välttämättä ole oltu täysin tyytyväisiä. Jotkut ovat nähneet sanaston muuttuneen prosessin myötä käyttäjien kannalta epäedulliseen suuntaan.
- 4) Kaikilla organisaatioilla ei ole ollut käytössään teknistä osaamista ja resursseja, joilla ontologia olisi voitu ottaa käyttöön heidän järjestelmässään;
- 5) Osalla ei ole ollut selvää näkemystä siitä, mitä konkreettisia hyötyjä ontologian käyttöönotto tuo heidän järjestelmäänsä asiasanastoon verrattuna.
- 6) FinnONTO-hankkeen loppuvaiheessa ennen Kansalliskirjaston ONKI-projektin aloittamista ontologiapalvelun jatkuvuus ja rahoitus oli pitkään epäselvää. Toisin sanoen ei ollut varmuutta siitä, että erikoisontologioiden kehitystyön käytännössä edellyttämä keskitetty palvelu olisi ylipäätään pystyssä. Nämä tekijät erikseen ja yhdessä ovat todennäköisesti vaikuttaneet siihen, että ontologia on jäänyt käyttämättä ja ylläpitämättä.

”Kyllä välillä tulee aina kyselyjä, kun ministeriöt uusivat verkkosivujaan, että onko teillä ontologiaa tarjolla. Se on aina kilpistynyt päivityksen puutteeseen ja tekniseen tukeen. Tilanne on ollut tällainen. JUHO on olemassa, mutta kun sitä ei päivitetä niin ihmiset ovat sitä vierastaneet. Sitä saatetaan harkita, kun saadaan päivitys kuntoon.”

”Ontologian tilanne on hunningolla. - - Olisi aikamoinen lista päivitettävää. Mutta se oli aika turha alkaa niitä päivittämään ennen kuin sain sen ONKI-YSON, koska siellä ONKI-YSOssa oli moni asia eri lailla kuin Protégé-YSOssa. Tilanteessa heijastui hyvin se, että oli tutkimushanke, jolla on omat intressit, ja nyt on tuotantokäytössä oleva palvelu, jolla on omat tavoitteensa.”

4.3. Teknisistä ja organisatorisista haasteista

Haastatteluissa kartoitettiin erilaisia nykyisten järjestelmien haasteita ja erityisesti sen tyyppisiä haasteita, joita ONKI-projektissa olisi syytä huomioida ja joihin ONKI-projektin kautta voisi etsiä ratkaisua. Tekniikan ja järjestelmien luomien haasteiden lisäksi nousi esiin haasteita, jotka liittyvät organisaation toimintakulttuuriin, historiaan tai rakenteeseen.

Yleisiä järjestelmiin liittyviä ongelmia on esim. se, että ne saattavat olla jo iäkkäitä ja nykyhetken tarpeisiin nähden vanhentuneita. Organisaation sisällä saattaa olla useampi eri järjestelmä, kun tehokkaampaa ja yhteistyötä helpottavaa olisi hankkia kaikille sama järjestelmä. Järjestelmä uudistukset ovat kuitenkin pitkällisiä projekteja ja päätöksenteko monipolvista. Toinen yleinen ongelma on se, että järjestelmien ylläpitoon ei ole resursseja. Järjestelmä on usein tilattu ulkopuolelta ja yhteistyö järjestelmäkehittäjän kanssa on päättynyt tai se on haastavaa erilaisista syistä. Järjestelmään ei voida tehdä haluttuja korjauksia tai lisäyksiä: tästä seuraakin helposti edellä mainittu järjestelmän ikääntyminen.

”Meillä on käytännössä yksi koodaaja, joka on täysin ylityöllistetty. Se hidastaa tosi paljon. Kuukausia ollaan odotettu uutta rajapintaa. Järjestelmä on tämän tekijän takana ja toivoisin, että hän palkkaisi apulaisen.”

Ilmeisen selvänä teknisenä puutteena eräät tahot mainitsivat sen, että järjestelmä ei ole yhteydessä ONKI-palveluun, vaikka ontologiaa käytetäänkin referenssinä: tällaisesta kontrolloituun sanastoon linkittymättömästä sisällönkuvailusta seuraa monia ongelmia kuten kirjoitusvirheet ja erilaisten kirjoitusasujen esiintyminen. Tällaisissa järjestelmissä ja eräissä muissa todettiin sisällönkuvailun kannalta ongelmaksi myös, että järjestelmä ei tue autocomplete-toimintoa eikä ontologian avulla tapahtuvaa hierarkkisten tai assosiativisten käsitteiden ehdotusta. Näiden toimintojen lisääminen katsottiin hyödylliseksi.

Eräät tahot mainitsivat haasteeksi metatiedon heterogeenisyyden organisaation sisällä tai yhteistyötä tekevien organisaatioiden välillä. Metatiedon harmonisointi karsisi päällekkäistä työtä osastojen ja organisaatioiden välillä.

”Monelle asialle on oma tietojärjestelmänsä. Se on konkreettinen syy, miksi uutta sanastoa ei voida ottaa käyttöön, kun ei haluta investoida jonkin yksittäisen pikkujärjestelmän muutokseen. Asiat saattaa jäädä roikkumaan sen takia, että odotetaan jotain kaiken pelastavaa järjestelmä uudistusta. Järjestelmien yhdistämisestä on ollut puhetta, mutta se ei ole hetkessä tapahtuva muutos.”

Yleinen huomio on, että organisaatiot saavat järjestelmistään melko vähän palautetta käyttäjiltä. Kyseessä ei välttämättä ole käyttäjien passiivisuus, vaan palautteen antaminen on saatettu toteuttaa monimutkaisesti, palautejärjestelmä ei toimi kunnolla tai sitä ei ole ylipäättäen. Eräs epäsuora tapa saada tietoa käyttäjistä on tarkkailla järjestelmässä tapahtuvaa hakua: tästä näkee, millä käsitteillä tai asiasanoilla järjestelmästä haetaan tietoa. Kiinnostavia ovat myös ns. nollahaut eli ne, jotka eivät tuota lainkaan hakutulosta. Näiden avulla voi seurata, minkälaisia (virheellisiä) kirjoitustapoja

hakijat käyttävät tai mitä käsitteitä sanastosta vielä puuttuu. Ne organisaatiot, jotka hakuseurantaa tekevät, ovat kokeneet asian hyödylliseksi.⁴

”Mehän katsomme koko ajan niin sanottuja nollahakuja meidän palvelussa. Katsotaan, millä ihmiset ovat yrittäneet löytää jotain, mutta eivät ole löytäneet mitään. Se on aika arvokastakin tietoa. Siinä tulee esiin, että millä nimellä jokin tietty asia pitäisi ihmisten mielestä siellä olla. Tämä ontologia voisi ehkä auttaakin asiassa: jos on talousarvioehdotus niin ne kirjoittavat sinne ihan varmasti budjetti. Ja sitten voisi antaa hakutulokseksi niitä talousarvioita.”

Hakutoimintoon liittyy erilaisia haasteita. Näitä ovat esim. hakunäkymän sekavuus, järjestelmän hitaus sekä hakutulosten esittäminen selkeästi ja hakijaa palvelevasti. Ongelmaksi on koettu, että hakujärjestelmää ei ole kytketty ontologiaan: järjestelmä ei tällöin ehdota hakutermejä eikä väärinkirjoitettuja hakuja korjata. Käyttäjät saattavat kokea haun sekavaksi eivätkä tiedä, mitä sanoja haussa pitäisi käyttää. Ontologiaa ei välttämättä ole integroitu hakuun, vaikka aineisto olisikin kuvailtu ontologialla. Myös ontologiahierarkian ja synonyymien hyödyntämättömyys haussa on koettu puutteeksi ja kiinnostusta näiden toimintojen lisäämiseen olisi. Monet ovat huomauttaneet myös, että ONKI-palvelu on toiminut hitaasti tai on ollut kokonaan poissa käytöstä. Tästä lisää luvussa 6.2.

”Järjestelmässä ei ole mitään sellaista, että jos hakee väärällä asiasanalla niin se sumeasti ehdottaisi, että ’kirjoitit näin mutta tarkoittiko tätä?’. Tai ’tällä haulla ei löytynyt mitään mutta hae tällä’. Järjestelmä ei vie tiedonhakijaa eteenpäin eli siinä on samat synnit kuin muissakin järjestelmissä.”

Osa selailujärjestelmistä käyttää suosituksia eli ehdottaa selaajalle samankaltaista aineistoa. Järjestelmä poimii ehdotettavan materiaalin samankaltaisen metatiedon, mm. samojen ontologiakäsitteiden, perusteella. Myös ontologian käsitesuhteita, hierarkkisia ja assosiativisia, hyödynnetään suosituksissa. Ontologian vieruskäsitesuhteita ei tällä hetkellä hyödynnetä järjestelmissä, koska YSOssa monet korkeammalla tasolla olevat vieruskäsitesuhteet eivät ole temaattisesti koherentteja; jollain tarkemmalla tasolla vieruskäsitteetkin saattaisivat nostaa mielekästä sisältöä suosituksiin. Niissä järjestelmissä, joissa suositukset ovat käytössä, kokemukset ovat olleet positiivista, mutta suositusjärjestelmät kaipaavat vielä kehittämistä.⁵

”Suosittelemetrit on monimutkaisia. Mutta niin karkea kuin se suosittelu onkin niin on hyvin toiminut. - - - Varsinkin jos rakentaa kokonaisuuksia, joissa on monta eri tekstiä. Jos laittaa samoja asiasanoja eri teksteihin niin saa hyvän listan suosituksia.”

⁴ ONKI-projektissa on seurattu VESA-verkkosanaston hakuja. Seuranta on paljastanut paljon kiinnostavaa tietoa hakutottumuksista: esimerkiksi VESAn katkaisuhakukäytäntö ei ole kaikille tuttuja ja yksikkömuodolla hakeminen on hyvin yleistä, vaikka monet asiasanat ovat monikkomuodossa.

⁵ Kirjasampon käyttäjäpalautteessa todettiin järjestelmän antamat automaattiset kirjasuosittelut hyödyllisiksi; kirjastonhoitajien antamat käsityönä tehdyt suositukset todettiin vielä hyödyllisemmiksi, mutta ero ei ollut dramaattinen. (Soikkeli, Virva 2013: ”Kirjasampokyselyn raportti”. Vaasan kaupunginkirjasto-maakuntakirjasto.)

Ongelmana on, että annetut suositukset ovat välillä epämielekkäitä, kontekstiin sopimattomia tai suorastaan epätoivottuja. Ontologiaa hyödynnettäessä assosiatiiviset tai alakäsitteisiin perustuvat suositukset tuottavat välillä epärelevantteja suosituksia. Osittain tilanteeseen vaikuttaa assosiatiivisten käsitteiden luonne yleisesti: käsitteiden yhteys ei ole absoluuttisesti tosi, vaan tulkinnallinen. Eri konteksteissa syntyy eri assosiaatioita, joita ei voi yleistää kaikkialle. Toisaalta YSOssa esiintyvät yleisluonteiset assosiaatiot eivät välttämättä ole relevantteja jonkin erityisalan kannalta:

”Nyt ollaan päätetty, että otetaan pois assosiatiiviset käsitteet, koska se on tuonut meille vaikeuksia, ikävä kyllä lähinnä YSON kautta. On noussut kysymys, että kuka päättää YSON assosiatiiviset suhteet. Onko ne määritelty vain haun tai kirjaston näkökulmasta? Henkilökuntaeduille on laitettu assosiatiivinen suhde palkat ja työterveyshuolto. Sinne nousee myös työterveyshuoltoasiaa esille, jota me emme halua, koska meillä on niille oma sivu.”

Myös sama käsite saattaa tuoda esiin epätoivottuja suositustuloksia: esim. ’suoramyynti’ liittyy kansainvälistyvissä yrityksissä aivan eri tematiikkaan kuin kotimaan pienyrityksissä ja suositus voi tämän käsitteen kautta tarjota epärelevanttia sisältöä. Tämänkaltaisissa haasteissa käyttöliittymään sisältyvä fasettirajaus tarkentaisi suositustuloksia paremmin käyttäjäryhmien tarpeita vastaaviksi: materiaali voitaisiin esim. luokitella sen mukaan, onko käyttäjä kiinnostunut kotimaan vai ulkomaiden markkinoista. Sisällönkuvailijat huomioivat usein suositukset siten, että pyrkivät käyttämään samoja käsitteitä kuvailussa, jotta voivat luoda yhteyden dokumenttien välille: *“Pyrin rakentamaan kokonaisuuksia, että on useita toisiinsa linkittyviä tekstejä.”* Varjopuolena voi olla, että suositukset voivat ohjata sisällönkuvailijaa välttämään tiettyjä käsitteitä tai rajaamaan käsitteen käyttöä vain tiettyyn kontekstiin, jotta ei-toivottuja suosituksia ei syntyisi.

”Meidän karkea suosittelu synnyttää näitä ongelmia. Puutarha-artikkelin yhteydessä suositeltiin uutista oopiumin viljelystä Afganistanissa. Otsikkokin oli muotoiltu niin, että melkein kehotti oopiuminviljelyyn.”

“Olisi hyvä, että asiasanoittajien ei tarvitsisi tietää käsitteiden taustaa (esim. lähikäsitteitä). Itse vastaan nyt asiasanoittamisesta ja tiedän, että käytän mieluummin sanaa ‘perehdytys’ kuin ‘perehdyttäminen’, joka tuo vääriä assosiaatioita. Kohderyhmien mukaan ottamisella saataisiin tilanne, että ei tarvi enää kikkailla käsitteiden kanssa.”

Teknisten haasteiden lisäksi haastatteluissa tuli esiin myös organisatorisia haasteita liittyen päätöksentekoon, resursseihin ja toimintamalleihin. Ontologian ylläpidon ja käyttöönnoton sekä järjestelmä uudistusten osalta näitä on jo käyty läpi. Lisäksi haastatteluissa todettiin, että ontologiat eivät ole kovin tuttuja joka alalla ja niihin saatetaan suhtautua skeptisesti tai vieroksuen: *“sitten toisaalta nämä ovat kuitenkin niin vaikeita asioita, että monelle tulee sellainen olo, että ‘tämä on liian teknistä, ei koske mua’”*. Suuremmissa organisaatioissa haasteena ovat eri osastojen erisuuntaiset tavoitteet ja mielipiteet: esimerkiksi dokumenttien säilyvyys tai löydettävyys ja sitä kautta hyvälaatuinen metatieto eivät välttämättä ole prioriteetti organisaation joka osastolla.

”Luulen, että kirjastoihmiset ovat aika skeptisiä koko asiaa kohtaan. Olen esitellyt ontologia-asioita kirjastolaisille ja vastaanotto on ollut aika hämmentynyttä. Toisaalta esittelin joskus myös VM:n virkamiehille, mutta ne olivatkin hirveän innoissaan. En ole missään saanut niin hyvää vastaanottoa kuin siellä. Jotenkin kaipaisin sellaista raikkautta kirjastoalallekin.”

Haastatteluissa onkin esitetty, että ontologioihin ajoittain liittyviin ennakkoluuloihin ja tiedon puutteeseen voidaan vaikuttaa antamalla aitoja esimerkkejä järjestelmistä, joissa ontologiaa jo sovelletaan rikkaasti. Erityisen tehokkaita olisivat kotimaiset esimerkit, joissa käytetään Finto-palvelua ja kotimaisia ontologioita. Tämä poistaisi epävarmuutta ja edistäisi päätöksentekoa.

”Jostain uusiutuvasta järjestelmästä voisi saada ensimmäinen tapauksen, jota voi sitten käyttää hyväksi jatkomarkkinoinnissa ja esitellä muille. Tällä tavalla lähtisin viemään markkinointia eteenpäin. Se puree meillä. Aina kysytään: ’Missä se on käytössä? Kuka sitä käyttää?’”

5. Erikoisontologioista ja -sanastosta sekä niiden ylläpidosta

Haastateltavista kahdestatoista tahosta kahdeksan on kehittänyt omiin tarpeisiinsa erikoisontologian. Sanastomassaltaan nämä ovat reilusta tuhannesta käsitteestä muutamaan tuhanteen. Näistä kahdeksasta erikoisontologiasta viisi on sellaisia, jotka perustuvat tiettyyn erikoisalan asiansastoon, jonka hallinnoinnista vastaa sama taho kuin erikoisontologiastakin. Viidestä asiansastosta neljää ylläpidetään ja päivitetään edelleen. Sen sijaan kahdeksasta erikoisontologiasta vain yhdellä (THL:n hallinnoima TERO-ontologia) on säännöllinen ylläpito. Muilla ontologioilla ei joko ole ylläpitoa lainkaan, tai ylläpito on haastatteluhetkellä ollut pitkään hoitamatta. Ainoa ylläpidettävä ontologia kuuluu niihin, joiden taustalla ei ole erikoisalan asiansastoa. Tiedot vahvistavat jo edellä esitettyä: asiansastojen parista ei ole vielä siirrytty ontologioihin. Ylläpidon puutteen vuoksi ontologioista monet ovat jo vanhentuneita. Esimerkiksi MAO-ontologiaa ei ole päivitetty sen jälkeen, kun se yli kymmenen vuotta sitten luotiin.

Erityisesti mediasektorilla on yleistä käyttää ja ylläpitää itse tuotettua karkeahkoa temaattista kategorisointia, joka sopii esimerkiksi uutisartikkelien lokerointiin. Tämä saattaa olla käytössä vain organisaation sisällä, julkaisemattomana, yhteisesti sovittuna sanalistana. Tällaisen kategorisoinnin katsotaan palvelevan tavallista käyttäjää, asiakaskuntaa, joka on tottunut selailemaan sisältöä laajempien teemojen kautta: ontologian katsotaan olevan tarkkuustasoltaan liian spesifinen. Kuitenkin esim. YLEn Elävän arkiston teemakategoriat kattavat noin tuhat käsitettä, mitä on pidetty tarkoitukseen nähden liian laajana kokonaisuutena.

“Ponduskategorier perustuu käytännön kokemukseen siitä, mitä aihealueita ihmiset yleensä haluavat hakea: kulttuuri, musiikki, politiikka, EU, lapset, vanhukset, nuoriso, eduskunta, presidentti. Tämän tyyppisiä. Olemme ajatelleet, että ihmiset haluavat käyttää näitä syötteissä, vähän niin kuin sanomalehden tai kirjaston eri osastot. “

Asiasanastojen ylläpidosta hahmottuu kaksi mallia: laajemmassa mallissa asiasanaston sisällöstä (lisättävistä ja muutettavista asiasanoista) päättää useamman sidosryhmän edustajista koostuva asiantuntijaryhmä, joka kokoontuu säännöllisesti, ja varsinaiselle käytännön ylläpidolle on osoitettu selkeä vastuutaho. Näin on varsinkin, kun sanastoa käytetään useammassa eri organisaatiossa. Toinen malli on, että sanaston sisältömuutoksista päättää ja ne toteuttaa yksi organisaatio itsenäisesti: työstä vastaa yksi tai useampi organisaation asiantuntija. Tämä malli esiintyy, kun sanastoa käytetään vain yhdessä organisaatiossa. Molemmissa malleissa tavoitteena on vastata kuvailtavan korpuksen ja uuden aineiston vaatimuksiin.

TERO-ontologia, jolla on säännöllinen ylläpito, on rakentunut monesta pohjalla olevasta asiasanastosta, joista useimpia ei enää ylläpidetä. TEROn ylläpito on tapahtunut tiiviissä yhteistyössä FinnONTO-projektin ja TSK:n Katri Seppälän kanssa: TEROn tulleet uudet käsitteet ja niihin mahdollisesti liittyvät ongelmakohdat on käyty yhdessä läpi. THL on itse järjestänyt ylläpitoon vaadittavan sanastollisen ja teknisen resurssin, jonka avulla ontologia on päivitetty ONKI-palvelimelle.

JUPO- ja LIITO-ontologioiden taustalla ei ole ylläpidettävää asiasanastoa, eikä niitä myöskään ole säännöllisesti ylläpidetty. Molemmissa ontologia on rakentunut projektiluontoisesti määräaikaista toimineen asiantuntijaryhmän johdolla. Rakennustyö on tehty FinnONTO-projektin ja Katri Seppälän johdolla tai avustuksella. LIITO-ontologian kohdalla ylläpidon haasteet ovat olleet lähinnä teknisiä, liittyen esim. ontologiaeditoriin. Henkilöresurssi ylläpitoon löytyy. JUPOn kohdalla ylläpidon konsepti on kompleksinen: ylläpidolta puuttuu sekä tekijä että sanaston sisällöstä päättävä asiantuntijaresurssi. Toisaalta ontologian käytöstä kiinnostuneet tahot eivät ole JUPOn sisältöön täysin tyytyväisiä ja kaipaavat edelleen kehitystyötä sen suhteen.

”Palvelukarttaa käytettiin JUPOn pohjana ja Katri Seppälä rakensi hierarkian. Se jäi kesken. Monesta kohdasta voisi vielä keskustella. Yritettiin tehdä liian täydellistä. Kellään ei ole selvää käsitystä, minkälaisen pitää olla kuntalaisen luokitus. Jokainen ajattelee sitä oman toimintansa kautta.”

5.1. Erikoisontologioiden ja YSO:n suhteesta

Kaikilla haastattelun aikaisilla erikoisontologioilla on jo itsessään tiivis suhde YSOon: luomisvaiheessa ontologia on sillattu YSOon ja yhtäläisyydet on tarkistettu. Lisäksi kaikilla erikoisontologian taustalla olevilla asiasanastoilla on läheisempi tai etäisempi suhde YSAan, jota on käytetty joko suorana tai epäsuorana referenssinä sanastoja laadittaessa. Esimerkiksi Valtioneuvoston asiasanastoon (VNAS) uudet asiasanat valitaan hyvin pitkälti YSAn esimerkin mukaan ja nykyisin ministeriöistä nousee vain harvoin tarvetta sellaisille asiasanoille, joita YSAssa

ei olisi. Yhtäläistä sanastoa erikoisontologioiden ja YSON välillä onkin paljon, ja arviot päällekkäisyydestä kunkin ontologian kohdalla liikkuvat 80 % molemmin puolin.⁶

Aina sama termi, sama kirjoitusasu ei kuitenkaan viittaa samaan käsitteeseen. Tällaisissa tapauksissa erikoisontologian ja YSON välille ei merkitä vastaavuussuhdetta.

”Tämäkin on tulkintajuttu, mutta olemme päättäneet, että valinta tarkoittaa meillä kasvin- ja eläimenjalostukseen liittyvää valintaa ja YSAssa se on yleiskäsite. Mutta tässä on se pulma sisällönkuvailun kannalta, että näissä viitteissä on paljon käytetty asiasanoja väärässä merkityksessä.”

Esiintyy myös hienovaraisempia eroja. Edellä mainittiin jo, että YSOssa olevat yleisluonteiset assosiatiiviset suhteet eivät välttämättä ole toimivia erikoisalan kontekstissa. Lisäksi YSOssa oleva käsite saattaa erikoisontologian näkökulmasta olla merkitykseltään liian yleinen tai tiettyyn kontekstiin rajattu. Edelleen erikoisontologian hierarkia saattaa olla ristiriitainen YSON näkökulmasta ja tällöin erikoisontologian hierarkiaa on muutettava YSOa vastaavaksi, mikäli käsitteet halutaan yhdistää. Toinen vaihtoehto on säilyttää erikoisontologian oma rakenne ja jättää yhdistämättä ristiriitaiset käsitteet YSOon, mutta tällöin menetetään yksi ontologisoinnin päähyödyistä: YSON kautta avautuvat semanttiset linkit muihin tietoresursseihin ja aineistomassoihin.

”Rakenne ei enää kerro käyttäjälle sitä informaatiota, mitä erikoisasiasanasto kertoi. Nyt ne termit ovat pitkin poikin ontologiaa. Siinä menetetään aika paljon. Kun se meidän perinnelaji on nyt aivan vieraiden asioiden keskellä samassa tasossa ja hirvittävän kaukana sellaisesta, joka meidän mielestä on aivan viereinen käsite, niin tavallaan siinä menetetään hirveän paljon informaatiota.”

”Katri [Seppälä] yritti muistuttaa, että puhutaan toimipisteistä, ei palveluista. Mutta jos uimahalli ei löydy palveluista niin pitääkö sinne laittaa ’uimapalvelu’? Sehän on kamala sana. Tällaisia pohdintoja jouduttiin käymään paljon.”

Merkitys- ja rakenne-eroistakin huolimatta erikoisontologioissa ja YSOssa on samoja käsitteitä paljon. Useamman erikoisontologian kohdalla tämä on ohjannut viime aikoina ylläpitäjiä karsimaan sellaisia yleisempiä käsitteitä, jotka jo sisältyvät YSOon. Vastaavasti on keskitytty selkeämmin oman erikoisalan käsitteistöön. Päällekkäistä kehitystyötä on pyritty karsimaan. YSO ja YSA sisältävät nykyisellään kuitenkin paljon myös erikoisalojen käsitteistöksi tulkittavaa. YSAa käytetään eri tieteenalojen väitöskirjojen ja tutkimusten kuvailuun, mitä kautta sanastoon kertyy paljon erikoisalojen tuoretta käsitteistöä.

⁶ JUHOssa on 6 546 käsitettä, joista 4 951 on merkattu ekvivalentiksi YSO-käsitteen kanssa, eli 75,6 % JUHO-käsitteistä.

”Usein käy niin, että kun meillä lisätään jotain sanastoon niin eipä aikaakaan, kun se on jo YSAssa. Tämä on sisällön rajaamisen kannalta aika turhauttavaa.”

Päällekkäisyys koetaan toisinaan turhauttavana, kun selkeää rajaa yleis- ja erikoissanastojen välille on käytännön syistä mahdoton tehdä, jos niitä halutaan käyttää yhdessä. KOKO eli YSON ja erikoisontologiat kokoava ontologia helpottaa kuitenkin mahdollisuutta hyödyntää yhtäällä tehtyä työtä muuallakin, ja siitä on apua niin yksittäisten ontologioiden kehittämisessä kuin sisällönkuvailussakin. YSA-päivityksissä huomioidaan erikoissanastojen ratkaisut aina kun on mahdollista. KOKOn laatuun vaikuttaa eri alojen välinen yhteistyö.

5.2. Erikoisontologian ylläpidon haasteista

Edellä jo esiteltyjen ylläpitohaasteiden lisäksi selvänä teknisenä haasteena on ollut ylläpidossa yleisesti käytetyn ontologiaeditori Protégén epävakaus, vaikeakäyttöisyys ja suoranaiset puutteet. Teknisten ongelmien selvittäminen on edellyttänyt yhteydenpitoa FinnONTO-projektin kanssa. Sisällöllisiä haasteita sanastoissa on mm. se, että jokin ala saattaa olla paremmin edustettu kuin toinen, mahdollisesti johtuen siitä, että tietyn alan toimijat eivät ole olleet aktiivisia ehdottamaan sisältöä. Tämän seurauksena sanastossa saattaa olla sisäisiä painotuseroja. Sanastossa saattaa myös olla liiankin spesifisiä termejä, joita tavallinen tiedonhakija ei tunnista: tässä on usein taustana se, että kuvailija on halunnut kuvaillun dokumentin erottuvan muista, mutta varjopuolena on, että dokumentin löydettävyyttä yleensä heikkenee. Edelleen jotkin termit saattavat jäädä ns. päiväperhosiksi eli ovat käytössä vain tietyn ajan ja unohtuvat sitten. Näilläkin on kuitenkin oma historiallinen arvonsa ja toisaalta termin yleistymistä on vaikea ennakoida.

”Kun jotain tapahtuu tieteessä niin tekee mieli laittaa uusi asia omaan sanastoon. Mutta ne saattavat jäädä päiväperhosiksi. Menevät nopeasti ohi. Toisaalta se ei haittaa, jos joku haluaa käyttää niitä myöhemmin. Mutta missä vaiheessa sen uuden termin sinne liittää?”

Päällekkäiseen työhön liittyen on kaivattu selkeämpää vastuutusta aihealueisiin liittyen: esim. Agriforestissa ja YSAssa on tällä hetkellä umpimähkäinen määrä eliölajeja. Viikin tiedekirjaston mukaan järkevämpää olisi laatia näille oma ontologiansa, jolle olisi järjestetty oma ylläpito esim. Luonnontieteellisestä museosta.

Oman erikoisontologian ylläpitoon liittyvät haasteet ovat saaneet useamman haastateltavan miettimään oman erikoisontologian olemassaolon tärkeyttä. Useampi haastateltava mainitsi, että erikoisontologian olemassaolo ”ei ole itsetarkoitus”. Monella organisaatiolla on resurssivajetta, vaikka tarvetta omalle sanastolle olisikin. Toisaalta jotkut eivät koe järkeväksi ylläpitää omassa ontologiassa käsitteistöä, joka on jo kertaalleen valittu YSAan tai YSOon. Kysymyksiin oman erikoisalan sanaston tärkeydestä on vaikuttanut myös se, että sisällönkuvailun asiantuntijaverkostossa on suositeltu käyttämään YSAA tilanteissa, joissa asia löytyy sekä YSAssa että erikoisalan sanastosta.

”Me olemme pitkään jutelleet VNAS-ryhmässä tästä. VNAS:ssa olevista sanoista 90 %⁷ on samoja kuin YSAssa niin miksi emme olisi suoraan YSAn käyttäjiä. Sama analogia on tässä ontologiassakin. - - Toisaalta sanasto pyörii edelleen Sharepointin kupeessa ja jos tässä nyt saadaan JUHO päivittymään niin totta kai me käytetään sitä. Mutta ollaan avoimia toisenlaisillekin ehdotuksille.”

Toisaalta monen erikoisontologian kehitystyö on keskeytynyt, kuten luvussa 4.2. esitettiin, ja päivityksen tarve sitä myötä kasvanut. Haastatteluissa esitettiin päivityksen nopeuden suhteen myös toiveita, jotka eivät ole resursseihin nähden realistisia.

“Sen ylläpidon pitää olla vakaa 24h vuorokaudessa. Päivityksen pitää olla todella nopeata, päivän kahden varoitusajalla täytyy voida päivittää. Pitäisi myös järjestää tapaamisia ja huoltaa sitä koko ajan. Hallinnon pitää olla voimakas, koska siihen voi tulla hajontaa.”

⁷ Tarkalleen ottaen JUHOn käsitteistä 75,6% on sellaisia, joilla on ekvivalenssi YSO-käsitteeseen.

6. Ontologioiden hyödyistä, haasteista ja kehitystoiveista

6.1. Näkemyksiä ontologioiden hyödyistä

Haastateltavat pitivät ontologioita monella tavalla hyödyllisinä. Eräitä hyötyjä, kuten autocompletoimintoa käsiteltiin jo luvussa 4.1. Ontologia ja sen hierarkia antavat älykkäämmät tiedonhakumahdollisuudet käyttäjille. Haunlaajennukset säästävät hakijaa: kaikkia relevantteja alakäsitteitä ei tarvitse kirjoittaa erikseen. Haunlaajennus on kuitenkin toteutettava selkeästi siten, että asiakas ymmärtää, mitä tapahtuu. Myös ontologiaan sisältyvä disambiguointi eli yksiselitteistäminen selkeyttää sanaston käyttöä niin tiedonhaussa kuin sisällönkuvailussakin: monimerkityksiset termit eriytetään ja hierarkia kertoo kunkin käsitteen semanttisen kontekstin.

”Toinen kulma olisi se, että jos ontologia olisi integroitu hakulomakkeeseen niin sitten ryhmittelyt voisivat olla käyttökelpoisia. Jos hakisi kaikkia jonkin alla olevia ja sitten sitä voisi fasetoida pienemmiksi palasiksi. - - Jos haluat hakea kaikilla tietyillä konkreettisilla käsitteillä niin niitä ei tarvitsisi erikseen kaikkia kirjoittaa.”

Ontologia hyödyttää myös selailussa (vrt. 4.1. ja 4.3.): *“Kun potentiaalisia hyötyjä ajattelee niin nimenomaan hakupuolella tällaisessa selailevassa haussa hyödyttäisiin ontologiasta enemmän kuin asiasanoista.”* Parhaimmillaan ontologia auttaa tiedonhakijaa löytämään tarvitsemansa tiedon nopeammin. On katsottu hyväksi, että hakijan ei tarvitse tuntea hierarkiaa, mutta hierarkia voi toimia taustalla ryhmittelemällä sisältöjä. Se voi myös toimia runkona verkkosivujen ryhmittelyn hierarkisoinnissa, tai sisältö voi ryhmittyä puoliautomaattisesti ontologian avulla.

”Lopullinen tavoite ontologialla on asiakkaan pallottelun lopettaminen. Asiakas voi käydä kymmenen eri sivua ennen kuin löytää. Ja pitää olla kunnollinen, selkeä pohja sivuille ja selailuun. - - Asioiden pitäisi löytyä omilla nimillä. Hierarkiaa ei tarvi tuntea. Mutta hierarkia auttaa ryhmittelemään asioita. Palvelukartta on opettanut, että ihmiset hakevat hierakiapuun kautta, eivät hakukentän kautta. Nykyään hakukentän käyttö on alkanut lisääntyä.”

Hierarkia hyödyttää myös sisällönkuvailussa: se vähentää päällekkäistä sisällönkuvailua, eli samaa dokumenttia ei tarvitse kuvailla sekä ’jääkiekko’- että ’urheilu’-käsitteillä. Hierarkia luo käsitteille merkitystä ja selventää niiden käyttöä sisällönkuvailussa. Selitteet ja käännökset selkeyttävät asiaa entisestään. Tämä helpottaa sisällönkuvailua varsinkin sellaiselle, jolle aihealue ei ole tuttu. Hierarkia helpottaa myös tietyn aihealueen käsitteiden (esim. tunteet) selailua yhdistämällä käsitteitä kategorisesti.

”On totta, että jos on tuttu alue, niin sisällönkuvailu ja termien löytäminen on helpompaa. Ja sama sisällönkuvailija kuvailee eri lailla eri päivinä. Mutta veikkaisin, että kaikissa olosuhteissa ontologia auttaa ja vähentää variaatiota. Työstä tulee tasalaatuisempaa. Ja usein puhutaan, että käsitteille ei ole selitteitä, mutta ontologia ja

sen hierarkia auttaa siinäkin. Sen avulla voi haarukoida termin merkityksiä aika paljon. Osa valita oikean.”

Kuten sisällönkuvailuprosessin yhteydessä (ks. luku 3) on todettu, ontologioiden monikielisyys on niin ikään koettu eduksi. Samaa aineistoa ei esimerkiksi tarvitse kuvailla eri kielillä uudestaan. Toisaalta tietynkielinen aineisto saattaa kiinnostaa muunkielisiä, jotka eivät löydä aineistoa kielimuurin ylitse: julkaisun ja hakijan kielen ei tarvitse enää olla sidoksissa toisiinsa. Tietovarannot avautuvat kielten kautta yli valtio- ja kielialuerajojen.

Tietovarantojen yhdistäminen taas liittyy tiiviisti linked open data -ajatteluun. Monet näkevätkin juuri tässä ontologioiden keskeisimmän hyödyn. Ontologian katsotaan auttavan tiedonsiirtotarpeeseen. Esimerkiksi valtiolla ja kunnilla on erillisiä palveluita ja tietovarantoja, joiden saattaminen yhteismitalliseksi olisi monella tavalla hyödyllistä. Ontologia auttaa harmonisoimaan metatietoa, minkä avulla tietovarannot ovat paremmin yhdistettävissä keskenään. Pitkällä tähtäimellä yhteiset sisällönkuvailuvälineet säästävät myös kuvailu- ja sanastoresursseja. Päälekkäinen työ vähenee.

”Ulkomaailmaan linkittäminen tulee olemaan tulevaisuudessa aivan kriittistä. Pitää miettiä tämänhetkisiä YSO-käyttäjätahoja, että löytävät toisensa ja pystyvät rikastamaan toistensa aineistoja.”

”Vieläkin ajatellaan, että asiakkaiden pitää tulla organisaatioiden sivuille, koska haluavat brändätä omaa aineistoaan. Muutaman vuoden kuluttua varmaan huomataan, että kaikki kuitenkin tulevat aineistoon FINNAn tai Googlen kautta. Ehkä ne jossain vaiheessa heräävät. Tämä on suuri ajatuksenmuutos.”

Jotkut haastateltavista pohtivat myös ontologioiden potentiaalia kokotekstihaussa: tulevaisuudessa ontologioita voisi hyödyntää myös siinä. Ontologioita voidaan hyödyntää myös automaattisen sisällönkuvailun työkaluna. Toisaalta pohdittiin myös vanhojen luokitusten yhdistämistä ontologiaan: tällöin saataisiin myös vanhat aineistot, joita ei ole asiasanoitettu, linkittymään uudempiin aineistoihin. Hyödyntämismahdollisuudet ovat kuitenkin sidoksissa ontologian monipuolisuuteen ja laatuun: *”mitä parempi ontologia taustalla, sitä paremmin pystytään vastaamaan tarpeisiin.”*

Konkreettisina ontologioiden sovelluskohteina monet mainitsivat omien organisaatioidensa verkkosivut ja tietojärjestelmät. Myös UKJ:ta ja FINNAA pidettiin erittäin hyödyllisinä ja kiinnostavina sovelluskohteina.

”Jos UKJ:ssä olisi YSO integroituna niin olisihan se mahtavaa sekä sisällönkuvailun että tiedonhaun kannalta. Ja sitten jossakin vaiheessa FINNAssakin apuvälineenä. Varsinkin jos siellä on vielä englannin- ja ruotsinkieliset termit mukana niin olisihan se mahtava. Helpottaisi kaikilla tavoin.”

6.2. Ontologioiden haasteista

Yksi kriittisimmistä ontologioihin kohdistuvista käytännön haasteista ja samalla kehitystoiveista, joka haastatteluista kävi ilmi, on palvelun vakaus ja nopeus. Monet ovat huomauttaneet, että ONKI3-palvelu on toiminut hitaasti tai ollut ajoittain kokonaan poissa käytöstä. Tämä on vaikeuttanut ja raskauttanut haastateltavien omien järjestelmien käyttöä. Se on myös ollut yksi painava syy siihen, miksi tietyt tahot eivät ole yhdistäneet järjestelmäänsä ONKI-palveluun: palvelua ei ole pidetty tarpeeksi vakaana.

”Sisällönkuvailussa on ollut teknisiä ongelmia. ONKI-palveluun ei ole saatu aina yhteyttä ja ONKI on saanut huonon maineen sen takia.”

”Emme näytä sanastoa verkkosivumme haussa teknis-historiallisista syistä. ONKI-palvelu oli tutkimusprojekti ja se oli toiminnassa tai ei. Kukaan ei uskaltanut sanoa, että mikä ONKIn käytettävyys oli. - - Asiasanoitusta varten pitää olla tehoja, että ei tarvi odottaa kauaa vastausta järjestelmästä. Vakaus on ehto asiasanoitukseen. Sitä tekevät henkilöt, joiden omiin intresseihin asiasanoitus ei juuri kuulu.”

Monia on myös askarruttanut ontologiapalvelun tulevaisuus. FinnONTO-projektin loppuvaiheilla palvelun jatko ja rahoitustilanne olivat pitkään epäselviä. Tämä veti monet tahot odotuskannalle, ellei jopa skeptisiksi, ontologioiden käyttöönoton ja kehittämisen suhteen. ONKI-projektin käynnistyminen 2013 Kansalliskirjastossa onkin otettu organisaatioissa innostuneesti vastaan.

“Ja tämä ONKIn rahoituksen saaminen oli erittäin tärkeätä. Jossain vaiheessa alkoi tuntua, että onko tällä ONKilla mitään tulevaisuutta, mutta nythän siitä on tulossa ihan oikea standardoitu palvelu. On erittäin tärkeä että se on osa valtion infrastruktuuria.”

”On hienoa, että Kansalliskirjasto on ottanut ONKIn ja siihen liittyvät asiat hallintaansa. Toivottavasti Kansalliskirjastolla, OKM:llä ja VM:llä on ymmärrystä jatkuvuudelle ja siihen liittyville tarpeille. Tämä on tärkeä infrastruktuurihanke, joka hyödyttää koko valtakuntaa. Jatkuvuus on todella tärkeää.”

Aiemmin on mainittu, että asiasanastoista ei ole siirrytty ontologioihin. Ilmiö heijastuu myös YSON ja YSAn suhteeseen. Monille nämä näyttäytyvät vaihtoehtoina, joista pitäisi valita jompikumpi. YSA tuttuna ja pitkään käytettynä sanastona saattaa tuntua jopa houkuttelevammalta vaihtoehdolta: *“Mulla on tiettyjä varauksia YSON suhteen ja YSA on mielestäni hyvä monissa asioissa niin mietin, että YSAnhan voisi ihan hyvin ottaa käyttöön.”*

Keräsimme haastatteluissa mielipiteitä YSON hierarkiasta. Varsinaisia mielipiteitä hierarkiarakenteen loogisesta tai semanttisesta pätevyydestä ei niinkään tullut. Yleisin kommentti sen sijaan oli, että ylähierarkiaa ei käytetä tai tunneta. Sitä ei ole tarkoitettu sisällönkuvailuun vaan luomaan ontologian rakennetta. Se onkin jäänyt monelle sisällönkuvailijalle vieraaksi ja abstraktiksi. Sisällönkuvailija liikkuu konkretian tasolla. Ylähierarkian käsitteiden nimitykset saattavat myös oudoksuttaa.

”Ylähierarkiasta (muuttuva, pysyvä jne) ei sisällönkuvailussa ole hyötyä. Menevät niin kauas itse sisällöstä. YLEllä materiaali on niin käytännönläheistä, että teoreettiset tietomallit, joilla jäsennellään kokonaisuutta, ovat monesti vain hieman hämmentäviä.”

”Termit on nimetty hassusti. Esimerkiksi ’kulttuuriset tuotokset’ – kukaan ei ole kiinnostunut kulttuurisista tuotoksista, vaikka olisikin kiinnostunut sen tyyppisistä asioista itsessään.”

Erikoisontologiasektorilla iso osa haastateltavista oli sitä mieltä, että YSON rakenne ei ole kovin funktionaalinen: se ei palvele hyvin tiedonhakijaa. YSON hierarkiat eivät ole temaattisia. Erikoisontologioiden ja YSON määppämisvaiheessa on ajoittain havaittu erikoisalan ja YSON edustaman yleisnäkökulman välistä ristiriitaa: kun kompaktista erikoissanastosta siirrytään laajaan yleisontologiaa, käsitteet hajautuvat ympäri hierarkiaa. Toisaalta hierarkia tuo vieruskäsitteiksi käsitteitä, joilla ei ole mitään temaattista yhteyttä keskenään: esim. ’verotus’ ja ’aseidenriisunta’. Toisaalta YSAN terminologinen historia näkyy ontologian taustalla: sanastoon on jäänyt vanhahtavia asiasanoja ja termisuhteita, joita ei ole käyty läpi systemaattisesti.

”Hautajaiset’ on YSAssa ja YSOssa perhejuhla. Meillä se on siirtymäriitti. Kulttuurientutkimuksen näkökulmasta on etnosentristä ajatella hautajaisia perhejuhlana. Esim. Afrikassa se on koko yhteisön juhla.”

”Syvemmällä tasolla vieruskäsitteet ovat relevantteja, mutta kun ollaan niin korkealla kuin esim. verotuksessa niin ei aseidenriisunta liity verotukseen mitenkään. Hierarkiaan ei voitu määritellä, että huomio hierarkia tietyillä tasoilla, koska hierarkia ei ole syvyydeltään konsistentti. Jos tähän löytyisi joku ratkaisu niin vieruskäsitteitäkin voisi hyödyntää.”

Yläkategorioiden visualisointi on myös koettu haasteelliseksi: kategoriset käsitteet eivät erotu sisällönkuvailukäsitteistä. Samoin erikoisontologiaa selatessa erikoisontologian käsitteet ja YSON käsitteet (esim. yläkategoriat, joita ei ole erikoisontologiassa) näkyvät selailussa yhtä aikaa ilman mitään visuaalista eroa. Tämä saattaa hämmentää selaajaa.

”YSON rakenne ei ole ihmiskatsottava. Siellä on aputasoja, joita ei ole tarkoitettu sisällönkuvailuun ja ne eivät millään tavalla visuaalisesti eroa muista käsitteistä. Sitä on vaikea selittää kirjaston asiakkaille.”

Esitettiin myös toiveita erikoisalojen asiantuntemuksen paremmasta hyödyntämisestä YSON kehitystyössä. Painotettiin kuitenkin, että myös kokonaisuus täytyy huomioida. YSON aihealueissa todettiin puutteita esimerkiksi urheilu- ja mediasanaston osalta. Media-alalla koettiin myös haasteeksi vastata ontologian avulla tuoreiden ilmiöiden ja päivän uutistapahtumien kuvailutarpeeseen. Tarkkuustasossa ja painotuksessa todettiin olevan eroa aloittain: jokin ala on tarkemmin edustettu kuin toinen. Tarkempaa rajausta toivottiin sen sijaan paikannimien, organisaatioiden ja tapahtumien suhteen: ehdotettiin niiden poistamista YSOsta ja siirtämistä

omiksi ontologiakokonaisuuksikseen. Kaivattiin myös mahdollisuutta päästä vaikuttamaan YSON assosiatiivisten suhteiden valintaan.

”Tarvittaisiin samanlainen organisaatio kuin meillä on koodistopalvelussa, että eri aihealueilla on omat asiantuntijansa, jotka käy sitä läpi, mutta sitten toisaalta se, keitä nämä asiantuntijat ovat, vaikuttaa siihen, mitä se sisältö on. Kaikkea valtaa ei voi antaa asiantuntijoille. Jonkun pitää pitää se kokonaisuus mielessä. Mutta kaikki alathan eivät voi olla yhtä laajoja, kun niistä ei ole dokumentaatiota, vaikka käsiteympäristöt voisivat olla yhtä laajoja.”

Kartoitimme myös mielipiteitä disambigoinnista eli monimerkityksisistä termeistä, jotka ontologiassa on jaettu useammaksi käsitteeksi. Eräät pitivät tärkeänä, että käsitteet on jaoteltu ja merkityserot huomioitu ontologiassa: *“On tärkeää, että asiat ovat eksakteja, eli haussa tulee esiin, jos sama merkkijono esiintyy eri merkityksissä”*. Useampi haastateltava kuitenkin koki erottelun liian hienojakoiseksi tiettyjen käsitteiden osalta. Tämä asettaa kuvailijan ja tiedonhakijan hankalan valinnan eteen ja saattaa ohjata käyttämään kaikkia jaettuja käsitteitä varmuuden vuoksi:

”Esimerkiksi ’lapset’ on hirveän vaikea. Pitää miettiä onko nyt kyseessä ikäryhmä vai perheenjäsen. Kun kuvailee lastenperinnettä niin se on oikeasti mahdotonta. Pitää ryhtyä syvällisiin filosofisiin pohdintoihin. Jossakin tapauksessa käsitejako voi auttaakin, mutta aika usein ei.”

Käsitteen jakamiselle likeinen (tosin ei vaihtoehtoinen) malli ontologiassa on monihierarkia: yksi käsite voidaan sijoittaa useampaan hierarkiaan, mikäli ratkaisu on ontologisesti ristiriidaton. Käsitettä ei tällöin jaeta, vaan sama käsite on löydettävissä useammasta paikasta ontologiassa. Monihierarkian tuoma käsitteen merkityksen laveus nähtiin helpottavan sisällönkuvailijaa: se poistaa valinnan vaikeuden usean näkökulmaltaan suppean käsitteen välillä (vrt. disambigointi). Monihierarkian visualisointi saattaa kuitenkin hämmentää käyttäjää, ellei sitä ole toteutettu selkeästi ja ymmärrettävästi käyttöliittymässä.

”Toisaalta, jos olisi helposti nähtävissä tietyn näkökulman mukaan asiat, joko yhden ja saman käsitteen kohdalla tai sitten tässä moniperintäkysymyksessä niin se voisi avata käyttäjille sitä syvyyttä ja voisivat pystyä paremmin kuvaamaan sisältöä riittävän tarkalla tasolla ottamalla huomioon nämäkin asiat. Se voi toisaalta tyypistää asiasanoittamista, jos hierarkiarakenne on liian yksioikoinen. Se on enemmän käyttöliittymähaaste.”

6.3. Kehitystoiveista

Haastatteluissa usein toistunut kehitystoive on assosiatiivisuuden ja horisontaalisuuden eli laajempien hierarkiapolkujen lisääminen ontologiaan. Haastateltavien mukaan näitä suhteita on YSOssa ja muissakin ontologioissa tällä hetkellä verrattain vähän: esimerkiksi ’vanhemmuus’ -käsite ei tällä hetkellä ole yhdistetty ’isät’ - tai ’äidit’ -käsitteisiin, mutta assosiatiivisella suhteella

nämä käsitteet voitaisiin yhdistää. Jotkut näkevät, että näissä jopa piilee *“suurin potentiaalinen hyöty ontologioitten suhteen.”*

”Hierarkiaa enemmän kaipaisi sivusuunnassa liikkumista, samalla konkretian tasolla olevia toimintoja ja toimijoita. Muuten joutuu kiipeämään ylös ja tulemaan alas, ja on taas kateissa.”

Assosiaatioiden nähdään hyödyttävän niin sisällönkuvailijaa kuin tiedonhakijaakin. Edellä puhuttiinkin jo assosiaatioiden hyödyntämisestä suosittelussa. YLEn Elävässä arkistossa todettiin, että assosiaatioiden avulla saisi luotua *”tosi herkkullisia yhteyksiä, samanlaisia, joita täällä yritetään tehdä toimituksellisilla koosteilla.”* Pohdittiin myös mahdollisuutta käyttää useita erilaisia assosiaatiosuhdetyppejä. Monet kuitenkin korostivat hyödyntämisen teknisiä haasteita: mikäli sinänsä hyvin muodostettuja assosiaatiosuhteita, tai jopa assosiaatioketjuja, on vaikea integroida järjestelmään teknisesti, tai toteuttaa niiden hyödyntäminen käytännössä, suhteet saattavat jäädä käyttämättä.

”Tekninen toteutus on yksi kysymys: miten hyvin pystyy hyödyntämään näitä järjestelmässä. Sehän riippuu myös siitä, miten suhteet rakennetaan ontologiaan: tuleeko sille assosiaatioketjulle esim. oma tunniste vai miten se toimii, pystyykö sen ylätunnisteen kautta hakemaan kaikki siihen kuuluvat asiasanat. Sehän voi olla hyvin loogisesti tehty ontologiaan, mutta jos se on vaikeasti teknisesti hyödynnettävä niin saattaa silti jäädä käyttämättä.”

Selkeäksi kehitystoiveeksi haastatteluissa nousi synonymia, erilaisten puhetapojen ja selkokielen huomiointi. Erityisesti korostettiin tavallisten käyttäjien, maallikkojen tapaa sanallistaa käsitteitä ja hakea sisältöä. Sanastot ovat tyypillisesti ammattilaisten muodostamia ja niissä korostuu ammattilaisterminologia. Rikkaampi synonymia kehittäisi kuitenkin ontologiaa palvelemaan paremmin myös maallikkokäyttäjiä.

”Suuhygienia on hyvä esimerkki. Miksei se yksinkertaisesti voi olla hammashoito? Hyvin harva kuntalainen puhuu suuhygieniasta. He haluavat mennä hammaslääkärille tai hammashoitoon. Kaikki meidän asiakkaat hakevat hammashoito-sanalla.”

Nykyistä laajempi määritysten ja selitysten käyttö on niin ikään monen haastateltavan toiveissa. Jopa siinä määrin, että joidenkin mukaan *“se on yksi YSO:n suurimmista ongelmista, että siellä ei ole määritelmiä.”* Määritelmät selkeyttäisivät käsitteiden käyttöalaa sisällönkuvailussa. YSO:n lisäksi määritelmiä kaivataan myös niihin erikoisontologioihin, joissa niitä ei tällä hetkellä ole.

Haastatteluissa kartoitettiin myös näkemyksiä prosessista, jolla ontologioita, erityisesti YSOa, rikastetaan uusilla käsitteillä. Suurin osa haastateltavista piti tärkeänä, että olisi olemassa kanava muutos- ja lisäysehdoituksille ja että toiveisiin reagoitaisiin nopeasti. Lisäksi erityisesti mediasektorilla kaivattiin mahdollisuutta käyttää uusia käsitteitä nopeasti ja huomioida myös loppukäyttäjien tarpeet. Käsitteiden ehdotus voisi tapahtua esim. kuvailujärjestelmän kautta. Toivottiin myös, että ehdotusprosessi olisi mahdollisimman interaktiivinen ja läpinäkyvä:

ehdotuksia voisi seurata ja kommentoida. Palautetta pitäisi voida antaa myös nykyisistä käsitteistä ja tehdä niihin tarvittaessa korjausehdotuksia.

“Uutisaineistossa uusia aiheita tulee nopeasti ja usein, ja meidän sanasto ei välttämättä ole ihan yksi yhteen kirjastomaailman käsitysten ja nimien kanssa. Asiakkaat eivät välttämättä sellaisia nimiä käytä. Jos tietty nimitys esiintyy lehdissä päivästä toiseen (esim. teknologi-industri), mutta ei löydy Allsosta niin se ei palvele meidän asiakkaita. Me tarvittaisiin joustavuutta ja mahdollisuutta jakaa YLEn kanssa nopeasti uusia termejä, joilla olisi URI, ja sen voisi sitten myöhemmin linkittää oikealle paikalleen ontologiaan. Vaatisi paljon työtä, mutta mikä estäisi meitä antamasta käyttäjille mahdollisuuden laittaa synonyymeja sanastoon? Kun käyttäjä laittaa sinne oman sanan niin se voi saada kannatusta ja asiakkaiden oma tapa nimetä asioita voisi nousta hakutuloksissa. Olisi hyvä ottaa talteen käyttäjien spontaanit reaktiot mahdollisimman nopeasti. Asiakkailta voi oppia, järjestelmä voi oppia. Termin voi vaihtaa, jos asiakkaat suosivat jotain tiettyä. Mutta ajallista aspektia ei saa hukata, kun nimet vaihtuvat. Ja Wikipediaan voisi linkittää, tai antaa asiakkaiden ehdottaa linkkejä. Merkityksen antamista pitäisi miettiä prosessina. Liukuva, dynaaminen prosessi.”

Yksi tapa kerätä epäsuoraa palautetta loppukäyttäjiltä on käyttäjien tekemien hakujen seuraaminen: tuloksettomat nollahaut kertovat tavallisen käyttäjän hakutarpeiden sanallistamisesta. Niistä voi olla apua ontologian termistöä ja synonymiaa rikastettaessa.

Päivitystarpeisiin vastaamisen ohessa monet kuitenkin muistuttivat myös, että yhteys vanhoihin aineistoihin ja niiden kuvailutietoihin on säilytettävä. Muuttuvien käsitteiden osalta pitäisi ilmaista muutosten ajankohta, esim. milloin on siirrytty tietyn termin käytöstä toiseen. Historiallista tutkimusta varten tarvitaan myös vanhentuneita termejä. Toisaalta voisi olla hyvä saada hakutoiminnosta tarvittaessa suljettua pois vanhentuneet termit, jottei haku olisi liian raskas.

Dokumentaatiota pidetään tärkeänä, jotta asioiden etenemistä pystytään seuraamaan. Laajemman dokumentoinnin lisäksi kaivataan wiki-tyyppistä tiedotusta: ”*raporttien julkaisussa voi mennä aikaa, mutta wikiä voi koko ajan korjata ja ylläpitää, viimeisimmätkin jutut.*” Selkeää ja perustavanlaatuista ohjeistusta kaivataan: tämän tärkeys korostuu varsinkin, kun tekninen toteutus tilataan ulkopuoliselta taholta, joka ei entuudestaan tunne ontologioita. Tiedotus olisi hyvä olla suomen lisäksi myös ruotsiksi ja englanniksi. Esimerkiksi Ruotsissa ollaan kiinnostuneita Suomen ontologiakehityksestä, mutta tietoa projektista on saatavilla hyvin vähän ruotsiksi.

Haastatteluissa esitettiin myös toivomuksia ontologiapalvelun graafisesta käyttöliittymästä, jossa käsitteitä ja niiden verkostoja voisi selata moneen suuntaan nykyisen listamallisen hierarkiaselauksen sijaan. Joidenkin haastateltavien mukaan tällainen helpottaisi käyttöä ja ontologiarakenteen hahmottamista olisi helpompaa. Käsitteverkkoa tai -karttaa voisi esimerkiksi zoomata sisään ja ulos. Tässä mallissa myös assosiativiset suhteet voitaisiin huomioida keskeisemmin kuin nykyisessä käyttöliittymässä. Lisäksi toivottiin FinnONTO-projektissa luodun ARPA-järjestelmän jatkokehitystä, sillä nykyisen järjestelmän antamat sisällönkuvailutulokset ovat suurelta osin vielä käyttökelvottomia.

”Meillä on ARPA käytössä, mutta se ei tuo kovin hyviä vastauksia. Olisi helpompi jos ARPA olisi luotettavampi varsinkin, kun jatkossa me emme täällä itse asiasanoita kaikkia materiaaleja.”

6.4. Instanssiontologioista

Temaattisen sisällönkuvailun lisäksi monet tahot tarvitsevat kuvailussaan myös instansseihin liittyvää tietoa kuten paikka- ja toimijatietoa. Erityisesti paikkatieto on ollut yksi keskeisimmistä kiinnostuksen kohteista sidosryhmien kanssa käydyissä keskusteluissa. Paikkatiedolle on tarvetta niin media-alalla, kulttuurintutkimuksessa kuin luonnontieteissäkin. Sidosryhmät pitävät tärkeänä, että paikkatieto-ontologia olisi tarjolla keskitetysti samassa palvelussa kuin muutkin ontologiat.

FinnONTO-projektissa kehitetty SAPO-ontologia sisältää ajallista tietoa paikkakuntien nimien ja rajojen muutoksista. Myös tätä tietoa pidetään arvokkaana ja hyvin käyttökelpoisena monessa organisaatiossa. Ontologiaa pidetään optimaalisena mallina tällaisen tiedon esittämiseen.

“On hyvä jos ONKIin yhdistetty järjestelmä tunnistaisi historialliset paikannimet automaattisesti. Ei tarvitse asiasanoittaa sekä vanhalla että uudella kuntanimellä. Nämä tällaiset ovat tärkeä tapa lähestyä arkistomateriaalia: ihmisiä kiinnostaa itseä koskevat asiat, oma elinpiiri, vanhempien kotiseutu.”

Ontologiaan voi sisällyttää myös spatiaalisia osa-kokonaisuus-suhteita esim. kaupunkien ja kaupunginosien välille. Myös tätä tietoa pidetään tärkeänä. Näillä voisi jatkossa korvata maantieteellisten tarkkuustasojen päällekkäisen kuvailun tarpeen. Toisaalta joillakin organisaatioilla paikkatiedon tarve ulottuu myös Suomen rajojen ulkopuolelle, kun taas esim. SAPO-ontologia on keskittynyt vain Suomen paikkoihin.

”Ollaan jouduttu tehoistamaan luettelointia eikä enää kirjoiteta esim. jokaista osoitetta kokonaisuudessaan, kaikkia maantieteellisiä tasoja. Nyt ajatellaan, että ontologia ratkaisee sen asian jonain päivänä. Maantieteelliset tiedot tulevat sieltä. Se olisi tärkeää: jos haet 'Nyland' niin saat kaikki osoitteet, jotka ovat sen alla.”

”Olisi hyvä jos erisnimiontologiat löytyisivät ONKI-palvelusta. Olisi hyvä keskittää tekniset ratkaisut samaan palveluun. Kilpailevista palveluista Geonames on hyvä, koska uutisissahan on pakko olla globaali palvelu. Suomesta on olemassa laadukasta paikkatietoa, mutta jos Moskova puuttuu niin emme voi käyttää yksistään tällaista palvelua.”

Paikkatiedon ohella myös toimijatieto on laajalti kysyttyä ja sille kaivataan omaa tietovarantoa, jota voisi käyttää sisällönkuvailussa: ”*tämä rekisteri on todellakin tarpeellinen. Käyttöä olisi varmasti.*” Paikkatiedon tapaan myös toimijatieto muuttuu ajan myötä. Tämä olisi tärkeä huomioida myös toimijaontologiassa mm. organisaatioiden muuttuvien nimien, linkaaren ja vaihtoehtoisten nimien muodossa. Myös tarkemmat tiedot kuten organisaation toimiala ovat haluttuja.

Auktoriteettitietokantoja, erityisesti Kansalliskirjaston ylläpitämää ASTERI-tietokantaa, pidetään hyvänä lähteenä toimijaontologialle. Toimijaontologia olisi myös mahdollista linkittää ulkopuolisiin tietovarantoihin, esim. DBpediaan, jolloin ontologia verkostoituisi globaalisti ja käyttäjät voisivat saada lisätietoa toimijoista muualta.

”Ilman muuta näistä olisi hyötyä, jos olisivat standardoituja. Henkilönnimillä hakeminen on tyypillinen hakutapa. Samoin tapahtumat. Toisaalta tässä pitää selvittää, mikä on ontologian suhde auktoriteettitietokantoihin. Auktoriteettitietokannan ylläpidon ratkaisut olisi hyvä mennä käsi kädessä henkilöontologian kanssa, jos ne on pakko pitää erillisinä järjestelminään.”

Tarvetta on koettu myös tapahtumaontologialle, jotta erilaisten merkittävien tapahtumien huomiointi sisällönkuvailussa olisi helpompaa ja tapahtumilla voisi myös hakea tietoa nykyistä sujuvammin. Toisaalta ontologia voisi myös olla ratkaisu vanhoihin teoksennimiin liittyviin ongelmiin. Esimerkiksi eepoksen nimissä on historiallista vaihtuvuutta: aiemmin niitä yritettiin suomentaa, mutta nykyisin suositaan alkukielistä nimeä.

”Haluaisin luoda uutistapahtuman, jolla on URI, ja siihen voi liittää henkilöt, paikan, ja sitten siihen voisi linkittää uutisia YLEltä ja meiltä. Se olisi tosi hyvä myöhempiä yhteenvetoja tai tutkimusta varten. Siinä on haasteena keston määrittäminen.”

6.5. Ontologioiden kehitystyön mallista

Sanastoylläpidon resurssihaasteista puhuttiin jo edellä (ks. 4.2.): erityisontologioiden ylläpito on jäänyt vähäiseksi. Haastatteluissa kysyttiin mielipiteitä siitä, mikä olisi järkevä ja realistinen malli ontologioiden ylläpidolle, ja miten haasteltavat näkisivät Kansalliskirjaston ja oman organisaationsa roolin tässä mallissa.

KOKO-ontologiaan suhtaudutaan yleisen myönteisesti. KOKOssa erikoisontologiat on julkaistu yhdessä YSOon ripustettuna. KOKO-ontologia edustaa monelle tietovarantoa, jonka tulevaisuus näyttää turvatulta ja josta ollaan kiinnostuneita varsinkin, kun joidenkin organisaatioiden omat sanastoresurssit ovat epäselvät. Resurssipaineiden alla useampi taho miettii tällä hetkellä oman sanastonsa tulevaisuutta ja tarpeellisuutta. Toisaalta asiaan vaikuttaa myös se, että verrattain suurella asiantuntijaryhmällä ja tiheähköllä päivitystahdilla ylläpidettävä YSO kattaa tällä hetkellä hyvin paljon erikoisontologioidenkin sanastoa. Pällekkäistä ylläpitotyötä ei pidetä mielekkäänä: ”*Suomen kokoisessa maassa valtakunnallinen palvelu on varmasti ekonomisempaa kuin jos kaikki tekisivät itse sanastoja.*”

Näkemykset KOKOn ylläpidon toteutusmallista ja Kansalliskirjaston roolista siinä jäivät haastatteluissa vielä abstraktille tasolle. Kuitenkin yleinen ehdotus KOKOn ylläpitomalliksi oli, että nykyisten erikoisontologioiden ylläpitotahot vastaavat jatkossakin näiden alojen sanastosta. Suurin osa haastateltavista pitää hyvänä mallia, jossa erikoisalojen ontologiat ovat jatkossakin erillisiä kokonaisuuksia, ontologioita, joita on tarpeen tullen mahdollista käsitellä muusta KOKOsta

erillisenä. Tämä selkeyttää myös ylläpidon vastuuttamista. Ontologioiden erottelun ohella tärkeänä pidetään kuitenkin, että ne ovat kootusti tarjolla samassa palvelussa ja yhteisindeksissä.

”Voisi olla hyvä, jos olisi vain yksi ontologia. Kyllähän nyt jo toimintatapa on sellainen, että me käydään yhdessä läpi YSA-väen kanssa uusia termiehdotuksia ja mekin ehdotellaan sinne sanoja jonkun verran. Tai he kysyvät meiltä, mitä mieltä ollaan. Jos olisi vain yksi sanasto niin se tehtäisiin juuri näin. Me vastattais meidän kampuksen aloista. Niin me se nähdään ja varmasti YSAssa asia nähdään myös näin.”

Pieni osa haastateltavista piti kuitenkin parempana mallia, jossa ontologioita ei ole koottu yhteen, vaan ne ovat olemassa erillisinä kokonaisuuksinaan. KOKO-mallissa nähdään haasteena erikoisalojen näkökulmien ja terminologian keskinäinen yhteensopimattomuus. Hajautetussa mallissa jotkut näkivät etuna myös sen, että jatkuvuus ei ole riippuvainen yhden ison projektin rahoitusnäkymistä.

”Sellainen käsitys on, että sosiaali- ja terveysala on itsessään jo niin iso ja vaikeasti hallittava kokonaisuus, että jos se yhdistettäisiin KOKO-ontologiaan niin ylläpito ja sanaston seuraaminen kävisi aika hankalaksi. On paljon sanaston alueita, joissa termeillä on yleisiä merkityksiä, mutta sitten on spesifimpiäkin termimerkityksiä, ja nämä helposti sotkeentuvat jokapäiväisiin merkityksiin.”

7. Johtopäätökset

Raportti perustuu vuoden 2013 jälkipuoliskolla tehtyihin haastatteluihin, jolloin Kansalliskirjaston Finto-palvelua ei ollut vielä julkaistu. Finto on dynaamisessa kehitysvaiheessa, ja osaan jo esille tulleista ongelmista on löydetty ratkaisu. Osa on työlistalla ja lisäselvityksen tai suunnittelun asteella.

Tässä luvussa kerrotaan sidosryhmähaastattelujen herättämiä erilaisia ajatuksia ja näkökulmia. Tässä ei siis käydä läpi haastateltavien omia ajatuksia, vaan niiden pohjalta syntyneitä näkemyksiä ja kokoavia linjauksia, jotka on syytä ottaa huomioon tulevassa kehitystyössä. Haastatteluaineistosta on hahmotettu yhteneväisiä suuntauksia ja toiveita. Tässä esitellään, minkälaisilla toimenpiteillä projekti voisi näihin toiveisiin vastata.

7.1. Sidosryhmien käytännöistä ja toiveista ja niihin vastaamisesta

Sisällönkuvailukentällä kehitetään uudenlaisia järjestelmiä ja sisällönkuvailijat ovat taustaltaan aiempaa heterogeenisempia toimijoita: sisällönkuvailua tuottavat semanttisessa webissä usein kuvailtavan substanssin osaajat, kuten tutkijat, virkamiehet, toimittajat ja myös harjoittelijat ja kesätyöntekijät. Uudenlaiset toimijat usein siis tuntevat sisällönkuvailemansa aihepiirin ja ovat sen asiantuntijoita, mutta eivät välttämättä tunne sisällönkuvailuperiaatteita ja -käytäntöjä.

Keskeiseksi haasteeksi koettiin myös asiasanojen funktion ja asiasanoittamisen periaatteiden tuntemattomuus ns. tavallisille loppukäyttäjille ja ei-ammattimaisille sisällönkuvailijoille. Haastateltavat kokivatkin tärkeäksi järjestelmän suunnittelemisen niin, että myös se osaltaan avaisi asiasanoitusta itsenäiskäyttäjälle ja toimisi käyttäjäänsä opastaen. Sisällönkuvailijan olisi hyvä saada kattavasti informaatiota esim. koskien suositeltua ja ohjaavaa termiä, kun taas tiedonhakijaa tulisi säästää hänelle liialliselta ja epärelevantilta informaatiolta. Sisällönkuvailun välineisiin kohdistuukin nyt uudenlaisten käyttäjien myötä uudenlaisia käyttötarpeita ja -odotuksia. Käytettävyyden pitäisi palvella paitsi substanssiltaan ja toimialoiltaan sekä kehysorganisaatioiltaan varsin erilaisia toimijoita, niin myös sekä sisällönkuvailun ammattilaisia että noviiseja, joiden tarpeet ja valmiudet voivat erota osin suurestikin.

Haastatteluissa kävi ilmi, että yhtäältä asiasanastoista ei ole juurikaan siirrytty ontologioihin ja tehdyt ontologiat ovat jääneet käyttämättä. Toisaalta ontologioista ollaan kiinnostuneita ja niistä katsotaan olevan hyötyä. Haastateltavien yleinen asenne oli myönteinen ja optimistinen. Asenteen ei kuitenkaan koettu olevan alalla välttämättä se vallitsevin. Ontologia-ajattelu on vielä varsin uutta ja ns. suurelle yleisölle vierasta, nyt aktiivisen kehitysjoukon luodessa uudenlaista toimintakulttuuria ja -välineistöä.

Kehittämisen ja käyttöönoton yleinen este on resurssipula - ontologisointiin ja järjestelmien jne uudistamiseen ei ole irrottaa ihmistyövoimaa, osaamista tulisi kehittää ja toimintaa rahoittaa ja tukea. Ontologiatyön rahoitus ja rahoituksen jatkuvuus todettiinkin elimellisen tärkeiksi. Hyvin moni järjestelmä jää tyhjän päälle, ja jo tehty työ valuu vähintäänkin osin hukkaan, jos rahoitusta ei

ole. Käyttö- ja kehityskatkos onkin ollut ikävä todellisuus Suomessa ontologiahankkeiden alkuvaiheessa, ja joissakin organisaatioissa se on johtanut siihen, että jo laadittu erikoisontologia on jäänyt ottamatta käyttöön. Jatkokehittämistä mietittäessä huolta on yleisesti aiheuttanut palvelun jatkuvuuden epäselvyys - mitä tapahtuu jo tehdyille ja suunnitteilla olevalle työlle. Resurssipuute on kohdistunut yleisesti niin sanastolliseen kuin tekniseenkin toimintaan. Toinen merkittävä syy on se, että ontologisointityötä ei ole toteutettu kylliksi aiemmat sisällönkuvailukäytännöt huomioiden eikä sen siten katsota palvelevan käytännössä sisällönkuvailua eikä tiedonhakua. Kansalliskirjaston ONKI-projektissa onkin ryhdytty lisäämään kaivattua temaattisuutta mm. palauttamalla YSA-ryhmät, ja ylähierarkian muutoksilla halutaan vaikuttaa myös temaattisesti aiempaa tarkoituksenmukaisempien ratkaisujen löytymiseen. Huolta ovat aiheuttaneet myös palvelun riippuvuusuhde FinnONTOon, silloisen projektin aikaiset palvelinongelmat ja näkökulma, jossa sisällönkuvailua ei painotettu. Haastateltavat myös kokivat olleensa toimijoina ja kehittäjinä usein melko yksinäisiä, ja kaipasivat selkeää keskitettyä koordinoijaa, joksi he ehdottivat Kansalliskirjastoa.

Ontologian integrointi sisällönkuvailu- ja tiedonhakujärjestelmään kannattaa. Ontologia yhdenmukaistaa sisällönkuvailua organisaatioiden välillä, mahdollistaa aineistojen yhdistelyn sekä asiasanaston ontologisointiin perustuvan hakujen laajentamisen ja supistamisen sekä poistaa semanttisesti turhaa päällekkäistä sisällönkuvailutyötä. Se myös karsii tiedon löydettävyyden kannalta turhan rikasta ilmaisujen kirjoa ohjaten useasta eri ilmaisutavasta samaan käsitteeseen ja samankaltaiseen aineistoon. Toisaalta on muistettava, että erilaiset diskurssit perustuvat usein myös erityyppiseen aineistoon, ja että haussa on tärkeä voida halutessaan kohdistaa kysely tietyyppiseen aineistoon tai nimenomaan tietyllä tavalla ilmaistuun käsitteeseen. Jo tehty luokitus- ja sisällönkuvailutyö itsessään mahdollistaa nykyistä monipuolisempien hakujärjestelmien suunnittelun ja käytön.

Linked open data -toimintamalli pitää sisällään ajatuksen siitä, että aineistoja voidaan yhdistellä yli organisaatorajojen. Tällöin yhtäällä jo laadittu metadata voidaan hyödyntää muuallakin. Ontologia voi siis tuoda myös kustannustehokkuutta vähentämällä turhaa päällekkäistä sisällönkuvailutyötä eri organisaatioissa.

Ontologioihin siirtyminen vaatii ONKI-projektilta ja muilta alan pioneereilta asennetyötä - ontologiat ja niiden mahdollisuudet ovat laajemmalla kentällä vielä varsin tuntemattomia asioita, ja niistä tulisi tiedottaa ahkerasti. Lisäksi tarvitaan käyttökokemuksia: tehdyn työn kannattavuus tulisi konkretisoida, ja siihen tarvitaan enemmän Yritys-Suomen tyyppisiä sovellusesimerkkejä, joissa ontologioita jo käytetään sujuvasti ja rikkaasti.

Ontologioiden käyttöönotto järjestelmissä on tähän mennessä edennyt melko hitaasti: on siis syytä suunnitella sekä tekniseltä että sisällölliseltä kannalta, miten tulisi auttaa ontologioiden parissa ensiaskeleitaan ottavia. Paitsi ontologisointiasiantuntemusta, myös teknistä tukea pitää olla saatavilla. Ylläpidon pitäisi olla teknisiltä välineiltään helppoa ja kehittäjäystävällistä. Paitsi että tarjolla tulisi olla teknistä ja sanastollista tukea, saatavilla pitäisi myös olla hyvä ohjeistus ja selkeät toimintatavat, joiden avulla sanastoa voisi ylläpitää ilman jatkuvaa yhteydenpitoa Kansalliskirjastoon. Ontologioiden integroimiseen pitäisi myös tarjota teknistä ohjeistusta ja tukea.

ONKI-projektissa onkin kiinnitetty jo huomiota tiedottamiseen ja ohjeistukseen: esim. ontologiaeditoreista on julkaistu raportti⁸, jossa vertaillaan editorien toiminnallisuutta. Projekti tiedottaa toiminnastaan Wiki-sivuillaan⁹ ja ONKI tiedotus -postituslistassa. Linjanvetopäätökset sekä sisällölliset ja tekniset ohjeistukset tullaan dokumentoimaan ja viestimään sidosryhmille.

Sidosryhmien omien haku- ja sisällönkuvailujärjestelmien ylläpidon ja kehityksen kannalta on hyvä, jos heillä on organisaation sisällä omia pysyviä teknisiä resursseja: yksinomaan ulkopuolisen resurssin varassa toimiminen on hankalampaa ja katkonaisempaa. Finto tarjoaa myös tukea, mutta on hyvä muistaa, että myös Fintolla on resurssirajoitteet.

Uusien toimijoiden ja käyttötarkoitusten sekä -tapojen myötä erilaisten kompromissien tarve lisääntyy. On hyvä muistaa, että eri erikoisryhmillä on oma näkökulmansa ontologian käsitteistöön, ja viime kädessä käsitteenmuodostus on aina myös yksilöllistä. Ontologiapalvelun kehittäjät joutuvat punnitsemaan painotuksia. ONKI-projektin on siis valittava näkökulma sanastoon ja palveluun. Tällä hetkellä esimerkiksi YSO on laajasti ymmärrettyä yleinen suomalainen ontologia, joka on palvelun alkuvaiheessa noudattanut käsitteistöltään ja päivityskäytännöiltään YSAa eli Yleistä suomalaista asiasanastoa. YSOon kohdistuu kuitenkin moninaistuvampia tarpeita ja myös itse ontologiarakenne edellyttää uudenlaisia ratkaisuja, joten YSO:n suhde YSAan tulisi Fintossa määritellä tarkemmin ja YSOon kohdistuvat uudenlaiset käyttötarpeet huomioiden. On huomattava, että siinä missä YSA:n kehitystyö on kirjastotarvelähtöistä, on YSO:n kehitystyön lähtökohtana laajempi käyttökonteksti eli semanttinen web, jossa kirjastojen tarpeet ovat yksi keskeinen osa, mutta mukana on muidenkin toimijoiden, kuten hallinnon ja median, näkökulmat ja tarpeet. YSOsta löytyy jatkossa siis YSA:n sisältö, mutta lisäksi paljon muutakin, kuten englannin käännökset, siltaus kirjastoalan keskeiseen sanastoon Library of Congress Subject Headings, ja laaja hierarkkinen konteksti.

Kuten jo todettua, niin tarpeet ja niiden priorisoinnit ovat usein eriäviä, ja jopa saman organisaation sisäiset toiveet ovat joskus ristiriidassa keskenään. Sidosryhmien olisikin hyvä muistaa, että toiveet kentällä ovat varsin kirjavia, mikä kohdistaa projektille tiettyjä kompromissipaineita. Erilaiset näkökulmat ja mielipiteet ovat kuitenkin arvokkaita ja niitä on syytä aktiivisesti esittää. Media-alan tarpeet ovat yhtä tärkeitä kuin perinteisen kirjastokentänkin tarpeet: media edustaa monessa asiassa tulevaisuuden toimintamallia.

Monikielisyydestä keskustellessa on hyvä muistaa, että YSO on nimensä mukaisesti nimenomaan yleisistä suomalaisista sisällönkuvailutarpeista lähtöisin laadittu. Se on siten monikielinen, mutta silti lähinnä yksikulttuurinen, mikä selittää tietyt hierarkkiset ratkaisut. Toisaalta erikoisontologian kannalta suomalaiseen kulttuuriympäristöön rajautuva näkökulma voi näyttäytyä puutteena. Ontologiatyössä on kuitenkin oltava selkeästi määritelty lähtökohta, sillä käsitteelle on periaatteessa usein monta ns. oikeaa hierarkiapaikkaa, mutta ontologiatyössä valittu näkökulma johdattaa sen kehitystä ja käyttöä.

⁸ Palonen, Tuomas; Suominen, Osmo; Pessala, Sini; Lappalainen, Mikko; Nykyri, Susanna; Frosterus, Matias: "ONKI-projektin ontologiaeditorien vertailuraportti" <http://urn.fi/URN:NBN:fi-fe201306244091>

⁹ URL: <https://wiki.helsinki.fi/display/ONKI/>

Haastatteluissa ilmeni, että monet erikoisalajat miettivät omasta sanastosta luopumista, koska sille ei ole osoitettu kyllin resursseja ja/tai tulevaisuus resurssipulan johdosta huolestuttaa. Siksi olisikin erityisen tärkeää, että myös sidosryhmien vastuu olisi näkyvää ja tiedossa: Finto-palvelu tarvitsee monenlaista asiantuntemusta. Kansalliskirjastolla ei ole erikoisalojen osaamista eikä se näin ollen voi vastata erikoisalojen sanastoista niiden substanssin kannalta. Resurssit siihen pitäisi osoittaa erikoisaloilta. Kansalliskirjaston rooli koordinoijana, ontologiatyön asiantuntemuksen jakajana ja verkostojen luojana on kuitenkin merkittävä, ja siinä sen tulisi toimia aktiivisesti. Kansallinen palveluväylä ja ontologiapalvelu tarvitsevat keskitettyä toimintamallia. ONKI-projektin tulee olla aktiivinen tiedottaja ja konsulttija, ja KOKO-ontologian ylläpidon konsepti on mietittävä tarkkaan ja tehtävä läpinäkyväksi. Keskeinen haaste konseptissa on, miten luoda prosessista kevyt ja sujuva, mutta kuitenkin kaikki sidosryhmät huomioon ottava.

7.2. YSON ja Finto-palvelun kehitysmahdollisuuksia

YSON hierarkiaa koskevissa toiveissa korostui selkeys, ymmärrettävyys ja funktionaalisuus. Onkin tärkeä saada hierarkia palvelemaan nykyistä enemmän tiedonhakijaa. Tähän asti hierarkiaa rakennettaessa on ollut päämääränä terminologinen ja semanttinen pätevyys. Tätä ihannetta ei sovi unohtaa, mutta on tärkeää ottaa huomioon aiempaa enemmän myös ontologian käyttökonteksti, tiedonhaku ja verkkoselailu, ja miettiä, voiko hierarkia vastata paremmin näiden tarpeisiin. Ontologian ylähierarkia on koettu vieraaksi ja abstraktiksi. Toisaalta ylähierarkiassa esiintyvät käsitteiden nimet ovat välillä kummastuttaneet. Ylähierarkiaan onkin jo tehty muutoksia rakenteen selkeyttämiseksi ja tarkempi läpikäynti on alkanut.¹⁰

ONKI-projekti on huomionut selkeyden ja funktionaalisuuden myös ontologiaselaimen suunnittelussa. Projektissa on valmistunut käyttäjätutkimuksen ensimmäinen osa (Pessala 2013), jossa ilmi tulleet havainnot ja haasteet on huomioitu selaimen rakentamisessa. Käyttäjystävällisyys huomioidaan niin ikään myös palvelun visuaalisessa ilmeessä, jota kehitetään yhteistyössä Hahmo Design Oy:n kanssa.

YSON assosiativisten suhteiden kehittäminen kiinnostaa monia. Tähän on syytä paneutua jatkossa ja kartoittaa erilaisia etenemistapoja. YSON assosiativisten suhteiden rikastaminen käsityönä vaatisi suuren resurssin. Koneellinen lisääminen käyttämällä bibliometrisiä menetelmiä voisi olla kustannustehokkaampi ratkaisu. Ottaen huomioon, että haastateltavien organisaatioiden käsitykset mielekkäistä assosiativisista suhteista poikkesivat YSON nykyisistä, lienee järkevintä pitää assosiaatiokokonaisuudet ontologiasta erillään olevina moduuleina. Näitä ns. assosiaatiomoduleita voitaisiin muodostaa aihealue- ja korpuspohjaisesti. Organisaatiot voisivat valita näistä ontologiaansa ja käyttökontekstiinsa parhaiten sopivan assosiaatiokokonaisuuden. Yhteistyö sidosryhmien kanssa olisikin tässä olennaista. Yhteistyöllä organisaatioiden sovelluskehittäjien kanssa voitaisiin myös tutkia erilaisia tapoja, joilla assosiativisia suhteita hyödynnettäisiin järjestelmissä. Toinen tärkeä sovelluskehittäjien kanssa käsiteltävä puheenaihe olisi

¹⁰ Ylähierarkiaan tehdyistä muutoksista ks. Mikko Lappalaisen (2013) esitys "Ehdotus Yleisen suomalaisen ontologian (YSO) uudeksi ylärakenteeksi"

haunlaajennuksen kehittäminen ja hierarkkisten suhteiden hyödyntäminen haussa. On kuitenkin muistettava, että tällaiset kehityssuuntaukset edellyttävät resursseja.

Myös synonymian ja erilaisten diskurssien kuten puhekielen mukaan tuominen on tärkeä kentän toive. Tässä bibliometrinen menetelmien soveltaminen on huomattavasti haastavampaa ja käsityölieneekin mutkattomampi vaihtoehto. On syytä tutkia, onko synonymiaa mahdollista tuoda ontologiaan lisää nykyisillä resursseilla, ja miten realistisesti mitoitettu lisäresurssi edistäisi asiaa. Resurssikysymys huomioiden on syytä tarkastella myös muita mahdollisuuksia kuten joukkoistamista.

Käsitteille kaivataan lisää määritelmiä. Tämäkin tarve on huomioitava, mutta se on niin ikään suhteutettava resursseihin. Tarkkojen, sanakirjamaisten määritelmien ja YSON kattavan läpikäynnin sijaan on keskityttävä avaamaan käytöltään epäselviä käsitteitä tai keskenään helposti sekoittuvia käsitejoukkoja.

Haastattelujen perusteella on otettava tavoitteeksi uusien käsitteiden ehdottamisprosessin tehostaminen ja notkeuttaminen. Prosessin olisi hyvä olla nopeampi kuin nykyinen YSA-prosessi, jossa asiasanasto päivittyy neljä tai viisi kertaa vuodessa käytävän kokouksen päätösten pohjalta. On selvitettävä, miten järjestelmissä voisi käyttää myös vielä hyväksymättömiä ehdotuksia sisällönkuvailussa ja miten nämä saisi myöhemmin yhdistettyä ontologiaan ilman, että yhteyttä jo kuvailtuun aineistoon menetetään. Ehdotusjärjestelmästä on syytä tehdä läpinäkyvä: ehdotukset voisivat olla seurattavissa ja kommentoitavissa. Ehdotuksilla voisi olla oma yksilöivä tunnistensa, mikä helpottaisi niiden käyttöä sisällönkuvailussa. Uusista hyväksytyistä käsitteistä olisi tiedotettava. Prosessin uudistaminen on iso askel, joka vaatii paljon valmistelutyötä ja suunnittelua sidosryhmien kanssa. Lisäksi se vaatii kehitystyötä myös sidosryhmien omissa järjestelmissä.

Ontologianäkymää on selkeytettävä ontologiapalvelun käyttöliittymään: erikoisontologiaa tarkasteltaessa on visuaalisesti erotettava ko. ontologian käsitteet YSON käsitteistä. Lisäksi ne YSON hierarkiaa luovat yläkäsitteet, joita ei ole tarkoitettu sisällönkuvailuun täytyy erottaa visuaalisesti muista käsitteistä. Selkeyttä on mietittävä myös jaettujen käsitteiden ja monihierarkian visuaalisessa toteutuksessa. Lisäksi jaetut käsitteet on syytä käydä tapauskohtaisesti läpi ja arvioida niiden käyttökelpoisuus tiedonhaun kannalta.

Yksi keskeisimmistä toiveista koskee instanssitietoa, erityisesti paikkatietoa. Finto-palvelussa on jo nyt paikkatietorekistereitä: Paikannimirekisteri ja Suomen ajallinen paikkaontologia (SAPO). Jatkossakin palvelussa on syytä tarjota mahdollisimman laadukasta paikka- ja toimijatietoa. On selvitettävä, mitkä ovat parhaita tällaisia tietovarantoja, ja neuvoteltava asianomaisten tahojen kanssa yhteistyö- ja julkaisukysymyksistä. ASTERI-auktoiteettitietokannan nimiauktoiteetit¹¹ olisivat yksi mahdollinen tietovaranto ja sen hyödyntämistä tässä yhteydessä olisi selvitettävä, kuten myös YSAssa olleiden paikannimien. Myös mahdolliset globaalit tietovarannot on syytä ottaa huomioon. Globaaliuteen liittyy myös monikielisyyden tarve, ja ONKI-projektin myötä YSO

¹¹ ASTERI-auktoiteettitietokannan nimiauktoiteetit käsittävät teosten tekijöinä ja kohteina olevien henkilöiden ja yhteisöjen auktorisoidut ja vaihtoehtoiset nimimuodot. ASTERI pohjautuu kansallisbibliografian auktoiteettitietuisiin.

täydentyä paitsi ruotsin kielellä, myös englannilla, ja se sillataan myös kansainvälisiin keskeisiin sanastoihin, kuten Library of Congress Subject Headings.

Haastatteluissa nousi esiin myös toive ontologiakäyttöliittymästä käsitekartta-tyyppisestä selailusta. Tämän konseptia ja käyttömahdollisuuksia olisi hyvä resurssien salliessa tutkia. Finto-käyttöliittymään on tuotu jo lisää visuaalisuutta avaamalla ontologian hierarkianäkymä kokonaisuudessaan selaajan näkyville ja selailtavaksi.

Erikoisontologiat yhteen liittävän KOKO-ontologian, sen yleisen konseptin ja ylläpitoprosessin suunnitteluun ja sidosryhmien kanssa keskusteluun on varattava riittävästi aikaa. Tavoitteeksi on asetettava päällekkäisen työn vähentäminen organisaatioissa, mutta samalla on painotettava, että kaikkien sidosryhmien näkemykset tulevat kuulluksi. Konseptin suunnittelussa on myös huomioitava resurssit ja asiantuntemuksen takaaminen erikoisaloilta. Ontologiapalvelussa on syytä tarjota mahdollisuus käyttää ja selata KOKOa sekä yhtenäisenä kokonaisuutena että yksittäisten ontologioiden osalta.

7.3. Suosituksia ja kehitysideoita sidosryhmille

Ontologioita kehittäviä ja järjestelmässään hyödyntäviä sidosryhmiä sekä aiheesta kiinnostuneita pyydetään seuraamaan Finto-palvelun kehittäjien wiki-sivustoa¹², tiedotusta ja ohjeistusta. Ongelmat ja kysymykset kannattaa suunnata ONKI-projektin palautesähköpostiin: onki-posti@helsinki.fi.

Järjestelmien yhdenmukaistaminen organisaation sisällä tai yhteistyöorganisaatioiden välillä on suositeltavaa. Samat sisällönkuvailuvälineet ja -standardit edesauttavat aineistojen yhdistyvyyttä. Järjestelmien yksilöllisten tarpeiden huomioimisessa on hyvä varmistaa, että ne tukevat muiden järjestelmien ja moninaisen metadatan ominaisuuksia ja siten linkitetyn datan ideaa.

Palauteprosessin kehittäminen on myös tärkeää ja se tulisi kehittää selkeämmäksi eri järjestelmissä: palautteen aktiivinen ja jatkuva kerääminen ja sen antoon rohkaiseminen on elimellisen tärkeää ontologiapalvelun laadun ja ajantasaisuuden parantamisessa. Haunseuranta omissa tiedonhakujärjestelmissä on suositeltavaa ontologian ylläpidon kannalta, jotta voidaan vastata puutteisiin. Tuloksettomien hakujen eli nollahakujen seuranta niissä järjestelmissä, joissa se on mahdollista, on myös kannattavaa: se antaa ajantasaista ja tärkeää tietoa ontologiasta puuttuvista käsitteistä ja käsitteiden ilmaisumuodoista sekä tiedonhakijoiden diskursseista. On kuitenkin huomioitava, että autocomplete-toiminto muuttaa nollahakujen määrittelyä ja näin ollen myös niiden seuraamista.

Oman ontologian rakentamista tai asiasanaston ontologisoitua suunnittelevien organisaatioiden on hyvä huomioida, että asiasanojen tai muun halutun käsitesisällön sisällyttäminen ontologiahierarkiaan ja YSO-kontekstiin ei välttämättä ole kivuton prosessi. Hierarkisoinnissa on

¹² URL: <https://wiki.helsinki.fi/display/ONKI/>

mietittävä käsitese-semantiikkaa. Hierarkia saattaa toteutua eri tavalla ontologiassa kuin asiasanastossa. Tähän on syytä varautua ontologisointiprosessissa.

Järjestelmiin suositellaan autocomplete-toiminnon sisällyttämistä niin sisällönkuvailu- kuin tiedonhakujärjestelmiinkin. Vaihtoehtoisten termien huomioimista autocomplete-toiminnossa on syytä harkita organisaatiokohtaisesti. Sisällönkuvailujärjestelmässä kannattaa harkita käsittehierarchy (osittaista) visualisoinnista, tai välitöntä linkkiä Finto-palveluun, josta hierarkia on nähtävissä: tämä selkeyttää esim. homonymiatapauksissa.

Järjestelmän kehitystyössä tarvitaan aktiivista keskustelua sisällöntuottajien ja teknisten ratkaisujen toteuttajien välillä. Esimerkiksi selailujärjestelmän suosittelutoiminto on todettu käyttökelpoiseksi, mutta sen toiminnallisuus on monimutkainen. Käyttäjät, kuten myös kehitystyössä mukana olevat, eivät välttämättä tunne taustalla olevia algoritmeja eivätkä tiedä miten suosittelu toimii ja mihin se perustuu. Lisäksi järjestelmän tiedonhakua suunniteltaessa on hyvä muistaa, että autocomplete-haku lienee parasta kohdistaa vain sisällönkuvailussa jo käytettyihin käsitteisiin, ei koko ontologiaan. Tiedonhaku voitaisiin haluttaessa laajentaa esimerkiksi sisällönkuvailussa käytettyjä käsitteitä laajempiin käsitteisiin ja sellaisiin käsitteisiin, joita ei ole suoraan dokumenteissa käytetty, mutta jotka ovat temaattisesti silti relevantteja. Erilaiset fasettirajaukset (esim. eri asiakaskunnat) voivat karsia ei-toivottuja haku- ja suosittelutuloksia.

Myös erikoisontologioiden ylläpitäjien ja kehittäjien on hyvä pohtia sisällönkuvailun ja sanaston kehitystyön osittaista joukkoistamista. Ns. maallikot voivat tarjota rikasta ja tuoretta synonymiaa, mutta ehdotusten tarkastaminen on työlästä. Joukkoistaminen ei kuitenkaan voi korvata laadukkaan sisällönkuvailun ja sanastotyön etuja.

Ontologiapalveluiden kehitystyötä on syytä tehdä myös tiedonhakujärjestelmistä vastaavien kanssa. Tiedonhaussa ontologia on (vain) yksi osa. Muita keskeisiä osia ovat hakustrategia ja tiedonhakujärjestelmä. Erikoisontologiat ovat perusluonteeltaan jo temaattisia eli tietyn aihepiirin sisältöjä ja käytäntöjä mukailevia, mutta yleisestä sanastosta jo lähtökohtaisesti puuttuva rajatumpi näkökulma voi tuottaa hankaluuksia haun puolella. Hakujärjestelmissä temaattisuutta voidaan usein toteuttaa muilla lisäkeinoin esimerkiksi tarjoamalla aihe-, julkaisija-, dokumenttityyppi- jne rajauksia.

Kaikista nykyaikaisista huolimatta YSON käyttö mahdollistaa jo nyt erilaisten tietovarantojen sisällönkuvailun linkittymisen toisiinsa. Monikielisyyden ulottaminen Englantiin ja siltaus Library of Congress Subject Headings -sanastoon on merkittävä laajennus esimerkiksi YSAan nähden. YSON kehitystyössä paneudutaan niin sisällöllisiin kuin teknisiin haasteisiin, ja nimenomaan tiivistä vuoropuhelua eri toimijoiden kanssa käyden. Vuoropuhelu onkin keskeinen edellytys sille, että palvelusta tulisi käyttäjiensä tarpeet huomioiva.

Raportin laatineet Susanna Nykyri ja Tuomas Palonen haluavat lopuksi kiittää haastateltaviaan arvokkaasta panostuksestaan Finto-palvelun kehittämistyöhön, ja myös laajennetun projektiryhmän jäseniä. Raportin syntymiseen ja muokkaamiseen kommentteillaan merkittävästi vaikuttaneita ovat

myös ONKI-projektin muut jäsenet. Kiitos Matias Frosterus, Sini Pessala, Mirja Anttila, Satu Niininen, Maria Forsén, Mikko Lappalainen, Osma Suominen, Henri Ylikotila ja Liisa Virtanen.

Lähteet

Lappalainen Mikko, Nykyri Susanna, Palonen Tuomas 2013: *Semanttisesta ja funktionaalisesta tiedonhaun ja –tallennuksen ontologiasta : suomalaisen yleisontologian laadinnan haasteita.*

Informaatiotutkimus-lehti Vol 32, Nro 3-4, 2013. URL:

<http://ojs.tsv.fi/index.php/inf/article/view/9445>

Lappalainen, Mikko 2013: *Ehdotus Yleisen suomalaisen ontologian (YSO) ylähierarkian uudeksi ylärakenteeksi.* Powerpoint-esitys ONKI-projektin laajennetun projektiryhmän kokouksessa 9.10.2013. URL:

<https://www.kiwi.fi/download/attachments/16758668/tuomassusannamikko10092013.ppt?version=6&modificationDate=1378799521893>

ONKI, Kansalliskirjasto: *ONKI-projekti ja suomalainen ontologiapalvelu Finto.* Päivitetty viimeksi 17.1.2014. URL: <http://wiki.helsinki.fi/display/ONKI/>

Palonen, Tuomas; Suominen, Osma; Pessala, Sini; Lappalainen, Mikko; Nykyri, Susanna; Frosterus, Matias 2013: *ONKI-projektin ontologiaeditorien vertailuraportti.* Kansalliskirjasto / Kirjastoverkkopalvelut, ONKI-projekti. URL: <http://urn.fi/URN:NBN:fi-fe201306244091>

Pessala Sini 2013: *ONKI Light -ontologiaselaimen käytettävyydestä : ensimmäinen testikierros.* Kansalliskirjasto / Kirjastoverkkopalvelut, ONKI-projekti. URL: <http://urn.fi/URN:NBN:fi-fe201312127628>

Soikkeli Virva 2013: *Kirjasampokyselyn raportti.* Vaasan kaupunginkirjasto-maakuntakirjasto. Ei julkinen.

Liite 1: Haastattelurunko

Aloitus: ammatilliset taustatiedot (toimenkuva, lyhyt historia)

Teema 1: Taustatiedot

Mikä on käsityksenne ontologioista? Mihin niitä käytetään?

Mihin organisaatiossanne nyt käytetään asiasanastoa / ontologiaa?

Teema 2: Sisällönkuvailu ja tiedonhaku

Mitkä ovat sisällönkuvailun tavoitteet organisaatiossanne?

Mietittekö sisällönkuvailussa tiedonhakijaa?

Miten määrittelette tiedonhakijan ja hänen tarpeensa?

Minkälaista tiedonhakua järjestelmässänne tehdään?

Millaista tiedonhakua itse teette työtänne varten?

Teema 3: Haasteet ja puutteet toimintatavoissa ja välineissä

Oletteko havainneet jotain puutteita nykyisessä järjestelmässänne / toimintatavassanne?

Oletteko havainneet haasteita asiasanaston ylläpidossa? – Tyypillisiä haasteita?

Millaisia ongelmia ajattelette ontologian ratkaisevan?

Käyttöliittymä ja käyttökonteksti? - Vaikutus?

Tekniset resurssit? Teknisen avun tarve?

Teema 4: Ontologian käyttö ja tuntemus

Käytättekö jo nykyisin ontologiaa? / Oletteko miettineet ontologioiden käyttöönottamista?

Jos ette käytä/suunnittele käyttävänne, miksi?

Oletteko tutustuneet YSOon?

Oletteko huomanneet siinä jotain puutteita?

Entä hyviä puolia (esim. asiasanastoihin nähden)?

Helpottaako YSO jollain tavalla työtänne?

Missä asioissa YSO ei helpota työtänne? - Millaisissa haittaa?

Millaisia semanttisia ominaisuuksia olisi hyvä lisätä? – Millaisia suhteita (esim. RT)? Entä instanssit ja niiden sijoittelu? Esim. paikannimet?

Millaisia toiveita laajentaa/rajata? Temaattisesti?

Käsitettä määrittelevien selitteiden yms. tarpeellisuus? Entä muut täsmennykset?

Teema 5: Tulevat tarpeet ja yhteistyö

Kuinka tärkeäksi koette, että on olemassa oman erikoisalan ontologia / kontrolloitu asiasanasto?

Oman erikoisaihepiirin itsenäisyys vs. KOKO? YSO:n ja erikoisontologian ero? Erilaisten käyttötärpeiden mukainen selailtavuus? Millainen, miten?

Olemme perustamassa ONKI-konsortiota. Oletteko halukkaita osallistumaan oman alan ontologian ylläpitoon tulevaisuudessa?

Minkälaiset ovat asiasanastonne / ontologianne ylläpidon resurssit tällä hetkellä? Millaiset arvelette niiden olevan lähitulevaisuudessa (esim. kolmen vuoden päästä)?

Oletteko miettineet julkaisevanne tulevaisuudessa tietokantanne avoimena datana?