



17.2.2017

Yhteinen tiedon hallinta -hanke

**"Sanastoeditorin" katselmointi -työpaja**

Aika 25.1.2017 klo 12.00–15.00

Paikka VM, nh Loppupeli / etäyhteys\*  
*etäyhteystiedot kalenterikutsussa*

## Osallistujat

Miika Alonen, CSC	Jaakko Laakso, Tilastokeskus
Kristiina Asp, UM	Jessica Parland-von Essen, CSC
Dea Crichton-Turley, TEM	Tarja Pykälä, MML
Matias Frosterus, Kansalliskirjasto	Suvi Remes, VM (koordinointi)
Katri Haverinen, Lingsoft Oy	Walter Rydman, CSC
Outi Hermans, Helsingin kaupunki	Katri Seppälä, TSK
Mikael af Hällström, Vero/YSR/KMR, klo 13 alkaen	Saku Seppälä, TSK
Virpi Kalliokuusi, THL	Teija Soini, Helsingin kaupunki
Simo Kankkunen, VNK, etänä	Osma Suominen, Kansalliskirjasto, etänä, klo 13.30
Elisa Kettunen, Kuntaliitto; etänä, klo 13.30 asti	asti
Kaisa Kuhmonen, VNK; etänä, klo 12-13	

Työpaja liittyy Yhteinen tiedon hallinta -kärkihankkeen kehittämistoimenpiteisiin (sanastotyön ja sanastojen käytön edistäminen ja tähän liittyvä työkalutuen kehittäminen, ns. Yhteentoimivuuden välineistö). Työpajassa on mahdollista katselmoida ja kommentoida kehittämissuunnitelmia.

**Muistio**

Käytiin läpi "Sanastoeditori"-työkalulle hahmoteltuja käyttötapauksia, materiaalit löydettävissä Yhteentoimivuuden välineistön Eduuni-työalueelta:

- Sanastojen ryhmittely
  - o ryhmäjako määritellään joustavasti sen mukaan mikä on julkisen hallinnon yhteinen sopimus; esimerkiksi toimialaryhmittelyn mukaan, jos sitä halutaan käyttää
  - o aineistotyyppin mukaan; näitä voidaan laajentaa, "luokituslaajenostyyppejä"
  - o käsitteiden esittäminen termeittäin
    - olisi tarpeen, että pystyy näkemään tietyn sanaston listana kaikkine käsitteineen; ei tarvitse tietää mitä tietokannassa on; eli näkymä kuin printtisanastossa
  - o liittyy hallintamalliasiaan
    - esimerkki palkkasanasto >> olisiko se tuolla selattavissa "palkkasanastona", ja sellaisena kaikille yhteinen, vai olisiko se ansiorekisterinä selattavissa vai molemmilla tavoilla?
      - Kuka on taho, joka hallinnoi/omistaa?
    - Jos yhdessä aihealueessa useita itsenäisiä sanastoja, niin miten niitä tarkastellaan?
      - Muodostetaan useita sanastoja tai käsitevalikoimien kautta

- Voi olla taloushallinnon sanasto, ja sen sisällä on palkkahallinnon sanasto
  - Mitä kuvataan soveltamisprofiilina ja mitä sanastona?
    - Samoja osia monissa?
    - Soveltamisprofiili on tietomalli, joka hyödyntää sanastoa
  - Kansallisia sanastoja, joiden omistajuus on selkeä
    - ”liiketoiminta-alue” hyödyntää
  - Olisiko sitten käsitteellistä päällekkäisyyttä?
    - Pitäisi näkyä työväliseen kautta; miten saadaan fiksusti kuvattua?
    - Miten johdetut käsitteet ilmaistaan?
  - Olisiko ”kahden luokan” sanastoja, ”kansallisesti sertfioituja”, joita kaikki käyttävät; muista ilmaistaan onko johdettu tms.
  - Resurssit pitäisi laittaa sisältötyöhön?
    - On jo nyt näkyvissä kaikille yhteisiä
      - Paikkatieto?
      - Palkkatieto?
  - Editorin näkökulmasta tukee sitä, että linkityksiä voidaan tukea, mutta ei ota kantaa hallintamalliin pitäisikö
    - Vrt. KOKO, joka yhdistää erikoisontologiat; voitaisiin valtuuttaa eri tahot huolehtimaan omastaan; mitä syvällisemmin sanastot linkittyvät toisiinsa, sitä kiintoisampaa tulee
    - THL >> alkuvaiheessa kaikki laitettu sisään editoriin ja sen jälkeen harmonisoidaan
  - Alkuvaiheessa, jos tiedetään, että on hyväksytty sanastoaineistoja, niin ne kannattaisi tuoda esille käyttäjälle; status/prioriteetti
    - Helposti olla löydettävissä standardiversio, johon kehittämistä nojataan; ei oteta pohjaksi jo valmiiksi muokattua
- Lisäys ryhmittelyyn: organisaatioiden mukaan?
    - Entä, kunoOrganisaatio toimii monella toimialalla? >> fasetti
    - Asettelu käyttöliittymäsuunnitelmassa on nurinkurinen?
      - ryhmien ja tyyppien pitäisi olla joko vasemmalla tai yläpuolella hakuun nähden; vrt. vanhassa ONKI-näkymässä on aika hyvä asettelu ja toimintalogiikka
    - Julkaisijan mukaan voisi poimia faseteista
    - Esimerkkis: aiheen mukaan olisi ”palkkasanasto”; julkaisijan mukaan VM (omistaja, tulennee siirtymään Tulorekisteriyksikölle)
- Käsitteiden selaaminen kuten Fintossa, voi jäsentää vielä hienojakoisemmin (ajan mukaan ehdotukset, ajan mukaan hyväksty)
    - Ryhmät = käsitevalikoimia
  - Käsitteen tiedot syksyllä määritellyn metatietomallin mukaan
    - Onko näin, että vain yhtä käsitettä kerrallaan pystyy katsomaan?
      - Näytä kaikki käsitteet –toiminto on nykyisessä THL:n editorissa
      - Olisi ikään kuin tulostusnäkyvä >> siitä voi valita, että editoi tätä; ylhäällä tai alhaalla voisi nopeasti editoida
    - Voisi olla mahdollista myös muokata useampaa käsitettä kerralla, mutta helpotaako?
      - Pointti on yhtä aikaa näkeminen; ei saa mennä, että avaa-sulje-tallenna; ei hyppimistä näkymien välillä
      - Editointitilan ja selauksilan ero? >> ei eroa, selausnäkyvä muuttuu editoitavaksi (kirjautuneelle käyttäjälle); inline-tyyppinen editointi, ei uutta ikkunaa
        - Kuulostaa hyvälle; ei kenttäkohtaisesti

- Sen sijaan, että muokkaa napilla muokattavaksi; klikkaus termiin voisi muuttaa termin editoitavaksi; muokkaa-nappi ehkä turha
    - Tallenna-nappi ehkä oltava, pelkkä enter voi aiheuttaa virheellisiä muutoksia
    - Tallentamisen yhteydessä ilmoitus oltava, että haluatko tallentaa tai poistaa
    - Käytettävyyskonventoiden huomioon; turhien klikkausten välttäminen; helppokäyttöisyyden ja virhealttiuden balanssi
    - Käyttäjäryhmä, miten heterogeeninen?
  - Tunnisteet ja tekninen metadata
- Muutoshallinta
  - jos tehdään muutos, niin meneekö suoraan dataan; voiko tietue jäädä muutostilaan ja voiko samalla tietueella olla kaksi eri tilaa?
    - Ei heti julkaistavaa ja täydellistä vaan esim. ryhmälle versio
    - Käytännössä sanaston versiointia
    - On yksi isoista kysymyksistä; aineistossa on eri käsittelyvaiheessa olevia tietoja
    - Vocbech toimii näin, eli säilyttää alkuperäisen tilan, ja tekee luonnosehdotuksia; tietää sekä vanhan ja uuden, ja vasta ehdotuksen hyväksymisen jälkeen ei ylikirjoita olemassa olevaa
    - Aiheutuu osaltaan hyväksymisbyrokratiaa
    - On erilaisia yhteisöjä, joissa erilainen osaaminen ja toimintatavat >> tarvittaneen useampia tiloja
    - Sanastot vain ovat iteratiivisia; täytyy pystyä erottamaan hyväksytyt ja muutoksen alla olevat
- Miten oli ajateltu julkaisu?
  - Julkaisua Finton kautta; virallinen sanasto Fintossa
  - Jos työstetään vain osa aiemmin julkaistusta, niin tarvitaan molemmat osat ("alkuperäinen" osa sanastoa ja uusi/uudet lisäykset)
  - Vain julkaisuvalmisversio julkaistaan Fintossa
  - Tarve silti pitää editori avoimena; voi johtaa muuten päällekkäiseen työhön
    - Oletusarvo täytyy olla että valmistelu ja editori on avoin
      - Mm. Fintossa YSO-ehdotusjärjestelmä tulossa
  - Jos tämä on avointa, niin entä kommentointi?
    - Vaikka ei pääse muokkaamaan, niin voiko kommentoida (käsitettä, ehdottaa uutta)?
    - Palautteenanto vaatii toki sitten prosessin käsitellä kommentti
    - Pitäisi olla työversio erikseen tallennettuna
    - YSR / KMR palautekierros >> moderni tapa kerätä kommentit
    - Ainakin sanastokohtaisesti pitäisi olla mahdollisuus valita avoimuus tai sulkeminen
  - Versiointi siten, että käsitteestä voisi tehdä "ehdotusversion"
  - Organisaatietieto julkisen hallinnon autentikaation kautta
    - Prosessilähtöisesti intressiryhmät, luvittaminen on kysymys
    - Käyttäjähallintaan suunniteltu vastaava ratkaisu kuin KaPA-palveluissa
- Sanastojen hallinta
  - Ei voi mielivaltaisesti linkittää
    - Kun sanastojen määrä kasvaa, niin samoja termejä/käsitteitä alkaa löytyä useista sanastoista
  - Käsitteiden muokkaus voidaan sitoa hallintamallin eli esim. ei voi poistaa, jos on määritelty suosituksiksi
    - Kannattaako poistamisen kieltä olla aineistotasolla?
    - On tyyppitasolla, sidotaan siihen hallintamalliin terminologiselle sanastolle
    - Tilanne, jossa sanastosta tulee uusi versio ja vanha on ollut suositus
      - Vanha voidaan merkitä poistuneeksi

- Kieliversioiden osalta voidaan termien tila säätää erikseen
      - Voi olla mm. EU-velvoittavuutta
- Kentät ja niiden pakollisuus tietomallin mukaisesti
  - Track and drop -toiminnallisuuksia helppokäyttöisyyden lisäämiseksi
- Käsitevalikoimat
  - Onko varauduttu siihen, että voivat olla hierekkisia >> valikoimat voivat olla
    - YSOssa ryhmät litteitä; metatietosanastossa ryhmät hierekkisia; UNESCON tesaurus concept-groupeina
    - OKSA-sanaston osalta otsikkohierekkia käytetty ryhmien muodostamiseen
    - Usein on tarve antaa ryhmille numerointi (YSO 2-numeroinen koodi) tai notaatiokoodi; ryhmien järjestys on käytännössä se hierarkiajärjestys
    - Entä aluejaot?
      - Luokituksille tarvitaan oma tietomalli; molemmissa samoja piirteitä
      - On mahdollista, että käyttöliittymä tukee suoraan luokituksille tehtävää tietomallia; huomioitava mm. luokitusavaimet
      - Jos käsitteiden linkitys toimii sanastojen välillä, niin saako fiksusta haettua?
      - Saako toinen sanasto tiedon, että siihen on linkitetty?
        - Ei suoraan, rajanpinnan kautta tai rdf:ssä linkit molempiin suuntiin, vrt YSO ja erikoisontologiat
        - "Ei saa keksiä omaa" ja siten linkittäminen on hyvin oleellista
          - Mitä tarkoittaa linkittäminen?
            - Ei saa linkittää "omiin virityksiin"
            - Mitä velvoitteita tulee esim. linkitetyille?
            - Linkittäjien täytyy saada tieto, ja voivat tehdä päätöksen pitävätkö linkin
      - Linkitys lienee ongelma myös jos kaikki aineisto ei ole julkista.
      - Tietojärjestelmien elinkaari ei välttämättä mene samassa ajassa käsitteiden elinkaaren kanssa; "kohde vanhentunut"; Fintossa ohjataan toiseen käsitteeseen, joka ei ole vanhentunut
        - Omistajan vastuulla vahtia linkit; jos vanhentaa, niin pitää määritellä tilalle korvaava käsite
          - Jos URIA käytetty, niin mukana, jos johtaisi aina kuranttiin käsitteeseen
          - Miten vanhentuneet näkyvät? Esim. vakiona voivat olla piilotettuna.
            - Vanhan termin ottaminen käyttöön uudessa merkityksessä?
            - Juuri tämän takia pitäisi olla näkyvissä.
            - Järjestelmä tarkastaa uutta lisättäessä, että ko. termiä ei ole olemassa (kuten tietomallieditori)? Erotettava visuaalisesti käytössä olevista käsitteistä = "vesileima päälle"
- Milloin käsite muuttuu niin paljon, että siitä tulee uusi käsite – perustavanlaatua ongelma
  - Prosessi editointiin täytyy olla
- TSK:n editorin esittely
  - Sanastokeskuksessa käytössä editori terminologien ammattikäyttöön, kehitetty hiljalleen eteenpäin ilmenneiden tarpeiden mukaan
    - XML-pohjainen, helppo muokata, joustava kielenä ollut
  - Yksi keskeinen ominaisuus, että sanastokokonaisuus nopeasti hahmotettavissa näkyvässä; tiivis ilmaisutapa (esim. tietueen tiedot nähtävissä kerralla)
  - Standardinmukainen notaatio

- Yhtä sanastoa voi editoida vain yksi henkilö kerrallaan; ammattikäytössä ei vielä suuri miinus toiminnallisuuksissa
- Muokkaus mahdollista open office/word muodossa ennen pdf-generointia julkaisuun
- Sähköinen julkaisuputki TEPA-termipankkiin olemassa
- SKOS-muoto Finto-julkaisua varten olemassa
- Käsittekaavioiden piirtoon erillinen sovellus (integroitu)
  - Aineistoon tehdyt muutokset saadaan automaattisesti kaavioihin
  - Käsitteellä tieto missä kaaviossa esiintyy
- Käsitteiden linkityksiin antaa automaattisesti ehdotuksia; vertailu termien perusteella
  - Antaa tiedon myös puuttuvista linkityksistä, termiperusteisesti ehdotukset
  - "bad reference" -toiminnallisuus >> ensisijaisten termin käyttö määritelmässä ja huomautuksissa
- Erilaisia import-toimintoja, mm. datan tuonti excelistä
- Joustava metatietokenttien lisäys/poisto (valikoima täytyy tosin tietää mitä saatavilla)
- Luo tunnisteiden (sanastolle ja käsitteelle)
- Erilainen editointiparadigma; ei hallintamallia, joka estäisi esim. muutokset (ammattikäyttö)
- kehitetty toiminnallisuuksille pikanäppäimet käyttöä helpottamaan ja nopeuttamaan; TSK-makrovalikoima kehitetty
- Spark-QL-tuki? >> ei agendalla, mutta mahdollinen; päivitysrajapinta vaikeampi

#### Liitteet

Projektisivusto ja kehittämismateriaalit:

<https://wiki.eduuni.fi/display/CSCYTI/Termieditorin+kayttotapaukset>

*Rautalangat kuvaavat karkeasti käyttöliittymän toiminnallisuuksia ja sovelluskomponenttien asemointia. Rautalangoissa ei ole kuvattu kaikkia mahdollisia käsitteisiin tai termeihin liittyviä kenttiä. Kaikki terminologisiin sanastoihin sisällytettävät tiedot löytyvät Terminologisen sanaston tietomallista osoitteesta: <http://iow.csc.fi/model/st>*

#### AP:t

- kehitystyö jatkuu, keskustelusta saatu tärkeitä havaintoja huomioitavaksi
- toiminnallisuuksien logiikka ja siihen liittyen myös hallinnan prosessit kuvattava
- järjestetään uusi katselmointi, kun kehitys mennyt eteenpäin niin, että on näytettävää keskusteltavaksi