



Auktoriteettitietokanta

Taustaraportti

6.9.2010

Sisältö

1. Johdanto	3
1.1. Nykytila	3
1.2. Tavoitetila	5
1.3. Auktoriteetit KDK:n asiakasliittymässä	6
2. Auktoriteettien määrät ja ylläpidon vaatima työmäärä	8
3. Toteutusvaihtoehdot	9
4. Tekninen toteutus ja käyttöoikeudet	11
4.1. Alephin auktoriteettimoduuli	11
4.2. Käyttö muissa järjestelmissä	11
Liite: Hankkeen tarkempi aikataulu	12

1. Johdanto

Kirjastojen käyttämiin auktoriteetteihin kuuluvat

- Asiasanat
- Nimiauktoriteetit (henkilöt, suvut/perheet ja yhteisöt)
- Yhtenäistetyt nimekkeet > teosauktoriteetit
- (Luokitukset) - *esitetty jätettäväksi auktoriteettitietokannan ulkopuolelle*

Kaavio: Kirjastojen auktoriteettitietokannan viitekehys



**) ISNI = International Standard Name Identifier; ISTC = International Standard Text Code; ISWC = International Standard Musical Work Code; FRBR = Functional Requirements for Bibliographic Records; FRAD = Functional Requirements for Authority Data; RDA = Resource Description and Access*

1.1. Nykytila

Tällä hetkellä auktoriteettitietoja ylläpidetään Voyager-järjestelmässä, jossa ei kuitenkaan ole erillistä auktoriteettimoduulia. Näin ollen ne ovat olleet tietokantakohtaisia ja niiden välittäminen muihin tietokantoihin - muista järjestelmistä puhumattakaan - on ollut enemmän tai vähemmän manuaalista puuhaa koska Voyagerin rajapinnatkaan eivät ole olleet kovin standardinmukaisia.

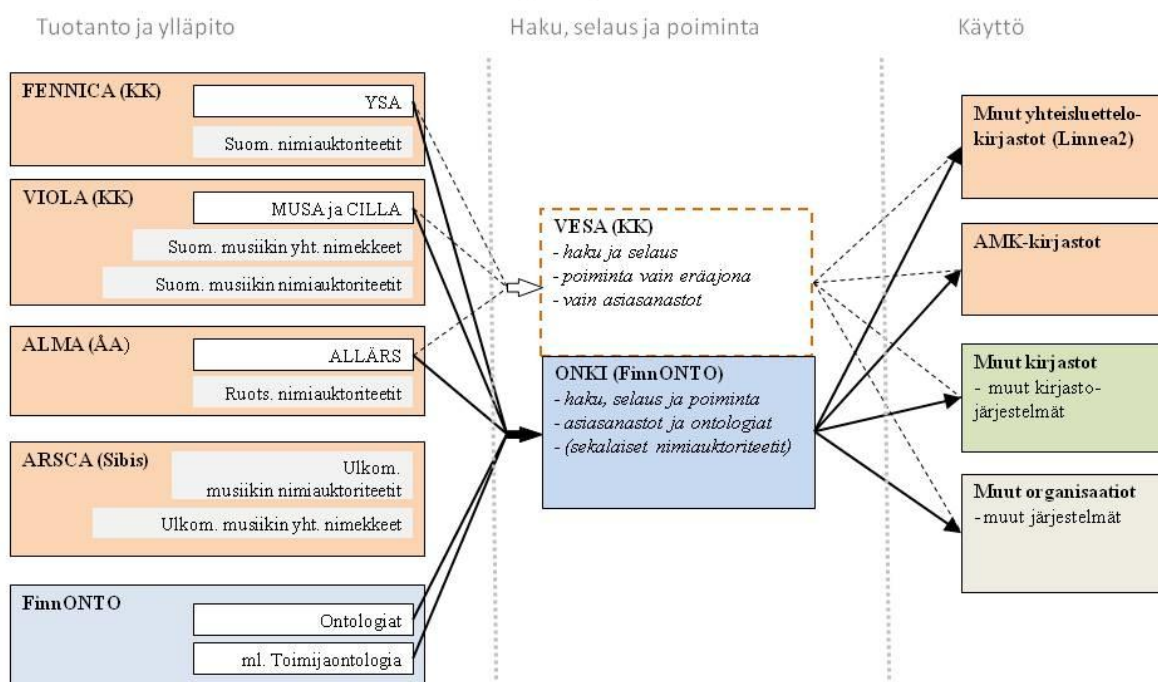
YSAa ja suomalaisia nimiauktoriteetteja on ylläpidetty FENNICA-tietokannassa ja niiden ruotsinkielisiä vastineita Åbo Akademin ALMAssa. Suomalaiset musiikkiauktoriteetit puolestaan ovat olleet VIOLAssa, kun taas ulkomaisia musiikkiauktoriteetteja löytyy Sibelius-Akatemian ylläpitämänä taidekorkeakoulujen ARSCAsta. Muilla kirjastoilla on ollut sekalaisesti omia auktoriteettitietueita omissa tietokannoissaan. Vaihto on tapahtunut pääasiassa eräajoina ja käyttö paljolti leikkaa/liimaa -toiminnolla.

Jo tekniikaltaan antiikkiseksi käyneen VESA-verkkosanaston tilalle on ratkaisuksi tarjoutunut ontologiahankkeen ONKI-palvelu. Sen etuna on paitsi linkitys ontologioihin, myös muihin kuin Kansalliskirjaston ylläpitämiin asiasanastoihin. Poiminta onnistuu leijukken avulla suoraan järjestelmiin, joissa on toimivat Web Services -rajapinnat - mitä Voyagerissa ei ole.

Lapsipuolen asemaan tässä kaikessa kehitystyössä ovat jääneet nimiauktoriteetit. RDA:n ja FRBR:n myötä kuvioon ovat tulleet myös teosauktoriteetit, jotka nekin ovat vailla omaa kotia. (Näitä kutsutaan nykyisin uudella termillä 'auktorisoidut nimimuodot'.)

Oheisessa kuvassa on tilanne tällä hetkellä, jolloin auktoriteetteja ylläpidetään Voyager-järjestelmässä:

Nykytila:



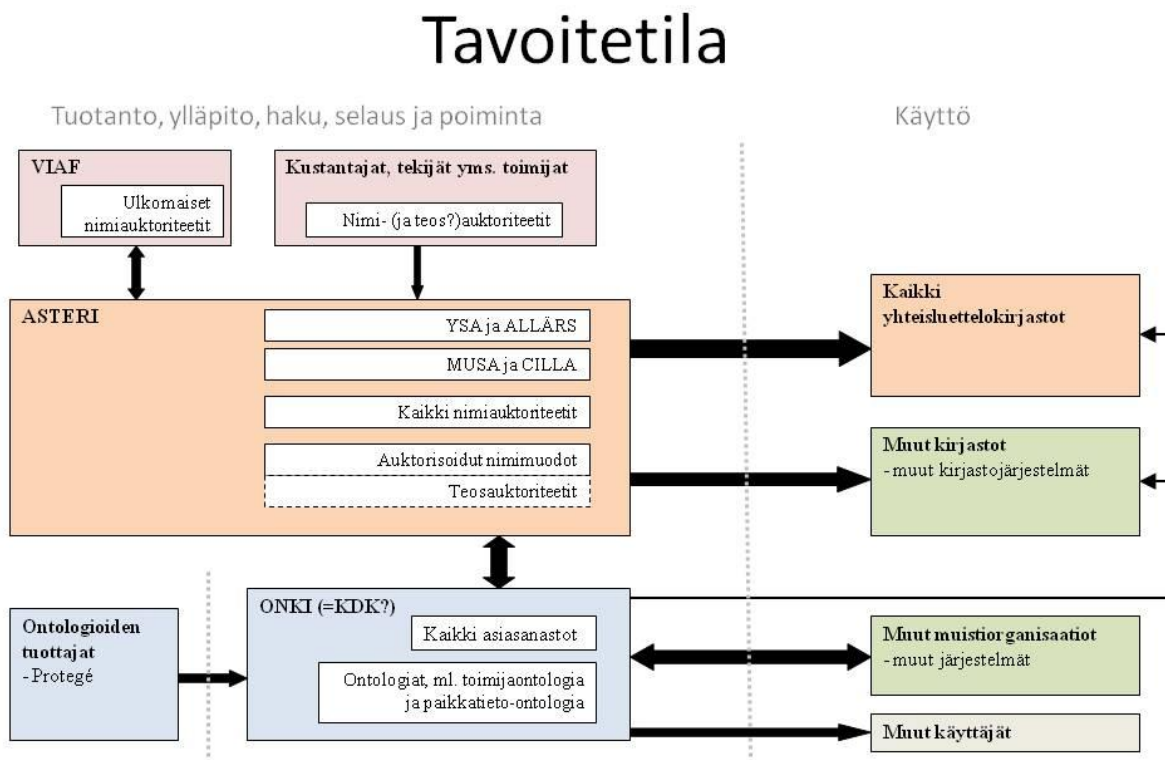
1.2. Tavoitetila

Erillisen auktoriteettitietokannan päällimmäinen hyöty on mahdollisuus yhdistää erikieliset auktoriteetit samaan tietokantaan, kun ne eivät ole suorassa yhteydessä bibliografisiin tietueisiin. Näin voidaan mahdollistaa useita hyväksytyjä muotoja (sekä suomen- että ruotsinkielinen muoto, sekä latinitsa että kyrillinen kirjoitusasu jne.). Nimiauktoriteettitiedoston avulla toisaalta erotetaan toisistaan samannimiset henkilöt tai yhteisöt ja toisaalta kootaan yhteen saman henkilön/yhteisön nimen eri variaatiot sekä luodaan suhteet nimen aiempiin ja myöhempisiin muotoihin.

Myös auktoriteettitietojen poiminta muihin järjestelmiin helpottuu, mikäli auktoriteettitietokanta on täysin erillinen ja avoin. Lisäksi se mahdollistaa auktoriteettien ylläpidon hajauttamisen.

Kun yhteisluettelo LINDA vuodenvaiheessa 2008/2009 siirrettiin Voyagerista Aleph-järjestelmään yksi radikaaleimmin muuttuneista työnkuluista koski juuri auktoriteetteja. Alephissa on niitä varten erillinen moduuli - siis valmis toimiva alusta auktoriteettitietokannalle. Suurin tekninen este poistui lähes vahingossa.

Seuraavassa on esitetty kuvaa 1 vastaava kaavio, kun auktoriteetit on erotettu bibliografisesta datasta.



Mahdollisia - ja toivottavia - yhteistyökumppaneita ovat lisäksi kustantajat sekä tutkijoiden julkaisuarkistot ja julkaisurekisterit. Kaikki nämä toimijat tuottavat valmista auktoriteettidataa, jota

ei kuitenkaan ole pystytty käyttämään suoraan hyödyksi kirjastoluetteloissa. Kustantajien käyttämän ONIX-formaatin ja julkaisuarkistojen käyttämän opinnäytteiden Dublin Core -formaatin molempien konvertointiin MARCiksi on olemassa valmiit mappaukset, vain toimivat rajapinnat tietojen vaihdolle ovat puuttuneet toistaiseksi.

Erillinen auktoriteettitietokanta helpottaa myös kansainvälistä auktoriteettitietojen vaihtoa. Kansalliskirjastossa on suunniteltu yhteistyötä VIAF (Virtual International Authority File) -hankkeen kanssa (<http://www.oclc.org/research/projects/viaf/default.htm>).

1.3. Auktoriteetit KDK:n asiakasliittymässä

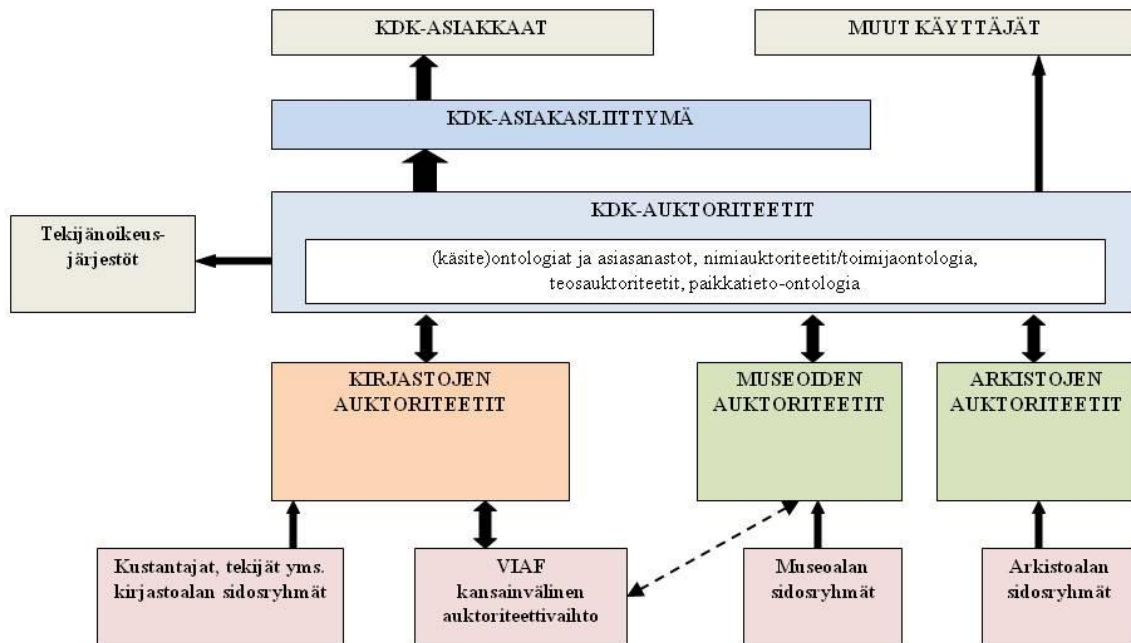
Kansallisen digitaalisen kirjaston asiakasliittymähanke puolestaan on nostanut auktoriteettitietokannan vaikeusastetta seuraavalle tasolle. Kun on saatu ratkaistua kirjastojen ja niiden sidosryhmien työnjako ja linjaukset, olisi KDK:n piirissä pystyttävä vaihtamaan näitä tietoja myös museoiden ja arkistojen kanssa.

Museoilla on esimerkiksi runsaasti auktoriteettitietoja suomalaisista taiteilijoista, jotka usein ovat aktuelleja myös kirjastoille joko teosten kuvittajina tai julkaisujen kohteina. Yhteisöjen kohdalla yhtymäkohtia on vielä enemmän. Arkistoilla puolestaan voi olla kokoelmissaan samojen henkilöiden henkilökohtaisia papereita tai yhteisöjen virallisia asiakirjoja. Kun tutkijan aiheena on Akseli Gallen-Kallela (jonka tunnetaan myös nimimuodolla Axel Gallen), hän on varmasti kiinnostunut niin taidemuseoissa olevista tauluista, kirjastoissa olevista kuvitetuista Kalevalan laitoksista ja taiteilijaelämäkerroista kuin mahdollisesti vain arkistoissa olevista henkilöhistoriaan liittyvistä dokumenteista.

Ensimmäinen ongelma on kuitenkin tallennusformaatti. Museot käyttävät pääsääntöisesti CIDOCin Conceptual Reference Modelia, Kansallisarkisto puolestaan käyttää oman alansa keskeiseen standardiin liittyvää formaattia Encoded Archival Context - Corporate Bodies, Persons, and Families (EAC-CPF). Ensimmäiseksi on löydettävä tapa mapata nämä keskenään niin että asiakasliittymän haku pystyy käyttämään niitä kaikkia hyväkseen. Tähän tarkoitukseen on esitetty KDK:n omaa auktoriteettitietokantaa.

Tässäkin työssä kuitenkin helpottaa, jos KDK:n tarvitsee keskustella vain yhden toimijan kanssa kullakin sektorilla. Ylläoleva kuva 2 näyttäisikin KDK:n näkökulmasta tältä:

Edellinen kuva KDK-asiakasliittymän näkökulmasta



Yhteisluettelon auktoriteettimoduuli toimii tässä mallissa kirjastojen auktoriteettien tuotantojärjestelmänä, jonka ytimen muodostavat kansallisbibliografian ja -diskografian auktoriteetit.

2. Auktoriteettien määrät ja ylläpidon vaatima työ määrä

Tulevaisuudessa (mahdollisesti) tehtävät uudet valinnat ja linjaukset luettelointikäytännöissä voivat tulla lisäämään auktoriteettien määrää. FRBRn ja RDAn käytön myötä auktoriteettityö korostuu ja tullee vaatimaan nykyistä enemmän aikaa.

Esimerkkinä kansallisbibliografian ja -diskografian nykyiset auktoriteettien määrät:

Asiasanat	määrä 02/2009	vuosikartunta
YSA	27.000 + 2.000 termiehdotusta	+ 1.000/v
Allärs	24.000	+ 1.000/v
MUSA/Cilla	1.800	
Nimiauktoriteetit		
Fennica: henkilöt	97.000	+ 3.500/v
Fennica: yhteisöt	34.000	+ 1.500/v
Fennica: kokoukset	1.000	
Viola: henkilö+yhteisö	40.000	+ 1.500/v
Teosauktoriteetit (tekijä+nimeke)		
Viola	17.000	

Auktoriteettejä yhteensä helmikuu 2009: 162.300 (Fennica)

Lisäksi muilla kirjastoilla on vaihtelevasti auktoriteettitietueita omissa kannoissaan. Laajimmin niitä on Sibelius-Akatemian ylläpitämänä ARSCA-tietokannassa, jossa niitä on yli 28.000 (2010). Määrä jakautuu seuraavasti:

- n. 4.700 henkilönnimeä (kenttä 100)
- n. 21.100 tekijä/nimekettä (kenttä 100 sisältäen osakentän t)
- n. 500 yhteisönnimeä (kenttä 110)
- n. 600 yhtenäistettyä nimekettä (kenttä 130)
- n. 1.300 asiasanaa (kenttä 150)
- Arscan auktoriteettitietueet sisältävät yhteensä n. 42.600 erilaista viittausta.

Yhden auktoriteettitietueen luominen/ylläpito vaatii nykyisellään keskimäärin 20 min. työaikaa, mutta vaihtelu on suurta riippuen tapauksesta. Primääriluetteloinnista se on noin kolmasosa. Nämä luvut kuvaavat karttuvan aineiston kuvailua. Takautuvaan luettelointiin liittyvä auktoriteettityö on kaikin puolin vaativampaa.

Mahdollinen auktoriteettien vaihto muiden kansallisten ja kansainvälisten toimijoiden kanssa sekä poiminta toisista järjestelmistä (esim. suoraan kustantajilta tai julkaisuarkistoista) tuo puolestaan säästöä.

3. Toteutusvaihtoehdot

Teknisten kysymysten lisäksi hankkeen kuluessa on tunnistettu puhtaasti "poliittisia" ongelmia, jotka myös vaativat ratkaisua ennen kuin uusi tietokanta voidaan ottaa täysimittaiseen tuotantokäyttöön.

Jos ylläpito hajautetaan, vaatii se kentällä lisäpanostuksen auktoriteettityöhön.

Yhteisluetteloraportin kommentteissa todetaan mm. että "auktoriteettien tallennukseen sitoutuminen on kirjastoille nykyisinä säästöaikoina todella haasteellista". Toisaalta primääriluetteloinnin osuus on esim. yliopistokirjastoissa vain noin 20% ja tästäkin osuudesta varmaan suuri osa on jo valmiiksi auktorisoitujen tekijöiden tuotantoa tai sitten nämä tiedot ovat poimittavissa valmiina muualta kuten esimerkiksi oppilaitoksen julkaisuarkistosta. Kirjastot, jotka primääriluetteloivat suurimman osan aineistostaan, kuten esimerkiksi Sibelius-Akatemian kirjasto, ovat jo ennestään panostaneet auktoriteettityöhön.

Jos puolestaan Kansalliskirjasto ottaa vastuun muistakin kuin kansallisbibliografian auktoriteeteista, aiheuttaa se lisäresurssien tarpeen Kansalliskirjastoon (joko lisää henkilöresursseja tai tiedontuottajasopimusten rahoituksen). Lisäksi esille nousee sellaisia riskitekijöitä kuin sitoutuminen ja laadunvarmistus. Erillinen auktoriteettitietokanta vaatii myös lisäpanostusta tekniseen ylläpitoon. Ohjeistus, koulutus ja yhdenmukaisuustavoitteet vaativat oman suunnittelun ja resursoinnin erityisesti hajautetussa mallissa.

Auktoriteettitietojen haravoiminen suoraan jostakin ulkopuolisesta lähteestä ei sekään ole ongelmatonta. Laadunvarmistus ja ohjeistus on siinä tapauksessa vieläkin tärkeämpää, mutta opastus käytännössä vaikeampaa, mistä on oltu huolestuneita mm. Kirjastokaapelin keskustelukanavalla alkukesästä 2010.

Oheisessa taulukossa näkökohdat on esitetty mukautettuna SWOT-kaaviona.

<i>vaihtoehto</i>	<i>hyödyt / mahdollisuudet</i>	<i>riskit / uhat</i>	<i>ylläpito / resurssitarve</i>
1. Kansallisbibliografian auktoriteetit (Fennica + Viola)	<ul style="list-style-type: none"> • lakisääteiset velvoitteet • kansainväliset suositukset • keskitetty palvelu kansallisesta aineistosta • kv. yhteistyön (esim. VIAF) suora hyödyntäminen ja siihen vaikuttaminen • palvelee myös yhteistyötä muiden muistiorganisaatioiden kanssa • palvelee myös julkaisualaa ja tekijänoikeusjärjestöjä • digitointihankkeissa tehtävä kehittämistyö saattaa tuoda mahdollisuuksia automaattiseen auktoriteettituotantoon sen aineiston osalta 	<ul style="list-style-type: none"> • jos ei voida erottaa yhteisluettelosta, ei palvele oikeana auktoriteettitietokantana • (LINDAn maksullisuus) • palvelee vain osittain kansallista yhteisluetteloa • resursointi riittämätöntä auktoriteettien osalta • VIOLAn musiikkiaineisto ei ole LINDAssa 	<ul style="list-style-type: none"> • ylläpidetään kokonaan KKssa (KKK+KVP) • verkostomainen tuotanto + vastuu KKssa (tiedontuottajasopimukset: esim. ÅA, Sibelius-Akatemia, ONIX-tuottajat)
2. Kaikki kansalliset auktoriteetit (kansainväliset muualta)	<ul style="list-style-type: none"> • työmäärän jakaminen kirjastoverkon kesken • kv. yhteistyön hyödyntäminen kirjastojen ja muiden muistiorganisaatioiden erilaisiin tarpeisiin • täydentää kansallisen aineiston kattavuutta • monikielisyyden ja erikoisalojen tuki 	<ul style="list-style-type: none"> • henkilötietolain rajoitukset (tiedot vapautukset koskevat vain kansallisbibliografiaa) • hajautettuun ylläpitoon sitoutuminen • osaaminen, laadunvarmistus • (LINDAn maksullisuus) • suomenkieliset yhtenäistetyt nimekkeet (huom! VIOLA ei ole toistaiseksi LINDAssa/Alephissa) 	<ul style="list-style-type: none"> • sisällöllinen ylläpito hajautettu > resurssitarve kentällä • tekninen ylläpito KK/KVP • ohjeistus ja koulutus vaativat lisäresursseja

Muiden kuin kansallisbibliografian auktoriteettien tuotantoa samassa järjestelmässä tullaan tutkimaan pilottihankkeen avulla, johon osallistuu 2-3 ulkopuolista tiedontuottajaa.

Keskeiset intressiristiriidat

Kansallisbibliografia:

- lakisääteiset veloitteet ja rajoitteet
- kansainvälisen vaihdon asettamat laatuvaatimukset ja reunaehdot
- kansalliset kielet (suomi, ruotsi)

Yhteisluettelo:

- yksi tietue / monta kuvailua
- useampi auktoriteettitietokanta taustalla mahdollinen, mutta ne on asetettava auktoriteettivalvonnassa prioriteettijärjestykseen
- monikielinen

4. Tekninen toteutus ja käyttöoikeudet

Opetusministeriön kanssa neuvotellaan parhaillaan toisaalta yhteisluettelon (pysyvää) vapaasta kansalaiskäytöstä ja toisaalta sen katteen laajentamisesta vuodesta 2010 alkaen. Jos näissä neuvotteluissa ei jostain syystä edetä, on myös auktoriteettitietokannan hyötyjä ja resursointia punnittava uusista lähtökohdista. Tavoitteena tulisi olla avoin tietokanta, joka palvelee kaikkia sektoreita ja myös muita yhteistyöorganisaatioita.

4.1. Alephin auktoriteettimoduuli

Ohjelmistolisenssi sisältyy yhteisluettelon (LINDA) Aleph-lisenssiin (yhteensä 5 instanssia), eli ei aiheuta lisäkustannuksia. Auktoriteettimoduuli tulee automaattisesti yhteisluettelossa mukana olevien kirjastojen käyttöön.

4.2. Käyttö muissa järjestelmissä

Muille auktoriteettitietueita voidaan tarjota osana poimintaluettelointia - joko maksutta tai maksua vastaan riippuen poimintaluettelointisopimuksen rahoituksesta. Auktoriteettitietojen selaus kuuluu vapaaseen kansalaiskäyttöön.

Auktoriteettitiedot on listattu keskeisiin kansallisiin tietovarantoihin, joita toivotaan avoimen viranomaistiedon piiriin (Open Linked Data julkishallinnossa). Ylläpidon resursointi saattaa ratketa tätä kautta.

Liite: Hankkeen tarkempi aikataulu

Q2/2010

- testikannan rakentaminen Alephiin
- FENNICAn auktoriteettien konvertointi ja loadaus

Q3/2010

- tuotantoprosessien tarkistaminen aloitetaan
- ruotsinkielisten auktoriteettien konvertointi ja loadaus kantaan ALMAsta
- monikielisten tietueiden tuotantoprosessin luominen
- musiikkiauktoriteettien konvertointi ja loadaus kantaan VIOLAsta
- uusien MARC-kenttien lisääminen ja käyttöönotto

Q4/2010

- Muiden yhteisluettelokirjastojen auktoriteetit: testauksen aloitus
- nimeke- ja teosauktoriteettien MARC-kenttien ja tuotantoprosessin tarkistus
- rajapintojen määrittely vaihtoa varten (Z39.50 yms.)
- webOPAC auktoriteettietokannalle
- indeksien määrittely hakukäyttöä varten

2011

- Muiden yhteisluettelokirjastojen auktoriteetit: tuotantoprosessien luominen
- rajapintojen määrittely vaihtoa varten (KDK, VIAF jne.)
- valmistautuminen uusien tunnisteiden (ISNI, ISTC) käyttöönottoon
- Kansallisbibliografian ja muiden kirjastojen auktoriteettien rinnakkaiskäytön pilotointi (poiminta, haku)
- päätös poliittisista linjauksista auktoriteettien tuotannossa

KDK-auktoriteettivaihto 2010-

(vastuutaho KDK-hanke)

- | | |
|------------------------------------------------------|------------------------|
| • teknisten vaihtoehtojen kartoitus | <i>aloitettu 2009</i> |
| • vaihdettavan datan laadun ja formaattien kartoitus | <i>aloitettu 2009</i> |
| • juridisten kysymysten ratkaisu (tietosuoja jne.) | <i>aloitetaan 2010</i> |
| • teknisen alustan valinta ja pystytys | <i>2011</i> |
| • tuotantokäytön vastuut ja työnjako | <i>2011</i> |