

Nimitietopalvelun projektiryhmän etäkokous alustukset

Perjantai 2020-08-28
Jarmo Saarikko, Okko Vainonen

Tilannekatsaus - Tietomallin tilanne

- NTP core - ydinmalli luotu Y-työkaluun <https://tietomallit.suomi.fi/model/ntp/>
 - Kaikki luokat ja propertyt on linkitetty vastaavaan RDA-registryssä <http://www.rdaregistry.info/>
 - Nämä viety myös Wikibaseen
- Y-työkaluun luotu myös sanasto sekä koodisto
 - KAM-sektorin nimitietopalvelun sanasto (keskeneräinen) <http://uri.suomi.fi/terminology/ntp/>
 - Kulttuurin toimijakuvailu <https://koodistot.suomi.fi/registry;registryCode=kamactor>
- Koko tietomalli
 - alunperin Word-tiedostona. Jarmo laatinut Excel-muotoon, jonka pohjalta on luotu graafi
 - graafi kokouksen agendan liitteenä
 - Graafista valittu “NTP core” propertyt, jotka viety Y-työkaluun

Tilannekatsaus - aineistojen mappauksien tilanne

Saatu aineistoja

- AHAA (EAC), Kansallisarkisto 4 yhteisöä, 4 hlö
- Asteri (Marc), Kansalliskirjasto n. 10.000 henkilöä
- Kauko (Profium sense) Espoon kaupunginmuseo n. 26.000 (suodattamaton)
- Museoliitto (MuseumPlus, LIDO export) n. 90 yht, 16.000 hlö

Heinäkuussa tehty mappaukset saatujen aineistojen sarakkeille

- tulokset Google-sheet taulukkona
- kirjattiin aineistojen yhdistelyä vaikeuttavia ongelmia sisällön rakenteessa

Tilannekatsaus - Wikibasen tilanne

- SPARQL -haku toimii <https://query.ntp-test.kansalliskirjasto.fi/> (toimii vain HY verkossa)
- API toimii - <https://ntp-test.kansalliskirjasto.fi/w/api.php>
- Tietojen lataustyökalujen testaus
 - Quickstatements - itemien luonti onnistuu, propertyjä ei
 - Wikidata Integrator testauksessa
 - pywikibot latausta ei ole testattu vielä
 - Cradle-tallennuslomaketta ei saa asennettua konttiympäristössä (toimii wbstack-wikeissä)
- Quality constraints (propertyjen tarkenteet)
 - laajennus ladattu, toimintaa testataan
- Kaikki ydinmallin (NTP core) propertyt on luotu Wikibaseen
- Wikibaseen luotu esimerkkitapauksia henkilö, yhteisö ja nomen -entiteeteistä

Tilannekatsaus - muu toiminta

Lähetetty abstraktiehdotus SWIB20 virtuaalikonferenssiin

- aiheena KAM-tietomallin RDF ja sen linkitykset, Wikibase ja kv. yhteistyö

Oltu yhteydessä Saksan kansalliskirjastoon

- tavoitteena Wikibase-yhteistyömahdollisuuksien selvittäminen
- Saksassa on testattu Wikibasea KAM-sektorin tarpeisiin yhdessä Wikimedia Deutschlandin kanssa.
- Wikibasen ja kirjastojen yhteistyö laajenemassa IFLAn Wikidatatyöryhmän kautta

2020 Q3 tavoitteet ja niiden toteutuminen

Pilottitoteutuksen tietomalli valmis

- Toteutunee: dokumentointi ja julkaisu Y-työkalussa ja RDF:muodossa

Tietomallin tarkistuspiste

- Palveleeko tietomalli nimitietopalvelun alkuperäisiä tavoitteita?
- Selvitettävissä: projektiryhmä päättää

Asteri-tietueesta Wikibaseen konvertteri

- Miten Marc-tietuenumero voidaan merkitä wikibaseen talteen (tietolähde-tietokohde)? Vaatii että WB:n Lähde-kenttään property, jossa esim. lähdejärjestelmä ja tietueen ID.
- Miten tehdään tietojen poisto, jos ladataan virheellistä tietoa.
- Selvitettävä pitäisikö olla myös toisin päin wikibasesta marcxml-muotoon? poiminta API vai sparql vai botti+export?

Päällekkäisten toimijatietojen (kuolleita henkilöitä, yhteisöjä) hankkiminen

- Ei ehditä tekemään Päällekkäisyyksien testausta
- Pohdittava miten deduplikointi tapahtuisi?

Wikibase-ratkaisun tarkistuspiste

- Toteutunee: mutta onko sovittu mitkä ovat tarkistettavat (kriittiset) asiat? Suorituskyvyn testaus?

2020 Q4 tavoitteet ja niiden toteutuminen

Tietomallin julkaisu

- onnistunee: julkaisukanavat: Y-työkalu, Github vai joku muu?
- Saadaanko RDF-malliin taulukoissa ja Wordissä listatut propertyt
- miten linkitys malleihin, joita ei ole julkaistu (kommenttiin viittaus järjestelmä kenttään?)

Päällekkäisten toimijatietojen alustava mäppäys yhteiseen tietomalliin ja yhdistelytarpeen osoittaminen

- tietueet on saatu ja kentät mäpätty taulukkona
- testataan päällekkäisyys case-tasolla

Konkreettiset yhteistyösuunnitelmat Tutkimustietovarantoon ja CSC:n PID-palveluun

- riittäisikö mäppäysten julkaisu taulukkona?

Projektin loppuraportti ja ehdotus jatkosta

- pitäisi olla valmis jo marraskuun lopussa

Tahtotila-dokumentti

OKM rahoitusta jatkokehittämiseen 2021-2024 ei myönnetty (tiedeosasto)

Samaan aikaan on todettu työryhmässä (kulttuuriosastolla), että tekijänoikeuksien hallinnassa tarvittaisiin yhteisiä toimijakuvauksia (liittyy EU-hankkeisiin)

Kootaan yhteisiä näkemyksiä KAM-sektorilla käytettävien toimijatietojen hallinnasta jatkossa (myös paikkatiedot?)

- Tavoite: rahoitus toimijatietojen yhteisen hallinnan kehittämiseen ja ylläpitoon
- Allekirjoittajat: KAM-sektorin keskeiset organisaatiot / johtajat
- Kohderyhmä: OKM?
- Tyyppi: muistio, julkilausuma, kirje vai joku muu?
 - työstetään yhdessä google docsina

Tietomallin edelleen työstäminen ja julkaisu

- Ratkaistavana:
- Viedäänkö koko malli Y-työkaluun? Julkaisu RDF-muodossa Githubissa?
- Toistettavuuden ja pakollisuuden merkitseminen malliin?
- Propertyihin Domain ja Range?
- Mihin muihin malleihin RDF:n entiteetit linkitetään?
 - <http://uri.suomi.fi/datamodel/ns/ntp#persistentIdentifierForPerson>
 - <http://rdaregistry.info/Elements/a/datatype/P50094>
 - skos:equalMatch seuraaviin LIDO, EAC, Spectrum, CIDOC CRM, Bibframe,

Käsittely yhteisessä työpajassa?

Aineistojen mappausongelmia 1

Etunimi/sukunimi-jaottelu - yksiosaiset nimet

- pannaanko aina sukunimi-kenttään? Taiteilijanimen kenttä?

Ensisijainen nimi

- sektorikohtainen: RDAn mukainen ensisijainen nimi
 -

Aineistojen mappausongelmia 2

Suluissa olevat dataa tarkentavat tekstit - esim. nimikentässä yhteisön tyyppi

- onko käytetty yhtenäisesti? Voiko siirtää omaan kenttäänsä?

Tietueiden mappaus (todellinen identiteetti, toinen identiteetti), (edeltäjä-seuraaja)

Taiteilijan nimi on varianttiniimenä eikä erillisenä tietueena

- miten todellisen henkilön ja taiteilijan nimen tiedot jaotellaan NTP-Wikibaseen
- miten tieto palautuu takaisinpäin (liitetäänkö NTP:n pseudonyymien id museon taiteilijan nimeen ja mitä muita ongelmia seuraa)?
- Vai johtaako siihen, että kaikilla oltava molemmat profiilit?

Aineistojen mappausongelmia 3

Jos kahdessa organisaatiossa eri ensisijainen nimi, kumpi valitaan yhdistettyyn tietueeseen ensisijaiseksi nimeksi

Henkilön / identiteetin varianttinitimet

- miten merkitään? onko tarpeen erikseen “identiteetin varianttিনি”?
- mallissa on nyt vain “henkilön varianttিনি”?

Aineistojen mappausongelmia 4

- vuosiluvut - EDTF formaatti? huomautukset eri propertyyn?
- kansalaisuus - MARCissa ei kenttä, Museoilla ei asuinpaikkaa
- sukulaisuussuhteet - ei ole sanastoa. Jos tätä tarvitaan, niin voitaisiin kääntää GND relation ontology?
- verkkosivut - kotisivu, tietolähde vai lisätietolinkki?
- lähdeviitteet - hieman eri kentissä eri järjestelmissä
- pysyvät tunnisteet - tyypeittäin vain yleinen tunniste-kenttä?
- MuseumPlus - yhteisön joukossa henkilöitä? Yhteisölle merkitty sukupuoli? Yhteisölle on merkitty ammatti, kun pitäisi olla toimiala?

Ehdotus: työpaja toimijatietojen yhdistelystä

- Mappaus ei onnistunut kattavasti
- Tarvitaan kunkin sektorin kuvailuohjeiden sekä järjestelmän tuntevien panostusta
- Laaditaan taulukko esim. google-sheetinä
- Työstetään kukin oman aineistonsa (sarakkeen) osalta jonka jälkeen työpaja jossa nämä mappaukset käytäisiin läpi ja tarkistettaisiin yhdessä
- Valmis mappaustaulukko voitaisiin julkaista projektiryhmän sivulla kiwissä.

Wikibase esimerkkejä

Nimien ja tunnisteiden hallinta

- propertyjen tyypit ja linkittäminen

Organisaatioiden väliset suhteet

Wikibase: nimien hallinta

- Tietomallissa nimi-ominaisuudet ovat linkkejä objektiin - seurauksena paljon erillisiä objekteja, jotka pitää linkittää keskenään.

Muutosehdotus

- Vain henkilön ensisijainen nimi luodaan kohteena (RDA object property)
- Muut nimet yksikielisinä merkkijonoina (RDAn “datatype” alaproperty)
- Osalle nimistä on oma property (etunimi, sukunimi, numero, jne.)
- Varianttiniimen tyyppi lisättäisiin tarkenteena olevaan propertyyn

Mahdollisuus: voidaan lisätä GND ontologian vastaavat luokat ja propertyt

Wikibase: auktorisoitu hakutieto

Auktorisoitu hakutieto on nyt omana propertynään merkkijonona

Muutosehdotus: Auktorisoitu hakutieto Wikibasessa: “Label + Description”

Koostaminen Wikibasessa (RDA:n ohje)

- Label-kenttään talletetaan MARC \$a osakentän tiedot
- Definition-kenttään talletetaan MARC 100-kentän muut osakentät (Hakutietojen lisäelementit)
- Lisäksi osakenttien tiedot talletetaan kukin omaan propertyynsä
- Mappaustaulukossa selvitettävä miten kootaan muista järjestelmistä
- *RDA 9.19.1 Henkilön auktorisoitu hakutieto*

Wikibase: tunnisteiden hallinta

Tunnisteeseen viittaaminen

- Luodaanko kustakin erillinen kohde johon viitataan propertyllä
- vai riittääkö merkkijono tyyppinen property

Eri tunnisteiden merkitseminen

1. Yleinen “henkilön tunniste” property
 - propertylle tarkenteeksi “tunnisteen tyyppi” (ISNI, ORCID, hetu, jne.) sekä URI.
2. Erilliset propertyt kullekin tunnisteelle (voi olla jopa 40 erilaista)
 - Luodaan jokaiselle tunnisteelle erikseen (yleisempi ratkaisu)
 - ISNI-tunniste, ORCID-tunniste, henkilötunnus, jne. Tarkenteena esim. URI

Wikibase: esimerkki tunnisteista 2 eri tapaa

ISNI-tunniste	 0000000484173266	merkkijono	 muokkaa
	URL	http://www.isni.org/isni/0000000484173266	
	tunnisteen tyyppi	ISNI-tunnus	
	▼ 0 lähdettä		+ lisää lähde
			+ lisää arvo

henkilön tunniste	 0000000484173266	tunniste kohteena	 muokkaa
	tunnisteen tyyppi	ISNI-tunnus	
	▼ 0 lähdettä		+ lisää lähde
			+ lisää arvo

Wikibase esimerkki Tunniste kohteena

0000000484173266 (Q315)

ISNI henkilölle Jarmo Saarikko

▼ Muilla kielillä ^{Maarita}

Kieli	Nimi	Kuvaus
suomi	0000000484173266	ISNI henkilölle Jarmo Saarikko
englanti	0000000484173266	Kuvausta ei ole lisätty
ruotsi	0000000484173266	Kuvausta ei ole lisätty
saksa	Nimeä ei ole lisätty	Kuvausta ei ole lisätty

esiintymä kohteesta

 pysyvä tunniste

▼ 0 lähdetä

 henkilön tunniste

▼ 0 lähdetä

tunniste henkilölle

 Jarmo Saarikko

▼ 0 lähdetä

tunnisteen tyyppi

 ISNI-tunnus

▼ 0 lähdetä

URL

 <http://www.isni.org/isni/0000000484173266>

▼ 0 lähdetä

Wikibase: esimerkki nimistä: Mika Waltari

Henkilö kohde Q64 - esimerkkejä nimien merkitsemisestä eri tavoilla

- Label: “Waltari, Mika” (merkkijono)
- Definition: ammatti (372) tai arvonimi (368), vuodet, (100\$d), (100\$\$q) tai (378)
- Alias: - eri varianttinitimet voidaan kopioida tähän kenttään wb:n hakutoimintoa varten
- nimi yksikielinen: “Wultari, Mayka” (merkkijono, henkilön muu nimi)
- nimi yksikielinen: “Mika Toimi Waltari” (merkkijono, nimen täydellisempi muoto)
- henkilön ensisijainen nimi: “Waltari, Mika” (Q256)
- henkilön varianttiniimi: Waltari, Mika T. (Q212),
- henkilön varianttiniimi: “Mika Waltari”
- etunimi: “Mika” (Q224), “Toimi” (Q225)
- toinen identiteetti: “Arne, Leo” (Q226)

Wikibase esimerkki: Seuraajat ja edeltäjät

Opiskelijakirjasto (Q335)

<yhdistymisestä muodostunut> (Helsingin yliopiston kirjasto (2010-)) (Q248)

Kansalliskirjasto (Q239)

<yhteisön edeltäjä> Helsingin yliopiston kirjasto (1919-2006) (Q240)

Helsingin yliopiston kirjasto (1919-2006) (Q240)

<yhteisön seuraaja> Kansalliskirjasto (Q239)

<yhteisön edeltäjä> Keisarillisen Aleksanterin yliopiston kirjasto (Q242)

Keisarillisen Aleksanterin yliopiston kirjasto (Q242)

<yhteisön seuraaja> Helsingin yliopiston kirjasto (1919-2006) (Q240)

<yhteisön edeltäjä> Kuninkaallisen Turun akatemian kirjasto (Q241)