

2020-03-13 Nimitietopalvelun tietomalli

Paikka: Kansalliskirjasto, Yliopistonkatu 1, 1. krs **Pohto**, (koputa ikkunaan oven oikealla puolella, jos tulet myöhässä). Etäkokous Zoomilla <linkki lähetetty osallistujille>

Aika: klo 12-14.

Ilmoittautunut: [Jarmo Saarikko](#) , [Okko Vainonen user-99b6a](#) , Anssi Lampela (Kansallisarkisto), Tommi Suominen (CSC), [Mikko Lappalainen](#) , [user-ad065](#) , [Jerry Jantunen](#) , [Piia Johanna Naukkarinen](#)

Poissa: [Matias Frosterus](#) , [Marja-Liisa Seppälä](#) , Salla Rimpinen, Leena Furu-Kallio, [Katerina Sornová](#)

Aihe

Käydään läpi tietomalli ja siitä Jarmo Saarikon tekemä kaavio mallin luokista ja propertyistä. Keskeisenä aiheena **henkilön**, henkilön **identiteettien** ja henkilön **nimien** väliset suhteet. Alla joitakin toteamuksia luonnostellusta mallista sekä niihin liittyviä kysymyksiä.

Tavoite: (alustava) päätös: Miten identiteetit kuvataan NTP-tietomallissa? Mitkä RDA:n nimientiteetit otetaan mukaan malliin?

Asialista - voi täydentyä tai tarkentua ennen kokousta

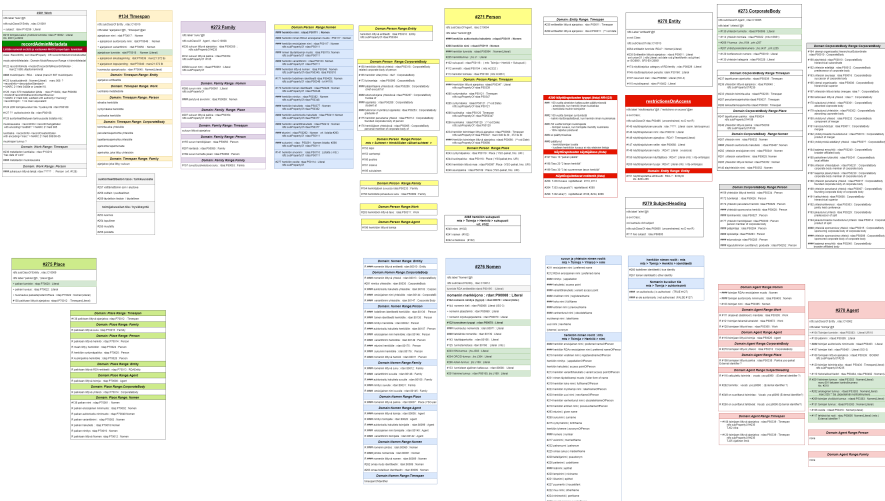
Taustaa

- Leena Furu-Kallio tiedusteli onko ollut keskusteluja Seco:n käyttämästä tietomallista ja järjestelmistä?
 - *JarmoS:* Matias on esitellyt projektia Heldigissä. Biografiasampo, Kirjasampo ja Sotasampo ainakin käsittelevät henkilöitä laajasti ja pohjautuvat mm. kirjastodataan. Ensivaikutelmani on, että ainakaan pseudonyymejä ei ole käsitelty. Jatkokeskusteluja käydään ja vertaillaan malleja, kun oma yhteinen RDA:han pohjautuva mallimme on tarkemmin kuvattu.
- Marja-Liisa Seppälä toivoo, että yhtenä kokouksen tuloksena saataisiin lista niistä tietomalliasioista, joita voitaisiin käsitellä KAM-kuvailuryhmän kokouksessa 23.3.
- Huom. katso myös kommenttiketjut sivun lopussa.

Alla kuvatussa kaaviossa olevan tietomallin selitys Nomenin osalta

- RDA-mallissa person -> nomen -> identiteetti. Personista viitataan nomeniin
- identiteetit ovat vain nomeneita
- jokaisesta merkkijonosta tehdään oma kohde, joka voidaan linkittää henkilöön
 - etu- ja sukunimi voidaan linkittää henkilöön
 - voiko sama merkkijono olla useampaan tyyppiä: esim. lempi- ja etunimi samat?
 - sama nimi, joka yhdistyy useampaan kaimaan? Nopeampi hakea?
 - riippuu teknisistä ratkaisuista: SQL, Wikidata, Opendata
- Jokainen nimiä käytetty merkkijono muodostaa oman Nomen-luokkaan kuuluvan kohteen.
 - Kysymys: kirjataanko kukin "merkkijono" malliin vain kerran? Esim. sukunimi.
- Henkilöllä on ominaisuuksia, joiden Range on Nomen.
 - Nämä relaatiot tehdään Nomeneihin, joilla on vastaava tyyppi.
 - Mikäli Nomen on tyyppiä, jolla ei ole vastaavan nimistä relaatiota, niin relaatio tehdään varianttinenä.
 - Kysymys: Voiko sama nomenin merkkijono olla yhtäkäsi eri tyyppiä? (kts. esimerkki alla #n6 vs #n25 sekä #n16)
 - Ellei voi, niin miten nimen tyyppi ilmaistaan luomatta jokaiselle omaa propertyä, mikä olisi loogisin ratkaisu. Kaikki voidaan luoda varianttinenä propertyn alaproperytinä.

Liitteenä keskeneräinen kaavio tietomallin luokista ja ominaisuuksista (3.3.2020)



RDA

- RDA:n mukaan kuvailtaisiin vain yksi **Person** -tyypin entiteetti ja jokainen nimi omana **Nomen** entiteettinä. Kaikki muut tiedot olisivat näiden liittyviä tai näiden välisiä ominaisuuksia (**propertyjä**).
- Nimien "perustyyppit": nimi, ensisijainen nimi, varianttiniimi, hakutieto, varianttihakutieto, auktorisoitu hakutieto.
- RDA <http://rdaregistry.info/Elements/a/P50111>: A "name of person" may be a real name, a pseudonym, a term of rank of nobility, a nickname, initials of a name, an assigned name.
- A name may be categorized as: Person: *preferred name of person* or Person: *variant name of person*. A name not chosen as a preferred name may be recorded as a variant name. Record this element as a value of a **Nomen entity**.
- Marja-Liisa Seppälä ehdottaa, että RDA:n henkilön nimestä otettaisiin malliin "ensisijainen nimi", "varianttiniimi" sekä "auktorisoiu hakutieto".

Identiteettien käsittely

- Henkilöllä on yksi "todellinen identiteetti". Muut henkilön identiteetit ovat kaikki tyyppiä "toinen identiteetti"
 - Huom! Kirjastosektorilla voi myös pseudonyymi olla henkilön "pääidentiteetti". Todellinen nimi olisi tällöin identiteetin varianttiniimi.
 - Kuhunkin identiteettiin liittyy prefLabel (auktorisoiu hakumuoto) ja mahdollisesti muita nimiä
 - Kysymys: Onko "identiteetti" -property "henkilön" vai "nimen" ominaisuus? Tämä kysymys toistuu tässä useita kertoja.
- Auktorisoidut nimet vs. muut nimet
 - Rajoitus: **henkilön** auktorisoitu nimimuoto voi liittyä vain yhteen henkilöön.
 - Erilaiset nimityypit.
 - Henkilöllä on vain yksi auktorisoitu nimimuoto.
 - Vain osa nimestä auktorisoidaan, muut ovat varianttiniimiä.
 - Nimen tyyppi voidaan ilmoittaa omassa propertyssä
 - Pitääkö jokaisen nimen kuulua johonkin henkilön "identiteettiin"?
 - Entä eri henkilöiden yhteiset nimet Jaetut identiteetit!
 - Sama "Nomen" voi liittyä useaan henkilöön tai olla yhden henkilön todellinen identiteetti ja toisen henkilön pseudonyymi
 - kompleksiset esitykset: yksi entiteetti, joka kuvaa useita erillisiä identiteettejä?
- Kuvataanko henkilön identiteetit nimitietopalvelun tietomallissa omana luokkana vai henkilöiden ja nimien välisinä viittauksina (propertyinä)?
- RDA
 - **Person** entiteettiin propertyt, henkilöstä viitataan nimeen (Range: Nomen)
 - **P50428** "henkilön toinen identiteetti", "has alternate identity of person", "alternate identity". — Nomen, joka on pseudonyymi tai muu henkilön käyttämä nimitys. "Aquired identity"
 - **P50429** "henkilön todellinen identiteetti", "real identity of person". — Nomen, joka on toista nomenia käyttävän henkilön todellinen nimitys. "Given identity"
 - P50428 ja P50429 ovat subPropertyjä propertylle **P50348**: "henkilöön liittyvä nomen", "related nomen of person" — Nomen, joka liittyy henkilöön.
 - **Nomen** entiteetin propertyt, nimestä viitataan henkilöön (Range: Person)
 - **P80157** "toinen identiteetti henkilölle", "alternate identity of person of". — Henkilö, jonka nimitys on pseudonyymi tai muu oletettu identiteetti.
 - **P80158** "todellinen identiteetti henkilölle", "real identity of person". — Henkilö, jonka nimitys on todellinen identiteetti.
 - JS: Huomioita
 - Henkilöstä on viittauksia monenlaisiin nomeneihin.
 - Voiko Nomen olla samalla kertaa "toinen identiteetti henkilölle" ja esim. varianttiniimi?
 - Voidaanko yksi nimi liittää vain yhteen henkilöön? Miten kuvataan, kun usealla henkilöllä on yhteinen jaettu pseudonyymi (esim. Outsider, Erin Hunter)?
 - Miten samaan identiteettiin kuuluvat varianttiniimet voidaan ryhmitellä yhteen?
 - voidaanko tällöin käyttää esim. Nomenin propertyjä
 - rdan:P80060 "johdos", "derivation" (Nomen, joka perustuu toiseen nomeniin)
 - rdan:P80061 "johdos nomenista", "#derivation of" (Nomen, joka on toisen nomenin perusta)
 - rdan:P80113 "vastaava nomen", "equivalent to" (Nomen, joka on saman entiteetin nimitys kuin toinen nomen)
- GND <https://d-nb.info/standards/elementset/gnd>
 - **:realIdentity** a owl:ObjectProperty , rdf:Property ; rdfs:subPropertyOf rdaa:P50106 , :relatedPerson ; rdfs:range :Person ; owl:inverseOf :pseudonym .
 - "Links an identity under which one or more persons act, e. g. write, compose or create art, but that is not their real name (i. e. a pseudonym) to their real identity."
 - **:pseudonym** a rdf:Property , owl:ObjectProperty ; rdfs:range :Person ; rdfs:subPropertyOf rdaa:P50105 , :relatedPerson ; rdfs:domain :Person ; owl:inverseOf :realIdentity ;
 - "Links a person's real identity to an identity under which one or more persons act, e. g. write, compose or create art, but that is not the person's real name (i. e. a pseudonym)."
 - **:relatedPerson** a owl:ObjectProperty , rdf:Property ; rdfs:domain :AuthorityResource ; rdfs:range :Person ; rdfs:subPropertyOf rdaa:P50220 agrelon#relatedAgent ,
 - **:Person** a rdfs:Class , owl:Class ; rdfs:subClassOf :AuthorityResource ; owl:equivalentClass foaf:Person , rdac:C10004 .
 - JS: Huomioita
 - identiteetti on Person luokan property, joka viittaa toiseen Person luokan entiteettiin. (yllä oleva on hieman muokattu kuvaus)
 - Mikäli identiteetti-property viittaa toiseen henkilöön, tämä property ei vielä kerro mikä on tuon toisen identiteetin käyttämä nimi.
- AHAA
 - Jokaisesta henkilötoimijan omaksumasta **julkisesta identiteetistä** voidaan laatia **erillinen toimijakuva**.
 - Toimijaan voi liittyä eri identiteettejä ja identiteetin tulee aina liittyä johonkin toimijaan.
 - **Todellinen identiteetti** = Toisen identiteetin omaksuneen henkilön todellinen identiteetti.
 - **Toinen identiteetti** = Henkilön omaksuma toinen (julkinen) identiteetti. Omaksutusta julkisesta identiteetistä voidaan luoda **erillinen kuva**, jolloin luodaan suhde henkilön todelliseen identiteettiin.
 - Toimijan kuvailussa kerrotaan parametrilla kumman tyyppinen identiteetti on kyseessä (Boolean muuttuja?)

Huom! Katso erillisellä alisivulla tarkempia [esimerkkejä ja keskustelua](#) [personien ja nimenien suhteista](#)

Yhteentoimivuusalusta

Projektiryhmän kokouksessa manitiin, että KAM-sektorin yhteisen toimija(nimi)tietopalvelun tietomalli voitaisiin julkaista myös yhteentoimivuusalustalla, kun täällä on saman tyyppisiä tietokomponentteja.

Yhteentoimivuusalustalla on Tommin laatima Tutkimustietovarannon tutkijan tiedot ym. aiheeseen liittyvää, johon voidaan linkittää mallia rakennettaessa.

- https://tietomallit.suomi.fi/model/ttv_tt/Tutkija/ - lisätietoja Tommi Suominen , CSC
 - tason verran abstraktimpi versio tutkimustietovarannon tietokantamallista, jota Tommi esitteli (kts. alla)
 - "Nimi"-propertyt: jhs:kokonimi, jhs:sukunimi, jhs:etunimi, jhs:etunimet, jhs:kutsumanimi, jhs:vaihtoehtoinennimi
 - kts. myös [semcerif:Researcher](#) , [mrd:Person](#)
 - tietomallissa tapahtumassa siirtymä entity-relationshipista linked opendataan
 - <https://tietomallit.suomi.fi/model/tutkimus> sisältää tietokomponentit
 - henkilö otettu JHS:stä
- yhteentoimivuuden API: <https://tietomallit.suomi.fi/model/iow/> (Interoperability Workbench)
- Suostumustenhallintapalvelun tietomalli: <https://tietomallit.suomi.fi/model/consent/>
- Julkishallinnon tietokomponentit:
 - Henkilö <https://tietomallit.suomi.fi/model/jhs/Henkilö/>
 - kaikki Nimi-propertyt ovat tyyppiä "xsd:string": jhs:kokonimi, jhs:etunimet, jsh:etunimi, jhs:sukunimi, jhs:kutsumanimi, jhs:vaihtoehtoinennimi, jhs:kayttajan nimi
 - kts. myös [schema:Person](#) , [cdmmd:Person](#) , [foaf:Person](#) , [corevoc:Person](#) , [person:Person](#)
 - Organisaatio <https://tietomallit.suomi.fi/model/jhs/Organisaatio/>, jhs:PublicOrganization, jhs:RekisteroityOrganisaatio, jhs:FormalOrganization
 - Kaikki Nimi-propertyt ovat tyyppiä "xsd:string": jhs:nimi, jhs:virallinennimi, jhs:preferredlabel; jhs:lyhenne, jhs:vaihtoehtoinennimi, jsh:toiminimi, jhs:rinnakkaistoiminimi, jhs:entinennimi, jhs:aputoiminimi, jhs:edeltajaorganisaationnimi, jhs:seuraajaorganisaationnimi.
 - kts myös [org:Organization](#)
 - Identiteetin hallinta <http://uri.suomi.fi/terminology/idha/>
 - [yksilöllisyys](#) (synonyymi "identiteetti")
 - Määritelmä: toimijaan tai objektiin liittyvien erottelevien ominaisuuksien joukko.
 - Huomautus: Yksilöllisyyden avulla yksilöt voidaan erotella muista yksilöistä.
 - Identiteetti on eri käsite kuin psykologian tai käyttäytymistieteiden identiteetti-käsite, joka on yksilön, lähinnä henkilön tai henkilöryhmien, käsitys omasta itsestään.
 - Liittyvä käsite: [yksilöinti](#), [yksilöintitieto](#)
 - [toimija](#) (luonnos)
 - Myös oppivat tekoälypohjaiset tekniset ratkaisut, joita ei nykyisin katsota oikeussubjekteiksi, voivat olla toimijoita.
 - [Objekti](#) voi muuttua toimijaksi, jos siihen liitetään oppiva tekoäly.
- Julkishallinnon yhteinen sanasto <http://uri.suomi.fi/terminology/jhs/>
 - henkilö, [jhs:J754](#) voimassa
 - toimija, joka on ihminen
 - luonnollinen henkilö [jhs:J7](#)
 - oikeussubjekti, joka on ihminen. Luonnollisen henkilön synonyyminä käytettävää henkilö-termiä voidaan käyttää myös toimija-käsitteen alakäsitettä edustavana terminä.
 - alakäsitteitä opiskelija, leski, asianhoitaja, haltija, oikeussubjekti, yksityinen elinkeinonharjoittaja
 - toimija, [jhs:J231](#) voimassa
 - Määritelmä: keskinäisessä toiminnassa mukana oleva aktiivinen osallinen, joka voidaan yksilöidä
 - Huomautus: Toimijat ovat yleensä oikeussubjekteja. Rekisteröimätön yhdistys on esimerkki toimijasta, joka ei ole oikeussubjekti.
 - alakäsitteitä henkilö, organisaatio, osallinen, palvelunkäyttäjä, palvelunantaja, asianosainen, laillinen eudstaja
 - organisaatio [jhs:J132](#)
 - Määritelmä: [toimija](#), joka on [henkilöiden](#) muodostama yhteenliittymä ja jolla on yhteinen tavoite ja tarkoituksenmukainen rakenne
 - Huomautus: Organisaatio voi muodostua useasta eri organisaatiosta. Esimerkkinä keskusjärjestö, joka koostuu useammasta alajärjestöstä tai kansanliike, joka koostuu sekä henkilöistä että organisaatioista. Organisaatio voi olla rakenteeltaan jäsentymätön, kuten kaljakellunta tai ravintolapäivä.
 - Muutoshistoria: YSR 16.12.2016: Vanha määritelmä oli laadittu jo vuosia sitten, viittausta sosiaaliseen järjestelmään ei koettu mielekkäänä. Käsite:organisaatio haluttiin käsite:toimija alikäsitteeksi vastinpariksi käsite:henkilö:lle. Vanhasta määritelmästä säilytettiin tavoite ja rakenne.
 - alakäsitteitä ryhmä, liikelaitos, aliorganisaatio, osasto

Keskustelua

Tommi S esitteli tutkimustietovarannon tietokantamallia:

- sisältää rahoituspäätökset
- PIDit keskeisessä roolissa, erityisesti ORCID
- ei identiteettitietoa vaan vain nimitieto rahoituspäätöksissä
 - jos ei ORCIDia on vain pelkkä nimi eikä muodostu identiteettiä
- laki tutkimustietovarannosta tullee voimaan ensi vuonna
- dim_name = identiteettömiä nimiä
 - tällaisia ilmentymismuotoa miljoonittain
 - kaksi henkilöä, jolla sama nimi, ei voi yhdistää, koska virhemahdollisuus
- dim_known_person
 - ORCID tiedossa
 - linkitetään dim_nameen
- dim_pid: yhden suhde moneen, henkilöllä voi olla monta tunnistetta
- yliopisto lisää tutkijan avulla ORCIDEja useampaan teokseen

- useampi dim_name yhdistetään dim_known_personiin
- fact_contribution
 - taulu, joka yhdistelee tekijän tieoja
 - dim_publication_id
 - dim_name_id
 - dim_organisation_id = affiliaatio
- Samalla id:llä voi olla useampia nimiä, ei tarvitse ottaa kantaa, mitä niistä käytetään
- Finto päälähteenä organisaatioiden edeltäjille ja seuraajille
- Kenneth A.: faktaa, että samoista henkilöstä tulee olemaan useampia tietueita, mutta niitä yhdistellään

Muu keskustelu:

- Tommi: onko mahdollista, että on nimiä, jota ei linkitetä identiteettiin?
- Piia: yksi vaihtoehto: jos tiedetään vain nimi, tehdään toimija, josta ei oteta kantaa onko todellinen/toinen identiteetti
- Tommi: jos ei luoda identiteettiä, kun ei tiedetä henkilöllisyyttä ei tehdä virheitä
- Piia loisi toimijan, mutta ei ottaisi kantaa identiteettiin
- Mikko: kirjastossa ei tehdä auktoriteettia jokaisesta tekijästä
- Kenneth: jos tehdään toimijatietokantaa, ei voida ulkoistaa identiteettikontrollia, kuten nyt tutkimustietovarannossa
- Anssi: mitä toimijoita viedään, mitkä on minimitiedot?
- Jarmo ajatellut, että kaikki tunnistetut tekijät auktorisoidaan
- Okko: ISNI-projektissa havaittu, että mahdollisimman täydellinen metadata auttaa yhdistelemään tekijöitä. ISNissä myös teostietoja, mutta niitä ei tässä NTP-mallissa ole
- Tommi: Google Scholar käyttää algoritmia: kenen kanssa joku julkaisee, näillä verkostoilla käytetään todennäköisyyspohjaista vertailua
- Jarmo: luodaan pseudonyymille myös todellinen identiteetti, kun sellainen löytyy
- Jokaiselle (etu-, suku- ja koko)nimelle oma kohteensa?
 - Kenneth: kuulostaa tekniseltä ratkaisulta
 - Mikko: mikä toteutettavissa master-tietokannassa ja mikä kuvailujärjestelmien kannalta?
 - ylläpito?
- Anssi: AHAA-palvelussa keskusteltua: kun arkistoon luovutettu aineisto, luovuttaja halunnut salata todellisen identiteettinsä

Päätökset

- Miten identiteetit kuvataan NTP-tietomallissa?
 - luodaan samaan tapaan kuin Saksan GND- ja AHAA-mallissa, joissa jokaisesta julkisesta identiteetistä laaditaan oma toimijakuva
 - ei oteta kantaa identiteettiin (todellinen/toinen), jos se ei ole tiedossa
- Mitkä RDA:n nimientiteetit otetaan mukaan malliin?
 - ensisijainen nimi, varianttiniimi, auktorisoitu hakutieto
 - varianttiniimien eri tyypit voidaan määritellä eri propertyillä
- Miia ja Kenneth esittelevät edellä olevat kaksi kohtaa KAM-kuvailuryhmän kokouksella 23.3.2020

Seuraava kokous

- 28.4. klo 10-12. Mitkä ominaisuudet muodostavat ehdottoman minimin yhteisten auktoriteettien kuvaamiseen?
- Eli mitkä olisivat "pakolliset kentät"? Osa näisät oliskin jo mainittuna tietomallissa (MVP, minimum viable product)