

Matias Frosterus ja Marja-Liisa Seppälä

# LINKITETTY DATA ON KIRJASTOJEN SUOLA

Linkitetty data muuttaa kirjastotiedon yhä herkullisemmaksi, myös kirjastomaailman ulkopuolisesta näkökulmasta. Samalla se tarkoittaa muutoksia kuvailijan toimenkuvaan ja uusia haasteita, jotka kehityksessä on huomioitava.

**L**inkitetty data tarkoittaa dataa, jossa asioihin viitataan toiminnallisilla tunnisteilla. Tässä yhteydessä sana ”asia” käsittää sekä tyypilliset tunnistettavat kohteet, kuten henkilöt, paikat, julkaisut jne., että näiden väliset suhteet. Eli jos kuvailun kohteena on Signum-artikkeli ja sen kirjoittaja, yksilöivät tunnisteet tarvitaan artikkelille, kirjoittajalle ja näiden väliselle ”tekijä”-suhteelle.

Toiminnallisuus puolestaan tarkoittaa, että tunniste toimii verkko-osoitteena, jota seuraamalla löytää tunnisteeseen liittyvää metatietoa. Metatiedon pitäisi myös olla koneluettavassa muodossa, jolloin se sisältää sääntöjä sille, minkälaisia päätelmiä kone saa metatiedon perusteella muodostaa.

Linkitetyn datan hyödyllisyys syntyy, kun linkeistä muodostuu laaja verkko, jossa tietoa voidaan uudelleenkäyttää. Kun henkilöön viitataan samalla tunnisteella päästään aina samoihin metatietoihin käsiksi ja varmistetaan, etteivät samannimiset toimijat sekoitu toisiinsa. Kun metatietoja rikastetaan, ne rikastuvat kaikissa yhteyksissä, missä tunnistetta on käytetty. Käytännössä tämä mahdollistaa uusia

näkökulmia hakuun, kun aihetta on mahdollista lähestyä periaatteessa mistä päin linkitetyn datan verkkoa tahansa.

## Kuvailija yhteyksien kutojana

Kuvailijalle linkitetty data tarkoittaa ohjaavampaa roolia kuvailuprosessissa. Automaation ja tekoälyn voi antaa tehdä tylsät työt, esimerkiksi tietojen jäljentämisen aineistosta. Kuvailija voi keskittyä analysointia vaativiin osa-alueisiin.

Kuvailu lähtee aina aineiston analysoimisesta ja vertaamisesta metatietojen mallinnukseen. Linkitetyn datan tietomalli pyrkii olemaan rakenteeltaan niin joustavasti laajennettava, että se nielee myös aineiston uudet variaatiot. Kirjastomainen (julkaistu) aineisto kannattaa kuitenkin lähtökohtaisesti kuvailla kirjastojen tietomalleilla ja vastaavasti arkistomainen ja museomainen aineisto näiden omilla tietomalleilla näkökulmien eroavuuksien vuoksi.

Kansainväliset kuvailusäännöt ohjaavat kuvailijaa valitsemaan kuvailuun sellaisia tietoja, jotka ovat hyödyllisiä tiedonhakijan kannalta. Tietojen merkitsemistä ohjaavat kuvailusääntöjen lisäksi standardit ja sanastot. Aineiston metatiedot tallennetaan useimmi-

ten kansalliseen tietokantaan yhteiseen käyttöön, ja aineiston metatietoihin linkitetään muissa kansallisissa tietokannoissa olevia tietoja tekijöistä, paikoista ja aiheista. Näin metatiedot ripustetaan osaksi metatietojen kansallista ja kansainvälistä verkostoa, mikä antaa tukea tiedonhakuun.

### Laajempi näkökulma linkitettyyn dataan

**L**inkitetty data yhdistelee tietoa eri lähteistä suuremmaksi kokonaisuudeksi. Jos linkkejä seurataan automaattisesti, nousee tiedon luotettavuuden merkitys yhä tärkeämmäksi. Käytännössä luotettavuutta lienee helpoin huomioida provenienssin alkuperämerkintöjen kautta. Tarvitaan käytäntöjä ja sääntöjä siitä, miten tiedon alkuperä merkitään ja miten se huomioidaan, kun tietoa laajennetaan linkeillä. Tärkeää on myös, että linkkejä ei tehdä harkitsemattomasti ja että datan konteksti säilyy eikä sitä käydetä virheellisesti väärissä yhteyksissä. Läpinäkyvyys loppukäyttäjän suuntaan nousee sitä tärkeämmäksi mitä älykkäämmin kone hakuja tulkitsee ja parantaa.

Kun tiedon uudelleenkäyttöarvo nousee, käy myös järkeen laajentaa yhteistyötä kuvailussa. Digime-yhteistyön kautta kulttuuriperintöaineistojen kuvailua mietitään soveltuvien osin yhä enemmän yhteisesti, mutta yhteistyös-

### Tarvitaan käytäntöjä ja sääntöjä siitä, miten tiedon alkuperä merkitään ja miten se huomioidaan, kun tietoa laajennetaan linkeillä.

sä on varmasti vielä tiivistämisen varaa. Edellä mainitut henkilöt, yhteisöt, paikat ja asiasanat ovat paljolti yhteisiä kirjastoille, arkistoille ja museoille, ja näiden yhteistuotanto ja -käyttö tehostaa kuvailua ja mahdollistaa paremmat haut tiedonhakijoille.

Linkitetty data tarjoaa toki myös uusia haasteita. Avautuvien mahdollisuuksien hyödyntäminen edellyttää järjestelmäkehitystä ja yhteistä sopimista. Kun hajallaan oleva data integroituu vahvemmin yhteen, nousee uudeksi haasteeksi myös linkkien ajan tasalla pitäminen. Pysyvät tunnisteet, joiden kautta tietoa löytyy, vaikka tiedon säilytyspaikkaa vaihdetaan, auttavat joiltain osin, mutta loppuviimein linkitetty data tarkoittaa kasvavaa vastuuta omasta datasta. Kolikon toinen puoli toki on, että linkitetty data on mahdollisuus nostaa kirjastodatan kiinnostavuutta kirjastomaailman ulkopuolelle.

Kun data on helpommin hyödynnettävissä ja linkittyy laajemmin, voi laadukas ja neutraalista näkökulmasta tuotettu metatietomme kasvattaa kirjastojen roolia suomalaisessa data-taloudessa. ❖

### Kirjoittajat

MATIAS FROSTERUS  
Kansalliskirjasto  
matias.frosterus@helsinki.fi

MARJA-LIISA SEPPÄLÄ  
Kansalliskirjasto  
marja-liisa.seppala@helsinki.fi