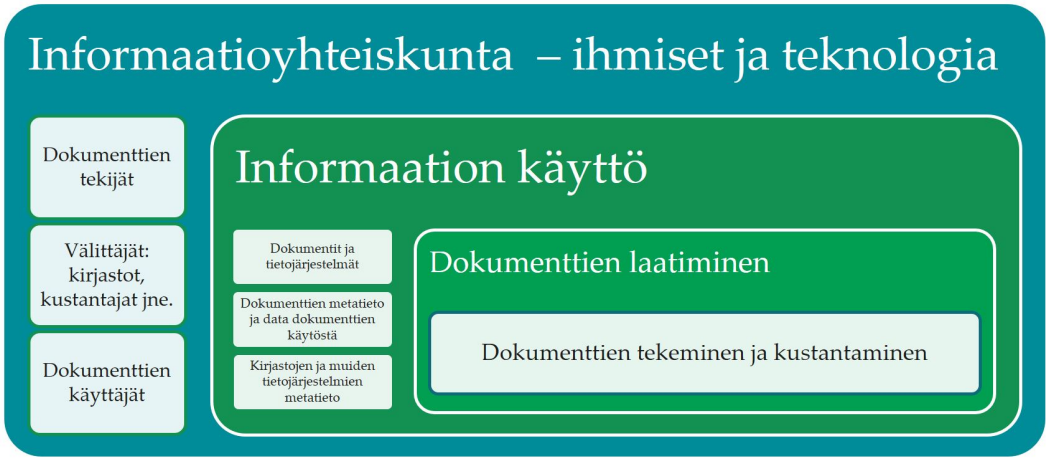
Artikkeli Tietolinjaan

Kuvitusehdotus artikkeliin: otsikkona Metatiedon käyttöympäristö



Teksti:

Katoavan tiedon jäljillä - metatietotyön merkitys digitoituvassa tietoyhteiskunnassa

Julkaisemisen kasvun myötä muistiorganisaatioiden - kirjastojen, museoiden ja arkistojen - rooli tiedon saatavuuden tukijoina on entistä tärkeämpi. Joidenkin arvioiden mukaan puolta tieteellisistä artikkeleista eivät lue muut kuin tekstin kirjoittajat ja lehtien toimittajat. Muistiorganisaatiot ovat luoneet sekä käytänteitä että tietojärjestelmiä tiedonhaun tueksi. Aineistojen kuvailutietojen (eli kuvailevan metatiedon) tuottaminen mahdollistaa tiedonhaun ja dokumenttien järjestämisen digitaalisiksi kokoelmiksi. Kansainväliset standardien noudattaminen kuvailussa tekee kuvailutiedoista yhteentoimivia, mikä parantaa hakutulosten kattavuutta ja datan siirtoa eri tietojärjestelmien välillä. Suomessa kuvailustandardiryhmä edistää kuvailustandardien noudattamista kirjastoissa ja näihin liittyvän osaamisen kehittämistä.

Suurissa digitaalisissa dokumenttimäärissä perinteinen tapa hallita aineistoja ja niiden sisältöjä ei enää yksistään riitä: tarvitaan lisää yhteistyötä metatiedontuotantoon, uusia tuotantotapoja ja metatiedon avointa jakamista.

Kuvailuyhteistyö

Kansainvälisten standardien mukaan tuotettua kuvailutietoa voi hyödyntää kuvailuyhteistyössä, mikä vähentää päällekkäistä työtä. Standardit myös määrittävät, minkä tyyppisiä kohteita kuvaillaan ja mistä näkökulmasta kuvailuyhteistyöhön tullaan. Esimerkiksi paikat, toimijat ja aiheet ovat luonnollisia kuvailuyhteistyökohteita muistiorganisaatioiden ja muiden kuvailutietoja tuottavien tahojen kanssa. Yhteistyömahdollisuudet näiden kohteiden kuvailutiedoissa ovat suuremmat kuin aineistojen kuvailussa. Paikkatietopalvelun selvityksessä ja Nimitietopalvelun selvityksissä on kartoitettu mm. kirjastojen, arkistojen ja museoiden yhteistyömahdollisuuksia paikkojen ja toimijoiden kuvailussa. Aiheiden kuvailuyhteistyössä keskeisiä ovat Finto-palvelun sanastot.

Kirjastoaineistojen kuvailutietojen tuotannossa voi yhteistyötä tehdä lähinnä julkaisu- ja tutkimusalan kanssa. Vaikka julkaisualalla noudatetaan eri kuvailustandardia (ONIX-standardin sovellusta) kuin kirjastoissa (RDA- ja MARC 21 -standardit), on kuvailun kohteena julkaistu aineisto. Kuvailun laatua parantavat yhteistyömahdollisuudet kirjastojen ja julkaisualan välillä ovat mm. ohjeistusten yhdenmukaistaminen, pakollisuuksien lisääminen ja yhteisten termien käyttö. Arkistojen ja museoiden aineistot ja tapa kuvailla aineistoja eroavat niin paljon kirjastojen tuottamasta kuvailusta, että yhteistyöhyötyä saada an lähinnä vain yksittäisten kuvailuelementtien termeissä. Yhteisten termien käyttöä tukee kansallinen Metatietosanasto.

Kirjastosektorin kansalliset kuvailuryhmät (mm. Kuvailustandardiryhmä, Kumea ja Sisällönkuvailun asiantuntijaryhmä sekä kirjastojen, arkistojen ja museoiden kuvailuyhteistyöverkostot ovat keskeisessä roolissa standardien mukaisen kuvailun toteutumisessa. Järkevästä hajautettu kuvailutietojen tuotanto yhdessä keskitettyjen kuvailutietovarantojen kanssa parantavat kuvailutietojen laatua ja yhteentoimivuutta.

Kuvailutyö tiedeyhteisöissä

Yhteistyöverkostot ovat tarpeellisia kaikenlaisissa organisaatioissa. Esimerkiksi korkeakouluissa, joissa tuotetaan suuria määriä dokumentteja ja muuta dataa, on keskeistä, että kaikki toimijat noudattavat standardinmukaisia metadatakäytänteitä. Tutkijat tuottavat julkaisuja yliopiston omiin sarjoihin tai ulkopuolisiin julkaisuihin. Nykyisin on tapana, että kirjoittajat tekevät myös omiin dokumentteihinsa kuvailun. Kirjaston roolina on ohjeistaa ja opastaa käyttäjiä käyttämään alansa työvälineitä ja samalla lisätä ymmärrystä laadukkaan metadatan merkityksestä omien julkaisujensa löytyvyydelle suurista julkaisu ja datamassoista.

Metadatan merkitys organisaatioissa ei rajoitu vain julkaisutoimintaan. Uusissa tietoympäristöissä, joissa viestitään ja keskustellaan digitaalisilla valineillä (esim. yhteisöjen omilla intraneteillä ja vastaavillä yhteiskäyttöisillä sovelluksilla), laadukas kuvailu ja metadata auttavat löytämään tarvitun tiedon ja parantavat viestien mahdollisuutta mennä perille vastaanottajalleen. Kirjastojen tulisikin tarjota rohkeammin metadataosaamistaan myös näissä uusissa tavoissa viestiä ja tuottaa metadataa. Voikin väittää, että perusosaaminen metadatasta ja sen tuottamisesta kuuluu jo lukiolaisten yleissivistykseen ja työelämätaitoihin, korkeakouluopinnoista puhumattakaan.

Kuvailutyö yleisissä kirjastoissa

Yleisissä kirjastoissa tehtyä kuvailutyötä on kehitetty ja systematisoitu viimeisinä vuosikymmeninä, ja työprosessit ovat läpikäyneet voimakkaan muutoksen. Yksittäisten kirjastojen luettelokorteille tekemistä kirjauksista on siirrytty sähköisiin tietojärjestelmiin, jotka mahdollistavat kuvailutietojen poiminnan useista eri tietokannoista. Kuvailutyötä on pyritty keskittämään, mutta yhä edelleen Suomen 284 kunnan- ja kaupunginkirjastossa tehdään päällekkäistä työtä. BTJ:n rooli kuvailutietojen tuottajana on yleisille kirjastoille edelleen merkittävä.

Kuvailun infrastruktuuri on yleisissä kirjastoissa pysynyt samana, eikä kuvailutyötä ole vielä organisoitu kansallisella tasolla niin tehokkaasti kuin järjestelmät antaisivat mahdollisuuden. Tästä huolimatta yleisten kirjastojen kuvailutyöhön käytettyjä resursseja on vähennetty, ja kuvailujen laatu ja määrä kärsivät riittämättömistä resursseista. Esimerkiksi tietokantojen yhdistämisissä syntyviä tuplatietueita ei ehditä yhdistelemään, ja tämä hankaloittaa verkkokirjastojen käyttöä ja näkyy myös kirjastojen yhteisissä verkkopalveluissa kuten Finnassa.

Yleisten kirjastojen kokoelmissa painottuvat painetut aineistot ja av-tallenteet, ja aineistoa hankitaan ennen kaikkea käytettäväksi ja lainattavaksi. Lainauksenvalvonta ei edellytä monipuolisia kuvailuja, mutta osana kirjastoverkostoa myös yleisten kirjastojen tulee tuottaa standardinmukaista kuvailutietoa. Standardinmukaisuus mahdollistaa myös sen, että yleisissä kirjastoissa voidaan hyödyntää muualla tuotettuja kuvailuja. Hyvin tuotettu metadata on myös erilaisten data-analyysien edellytys. Kokoelman sisällöistä, painotuksista ja käytöstä tehtävät analyysit edellyttävät laadukasta metadataa.

Vähitellen yleisiä kirjastoja on alkanut liittyä osaksi metatietovaranto Melindaa. PIKI-, Vaski- ja Anders-kirjastot tuottavat jo kuvailutietoja Melindaan, Kuopion ja Seinäjoen alueen kirjastoissa liittymisprosessi on meneillään. Melindassa toimiminen edellyttää sitoutumista yhteisiin pelisääntöihin, mm. RDA-kuvailusäännöstön käyttöönottoa, mutta kuuluminen isompaan kuvailuja tuottavaan yhteisöön antaa myös tukea ja auttaa parantamaan kuvailujen laatua. Jatkossa nähdään se, vähentääkö Melinda-yhteistyö kuvailutyön tarvetta tai muuttaako sen luonnetta yksittäisissä kirjastoissa.

Kansainvälinen yhteistyö

Kansainvälistä kuvailuyhteistyötä voidaan kuvailustandardityön lisäksi tehdä käsitepohjaisten sanastoissa sekä paikkojen, toimijoiden ja teosten kuvailussa. Yhteistyön edellytyksenä on kuvailukohteiden identifiointi kielineutraalien tunnisteiden kautta. Ideaalissa tilanteessa yhden kohteen kuvailu tehtäisiin vain kerran maailmassa ja tätä tietuetta hyödynnettäisiin tavalla tai toisella niiden kirjastojen järjestelmissä, joiden kokoelmissa ko. nimekettä on. Tarvittaessa kuvailutietoa voisi rikastaa esimerkiksi paikallisuuteen liittyvillä käsitteillä. Periaatteessa tämän toteuttamiseen on välineet, mutta lisäksi tarvittaisiin vahva tahtotila, prosessien uudelleenajattelua ja vähän rahaa.

Kuvailutyö arkipäiväistyy ja tuo alueelle mukaan erilaisia toimijoita. Standardit ja yhteiset käytänteet helpottavat myös heidän toimintaansa. Taulukossa 1. on kuvattu esimerkinomaisesta kuvailutyöhön vaikuttavia ohjeita, standardeja ja käytäntöjä. Siinä merkittävää on huomata, että kuvailutyötä ohjaavat myös monet kirjastoalan ulkopuoliset auktoriteetit. Näiden kaikkien noudattaminen vasta mahdollistaa tiedon haun ja datan siirtämisen digitaalisessa ympäristössä.

Kirjastoalan sisäiset ohjeet ja standardit	Julkaisualan ohjeet ja standardit	Alan ulkopuoliset ohjeet ja standardit
Kuvailusäännöt ja ohjeet	Standinumerot	ISO standardit
Sanastot, ontologiat ja auktoriteettitiedot	Sanastot, ontologiat ja auktoriteettitiedot	Sanastot, ontologiat ja auktoriteettitiedot
Tiedonsiirtoon liittyvät standardit	Tiedonsiirtoon liittyvät standardit	Tiedonsiirtoon liittyvät standardit

Taulukko 1. Kuvailutyöhön vaikuttavia ohjeita, standardeja ja käytäntöjä

Uudet menetelmät

Monet tiedonhakijat luottavat automaattisten hakukoneiden tarjoamiin listauksiin ja ajatukseen tekoälystä, joka korvaa ihmisen tekemän tulkinnan, kuvailu- ja analysointityön. Automatiikkaa hyödyntävien järjestelmien rajoitukset näkyvät tiedonhaussa: hakutulosten sisältöjen paikkansapitävyydestä tai laadusta ei ole taetta eikä hakutulosten relevanssi ole aina hyvä. Tiedonhakua voidaan myös manipuloida kaupallisilla tai poliittisilla tavoitteilla.

Kuvailu- ja analysointiosaamista tarvitaan myös ohjelmistokehityksessä: automaattisten sisällönkuvailu- ja tekoälysovellusten kehittäminen vaatii inhimillisen päättelyn ja tulkinnan mallintamista. Tämän kehitystyön myötä laajojen verkkoaineistojen automatisoitu kuvailu ja tiedonhaku voidaan tehdä analyyttisemmäksi, luotettavammaksi ja tehokkaammaksi.

Erilaisia alakohtaisia ohjaavia auktorisoituja sanastoja ja thesauruksia on aina käytetty kuvailun laatua- ja käytettävyyttä parantamaan. Näiden sijaan nykyisin on Finto-palvelun avoimen rajapinnan kautta käytettävissä sanastoja ja ontologioita. Ne päivittyvät säännöllisesti ja ovat ajantasaisia toisin kuin kuin painetut edeltäjänsä. Ne ovat myös integroitavissa kuvailijan työpintaan, mikäli käytettävissä oleva kuvailutyökalu sen mahdollistaa.

Kirjastot maailmanlaajuisesti ovat jo vuosia toivoneet kuvailun tuotantojärjestelmien pikaista uudistamista. Toiveista huolimatta järjestelmiä kehittävät tahot ovat olleet hitaita työlään uudistamisprosessin käynnistämisessä. Muutos nykyisistä järjestelmistä linkitetyn datan tuotantoa tukeviin modulaarisiin järjestelmiin on suuri ja toteutunee vasta 2020-luvulla. Kirjastojen vanhentuneet järjestelmät estävät osittain kuvailuyhteistyön toteutumisen muiden metatiedontuottajien kanssa.

Datan avaaminen

Dataa tuotetaan ja käytetään muuallakin kuin kirjastoissa. Yhä moninaisemmat datan käyttäjät ja käyttötavat edellyttävät datalta laatua ja standardienmukaisuutta. Avattua dataa saattaa käsitellä toinen tieto-organisaatio, kuten museo tai arkisto, tai vaikkapa yksittäinen sukututkija, jolla on tarvittava osaaminen avatun datan hyödyntämiseen.

Metatieto on myös tietoa itsessään. Se sisältää informaatiota kulttuurisisällöstä, merkityksistä ja niissä tapahtuneista muutoksista. Enää metatiedon kautta ei vain hakeuduta varsinaisten tiedonlähteiden ääreen, vaan metatieto itsessään voi toimia primäärilähteenä, jota eri tieteenaloja edustavat datan louhijat käyttävät. Esim. historioitsijat voivat saada metatiedosta paljon irti louhimalla ja yhdistelemällä julkaisuvuotta, julkaisupaikkaa ja julkaisukieltä koskevaa dataa. Tämän päivän kulttuuriperinnön säilyttämistä on myös metatiedon tuottaminen nykyaikaisten ja tulevaisuuden tutkimusmenetelmien hyödynnettäväksi.

Avaamalla kirjastossa tuotettua metatietoa sen kirjaston vaikuttavuus yhteiskunnassa kasvaa. Toisaalta, avaaminen tuo esiin puutteet metatiedon kattavuudessa ja laadussa. Avaaminen luo paineita kirjastojen tuottaman metatiedon rikastamiseen ja laadun parantamiseen.

Avatun, linkitetyn datan aikana tieto voi muodostua useasta eri lähteestä kerätystä datasta. Linkitetyn datan muodostamissa kokonaisuuksissa on tärkeää, että siirtyminen eri lähteistä kerätyn tiedon välillä käy sujuvasti. Ei-standardienmukainen metatieto voi avattunakin olla pahimmillaan käyttökelvotonta.

Tietoa ja näin ollen myös metadataa jaetaan yhä enemmän kansainvälisesti. Tässä kokonaisuudessa kotimaisten aineistojen kuvailulla on merkitystä. Ilman laadukkaita kuvailutietoja aineistot ja metadata jää kansainvälisen tutkijayhteisön ja yleisön ulottumattomiin. Kansainvälisten standardien noudattaminen tuo metatiedon käyttöön yli rajojen.

Data on merkityksetöntä ilman laadukasta metatietoa

Ilman laadukasta metadataa moni tieto jää hakijalta tavoittamatta. Metatieto on osa tietoyhteiskunnan infrastruktuuria: ohjelmistojen ja laitetekniikan lisäksi tarvitaan metatietoa sisältöjen tavoittamiseksi. Laadukas metatieto on myös ehdoton edellytys datan käytön mahdollistajana - pelkkä bittijono ei kanna pitkälle. Erityisesti digitaalisten ihmistieteiden kehitys on viime aikoina osoittanut, että kirjastojen ja muiden muistiorganisaatioiden tuottamista viitetietokannoista on muodostumassa keskeisiä primaarilähteitä tälle tieteenalalle. Hyvin laadittujen viitetietokantojen pohjalta, joissa metatieto on rikasta ja tarpeeksi erottelukykyistä, on jo nyt pystytty tekemään merkittäviä löydöksi esimerkiksi suomalaisesta kirjallisesta tuotannosta historian eri aikakausina.

Tämän hetken tilanne näyttää ristiriitaiselta: tutkijat ja dokumenttien käyttäjät tarvitsevat entistä rikkaampaa, analyyttisempaa ja standarinmukaista metatietoa ja kuvailutyötä, jotta he pystyisivät hallitsemaan suuria datamassoja ja jotta yksittäiset dokumentit erottuisivat suurista joukoista ja löytäisivät käyttäjänsä. Samalla kirjastoissa on yleistynyt kuvailun automatisointiajattelu ja pyrkimys kuvailun tason madaltamiseen.

Näyttäisi kuitenkin siltä, että muistiorganisaatioiden tulevaisuus on laadukkaan ja rikkaan metatiedon tuottamisessa. Tämä avaa uudenlaisia mahdollisuuksia kaikenlaiselle tutkimukselle ja helpottaa ihmisia tiedonhankinnassa. Perinteiset menetelmät eivät kuitenkaan enää riitä vaan uusia tapoja hyödyntää tietotekniikkaa ja automaatiota tulee kehittää. Alamme tulisi myös olla metatietotyön erityisasiantuntijoita ja tarjota osaamista kaikille toimijoille. Tämä vaatii alan koulutuksen ja tutkimuksen kehittämistä Suomessakin. Muuten muut ottavat asiantuntijuuden ja kehittäjän roolin metatietotyössä, ja kirjastot ovat vaarassa menettää se erityisasema ja osaaminen, joka niillä on pitkässä historiassa kehittynyt.

Kuvailustandardiryhmä:

Kaisa Hypen, Turun kaupunginkirjasto

Ulla Ikäheimo, Kansalliskirjasto

Mari Isokoski, Haaga-Helia ammattikorkeakoulu

Maria Kovero, Helsingin yliopiston kirjasto

Tarja Mäkinen, Kansalliskirjasto

Juha Piukkula, Kansallinen audiovisuaalinen instituutti KAVI

Marjut Puominen, Helsingin kaupunginkirjasto

Tatja Pusa, Kansalliskirjasto, ryhmän sihteeri

Jarmo Saarti, Itä-Suomen yliopiston kirjasto

Marja-Liisa Seppälä, Kansalliskirjasto, ryhmän puheenjohtaja

Ryhmän sähköpostilista: kuvailustandardit-posti ät helsinki.fi