

Metatietotyön uudet tuulet

– dokumenttien kuvailun merkitys digiyhteiskunnassa

Kaisa Hypén, Juha Piukkula ja Jarmo Saarti

Metatiedolla tarkoitetaan tietoa tiedosta, kirjastoympäristössä konkreettisesti erilaisiin julkaisuihin liitettävää kuvailutietoa. Kuvailutiedolla on kaksi tarkoitusta: ensinnäkin sen avulla voidaan luoda tietokantoja ja kertoa julkaisuista ja niiden historiasta. Toinen tarkoitus liittyy siihen tavoitteeseen, jossa lukija tai käyttäjä ja julkaisu autetaan löytämään toisensa. Metatietotyö on murroksessa. Tämä johtuu digitaalisen toimintaympäristön ja tietotekniikan kehittymisestä, julkaisujen määrän valtavasta kasvusta ja tieteellisessä ympäristössä uusien tiedonlouhintaan liittyvien tutkimusmenetelmien kehittymisestä. Näissä kaikissa tarvitaan hyvää ja laadukasta metatietoa.

Miksi ja millaista metatietoa?

Julkaisuista kertova ja digitaalisiin dokumentteihin nykyisin jo suoraan liitettävä metadata voi olla kuvailevaa, hallinnollista tai rakenteellista. Kuvailevalla metadataalla kerrotaan sekä aineiston sisällöstä ja luonteesta että kuvataan aineiston

fyysisiä ominaisuuksia. Sen pohjalta tiedonhakuja voi päätellä, vastaako kyseinen aineisto hänen tarpeitaan. Erityisesti sähköisten aineistojen saatavuutta, käyttöoikeuksia ja teknisiä käyttöehtoja määritellään hallinnollisen metadatan avulla. Rakenteellinen metadata liittyy aineiston rakenteeseen ja esimerkiksi osien järjestykseen. Tässä artikkelissa keskitytään kuvailevaan metadataan.

Maailmassa julkaistaan dokumentteja enemmän kuin koskaan. Omakustanteet mukaan luetuna vuosittain julkaistaan noin miljoona kirjaa. Lisäksi tulevat miljoonat lehtiartikkelit ja miljardit kuva- ja muut tiedostot, joita ihmiset jakavat. Kaiken julkaisemisen tarkoituksena on tulla kuulluksi. Muistiorganisaatiot – kirjastot, museot ja arkistot – ovat kehittäneet menetelmiä niin, että kukin käyttäjä löytäisi arjessa, työssä tai opinnoissa tarvitsemansa sisällön.

Muistiorganisaatiot ovat luoneet sekä käytänteitä että tietojärjestelmiä tiedonhaun tueksi. Merkittävä osa tästä työstä perustuu ihmisen ky-

Kuvailustandardiryhmä keskittyy kuvailua ohjaavien kansainvälisten kirjastoalaan liittyvien standardien jalkauttamiseen Suomeen. Tämä tehdään yhteistyössä muiden kuvailun asiantuntijaryhmien sekä Kansalliskirjaston Kuvailusääntöpalvelun kanssa.

Ryhmä vastaa standardeja koskeviin kansainvälisiin kommenttipyyntöihin. Se myös välittää muistiorganisaatioista tulevat standardeihin liittyvät kehittämissuositukset eteenpäin. Ryhmä jakaa kuvailustandardiosaamista koulutusten ja tiedottamisen avulla kotimaisille muistiorganisaatioille.

Kehittyviä standardeja ovat erityisesti linkittyvään metadataan liittyvät, mm. RDA, ISBD, FRBR-käsitelmä johdannaisineen ja Kuvailun kansainväliset periaatteet. Kokoonpanossa ovat edustettuina eri kirjastosektorit ja muut muistiorganisaatiot. Toimintaa ohjaa ja avustaa Kansalliskirjasto.

Kuvailuryhmän tarkoitus ja tehtävät

kyyn analysoida ja kuvailla aineistoja niitä tarvitseville. Näiden organisaatioiden perustehtävänä on saattaa erilaiset julkaisut ja niiden käyttäjät kohtaamaan, ja ottaa samalla huomioon erilaisen käyttäjien ja aineistojen ominaispiirteet. Aineistojen järjestämistyö, kuten niiden kuvailu käyttäjille, perustuu kansainvälisessä yhteistyössä luotuihin standardeihin, jotka edesauttavat tietojen välitystä.

Tiedonhaku ja sen tarvitsema metatieto digitaalisessa ekosysteemissä

Digitaalisessa ympäristössä suuria dokumenttimääriä ei voi enää hallita perinteisellä tavalla. Monet tiedonhakijat luottavat automaattisten hakukoneiden tarjoamiin listauksiin ja ajatukseen tekoälystä, joka korvaa ihmisen tekemän tulkinta-, kuvailu- ja ajattelutyön. Verkojulkaisemisen alussa ajatuksena oli myös, että käyttäjät hoitaisivat kuvailun ja haun itse.

Automaatiikkaa hyödyntävät järjestelmät eivät ainakaan toistaiseksi ole riittävän kehittyneitä ja siksi verkkoaineiston tiedonhaun ongelmat alkavatkin jo näkyä arjessa: hakutulosten sisältöjen paikkansapitävyydestä tai laadusta ei ole taetta, tiedonhakua manipuloidaan kaupallisilla tai poliittisilla tavoitteilla ja suurinta osaa julkaisuista ei edes huomata. On arvioitu, että puolta tieteellisistä artikkeleista ei lue kuin tekstin kirjoittaja itse ja lehden toimittaja. Käyttäjää kiinnostavat dokumentit ja niiden käyttäminen, ei niiden kuvailu.

Suomessa on poliittisessa puheessa panostettu vahvasti tietoyhteiskunnan infrastruktuurin rakentamiseen. Tämä työ on ikävä kyllä pitkälti ohjelmistojen ja laitetekniikan kehittämistä. Ihmiset hakevat tietoyhteiskunnassakin sisältöjä: tutkittua tietoa, merkityksiä ja elämyksiä. Sisältöihin liittyvä kuvailu- ja tallennustyö tulee ymmärtää vähintään yhtä tärkeäksi kuin fyysisen infrastruktuurin rakentaminen, ja sille on turvattava vastaavat resurssit.

Myös ohjelmistokehitys edellyttää jatkuvaa inhimillistä ajattelutyötä: automaattisten sisällön-

kuvailu- ja tekoälysovellusten kehittäminen vaatii järjestelmien opettamista, inhimillisen päätelyn ja tulkinnan mallintamista. Vasta kehitystyön myötä laajojen verkkoaineistojen automatisoitu kuvailu ja tiedonhaku voidaan tehdä luottavammaksi ja tehokkaammaksi.

Lyhyt vai pitkä muisti?

Kirjastojen, museoiden ja arkistojen toiminta taakaa kansakunnan muistin säilymisen ja antaa välineitä kansakunnan itseymmärrykselle. Tässä aikajänteessä ovat sadat ja tuhannet vuodet. Säilytetyillä aineistoilla ei ole merkitystä, jos niitä ei käytetä tai jos niistä ei ole saatavana kuvailutietoa siellä, missä ihmiset etsivät itselleen aineistoja.

Metatieto säilyttää ja välittää tietoa dokumenttien sisällöstä, myös niiden aineistojen osalta, jotka eivät ole aktiivisessa käytössä. Uudet tekniset ratkaisut mahdollistavat tiedon louhinnan myös standardinmukaisesta metatiedosta. Niinpä metatieto jo itsessään kertoo kulttuurisällöistä, merkityksistä ja niissä tapahtuneista muutoksista.

Metatiedon tuottamiseen liittyvä pitkäjänteinen kansainvälinen yhteistyö mahdollistaa kaikille käyttäjille samanlaisen tiedonkäytön ympäristön ja vähentää päällekkäistä työtä. Näin saamme koottua yhdenmukaisen, laadukkaan metatiedon tiedonhakijoiden tarpeita varten kaikista aineistoista.

Suomalaisen kulttuurin säilymisen ovat taanneet julkisesti rahoitetut muistiorganisaatiot. Niiden keskeinen tehtävä on varmistaa nyt sellaisen metatiedon tuottaminen, että-kulttuuria ja sen tuottamia aineistoja käytetään mahdollisimman kattavasti. Samalla suomalainen kulttuuri säilyy elinvoimaisena ja luo uutta suomalaisuutta.

Digitalisaation eturintamassa

Muistiorganisaatiot ovat olleet digitalisaation eturintamassa. Ne ovat ottaneet ensimmäiset sähköiset tietojärjestelmät käyttöön jo 1980-luvun lopulla ja uudistaneet niiden myötä tiedontallennukseen liittyviä työprosessejaan. Kunkin sektorin – arkistojen, museoiden ja kirjastojen

– aineistoilla on omia erityispiirteitään, jotka on otettava huomioon tiedontallennuksessa ja -organisoinnissa, ja jotka ovat määritelleet tietojärjestelmien kehittämistä.

Yhteisiäkin kuvailun kohteita näillä organisaatioilla on. Ne kaikki kuvailevat henkilöitä ja paikkoja – omista erityislähtökohdistaan arkistodokumentteihin, museoesineisiin ja kirjastojen aineistoihin. Yhtenäistämällä näiden elementtien kuvailua saadaan entistä eheämpi kuva esimerkiksi tiettyyn paikkaan liittyvästä esineistöstä, sitä kuvaavista arkistomateriaaleista ja siihen liittyvästä aineistoista kirjastoissa. Myös historiallisten henkilöiden toiminnasta saadaan monipuolisempi kuva yhdistelemällä dataa useista eri lähteistä. Tätä paikka- ja toimijatiedon yhtenäistämistä tehdään Kansalliskirjaston Finto-projektissa.

Näiden organisaatioiden verkkopalveluilla on myös jo pitkä ja monipolvinen historia. Käyttöliittymiä on muokattu ja parannettu aina sen mukaan miten kehittyvä teknologia on mahdollistanut. Käyttäjän kannalta ongelmana on se, että palvelut ovat hajautuneet ja sirpaloituneet useiden eri toimijoiden omiin verkkopalveluihin ja -sivustoihin, ja käyttäjän on pitänyt miettiä tiedonhakua organisaatiolähtöisesti: minkä toimijan tietokannasta mitään tietoa on järkevää hakea.

Eri toimijoiden palveluita on koottu laajemmiksi kokonaisuuksiksi, mikä osaltaan helpottaa tiedonhakijaa orientoitumaan:

Arkistot: <https://www.arkisto.fi/fi/aineistot/verkkopalvelut-ja-tietokannat>

Museot: <http://www.nba.fi/fi/tietopalvelut/tietojarjestelmat>

Kirjastot: <https://www.kirjastot.fi/kirjastot>

Sähköiset tietokannat ja hakuliittymät ovat edellytys sille, että muistiorganisaatioiden palvelut näkyvät ja ovat käytettävissä verkossa ajasta ja paikasta riippumatta. Laadukas metadata voi tämän lisäksi auttaa palveluiden käyttäjää navigoimaan myös fyysisessä tilassa. Museossa tai kirjastossa voi kävijää vastassa tulevaisuudessa olla pirteä Pepper-robotti, joka opastaa haluttujen ai-

neistojen luo tai antaa vinkkejä siitä, mitä kaikkea mielenkiintoista kyseisestä paikasta löytyykään. Se edellyttää, että Pepperillä on käytettävissään kokoelmiin liittyvä metadata ja ajantasainen tieto aineistojen sijainneista sekä ohjelmistot, jotka osaavat yhdistellä näitä tietoja ihmisen kannalta mielekkäiksi kokonaisuuksiksi.

Luettelokortista osaksi globaalia verkostoa

Kirjastojen kuvailutyötä on ohjattu yhtenäisillä kuvailusäännöillä. Kun kuvailutieto oli yksittäisillä pahvikorteilla yksittäisten kirjastojen kortistoissa, ei ollut väliä, jos ohjeita ei aina noudatettu tai jos kuvaillessa tuli virheitä. Kun tiedot kortistoista siirrettiin sähköisiin tietokantoihin ja kun näitä tietokantoja yhdistettiin suuremmiksi kokonaisuuksiksi, tilanne muuttui. Metadata on tuotettava standardin- ja formaatinmukaisesti, jotta se on vaihtokelpoista ja yhteensopivaa muualla tuotetun metadatan kanssa.

2010-luvulla käyttöön otettu, funktionaalisen luetteloinnin periaatteita soveltava RDA-kuvailusäännöstö (Resource Description and Access) mahdollistaa periaatteessa sen, että kukin teos kuvaillaan maailmassa vain kerran, siellä, missä se julkaistaan ensimmäisen kerran. Teokseen liittyvät uudet versiot, kuten käännökset, kuvaillaan puolestaan siellä, missä ne on julkaistu ja linkitetään alkuperäiseen teokseen, samoin eri laitokset, kuten pokkarit, ja niihin liittyvät kappaleet.

Tämä edellyttää tietysti yhteistä tahtotilaa ja visiota, jota kohti edetä. Se edellyttää valtavasti järjestelmäyhteistyötä ja MARCin korvaavaa formaattia (Bibframe). Lopputuloksena voisi olla globaali tietoverkko, jossa eri tietoelementit linkittyvät toisiinsa ja jossa kerran tehty työ on muiden hyödynnettävissä jatkuvasti. Keskeinen idea on linkittäminen: ei tiedon siirto tai kopiointi järjestelmästä toiseen, vaan yhdessä tietojärjestelmässä olevan tietoelementin hyödyntäminen linkittämällä ja rikastaminen uusilla tietoelementeillä muissa järjestelmissä.

Standardit varmistavat datan laadun

Ainakin toistaiseksi eletään erillisten tietojärjestelmien aikaa ja muistiaineistoja koskevaa metadataa tarvitsee väistämättä konvertoida tietojärjestelmien vanhentuuessa. Uusiin järjestelmiin siirtyminen tulee suunnitella huolella, jottei jo tehtyä työtä menetettäisi. Datan laadun ylläpito ja varmistaminen muutosvaiheessa edellyttää mitattavaa testausta.

Valitettavasti Suomessakin on esimerkkejä siitä, että konversioissa ei ole aina huomattu selvittää kaikkia datakonversion yksityiskohtia riittävän perusteellisesti ja huolellisesti. Dataa on joko kadonnut tai se on korruptoitunut ja saattanut muuttua osittain epäluotettavaksi.

Eri organisaatioiden tietokantoja on viime vuosikymmeninä yhdistelty laajemmiksi yhteisluetteloksi. Näitä luotaessa vanhat synnit ovat usein paljastuneet. On saatettu oikoa kuvailusääntöjen noudattamisessa ja luoda omia käytäntöjä, jotka sitten yhteislueteloissa heikentävät metadataan laatua. Kirjastojärjestelmät ovat myös omalta osaltaan ohjanneet tiedon tallentamista ja vaikuttaneet siihen, ettei standardien mukaista metadataa ole aina voitu tuottaa.

Standardien vajavainen tai olematon noudattaminen vaarantavat metadataan käytettävyyden. Esimerkiksi Finna-järjestelmä on pahimmillaan hyvinkin sotkuisen ja sekavan näköinen. Tietokannan ulkoasu on suora seuraus joissakin muistiorganisaatioissa harjoitetusta omintakeisesta kuvailupolitiikasta tai jopa suoranaisestä huolimattomuudesta.

Organisaatioiden henkilökuntaa ei välttämättä ole aina koulutettu tarpeeksi luettelointiin, sisällönkuvailutyöhön ja standardien huolelliseen noudattamiseen. Tietokantoihin on tällöin syntynyt virheellisiä tai vajavaisia tietueita, jotka osaltaan heikentävät datan laatua. Ne saattavat jopa estää tiedonhaussa halutun aineiston kattavan löytymisen.

Sisällönkuvailuosaamisen taso on laskenut alan koulutuksen muutosten myötä. Peruskoulutuk-

ssa aihepiiriin ei aina kiinnitetä riittävästi huomiota. Täydennyskoulutus on ajettu alas vuosittu- hannon alkupuoleen verrattuna. Tampereen yliopisto ei enää järjestä alan täydennyskoulutusta lainkaan. Osaamisen ylläpito on Kansalliskirjaston järjestämien koulutuspäivien varassa.

Kuvailustandardien osaamaton, luova (lue: oikova) soveltaminen aiheuttaa pahimmillaan mitattavia lisäkustannuksia. Yhteisluetteloa luotaessa syntyy väistämättä jossain määrin lisätyötä ja metadataan yhtenäistämisen ja siivoamisen tarvetta, mutta leväperäinen kuvailukäytäntö lisää tätä työtä entisestään. Kaikki virheelliset tai vajavaiset tietueet eivät myöskään välttämättä löydy. Niitä saattaa putkاهدella esiin vuosien saatossa tai ne katoavat yhteisluetteloiden uumeniin.

Metatietotyö ja digitaalinen tutkimus

Tietoteknisten välineiden käyttöönotto ja niiden kehittyminen on tarkoittanut myös kuvailutyön työprosessien keskittämistä ensin kirjastojen sisällä, sitten kirjastojen välillä ja kimpoissa, vähitellen myös valtakunnallisesti ja asteittain jopa globaalisti. Parhaillaan on käynnissä murros, jossa olemme asteittain siirtymässä massiivisten datamäärien aikakauteen. Tutkimuksessa tämä näkyy siinä, että kaikki tieteenalat ovat ottaneet käyttöön datan louhintaan liittyviä työkaluja ja menetelmiä.

Muistiorganisaatioiden tuottamat tietokannat ja niiden sisältämä laadukas metatieto ovatkin muodostumassa pikkuhiljaa primaarilähteiksi, kun ennen niitä jopa vähäteltiin vain polkuina alkuperäisen tiedon lähteille. Tutkijat luottavat muistiorganisaatioiden tuottamaan metatietoon ja tämän luottamuksen säilyttäminen on jatkossa entistä tärkeämpää. Vähättelevä puhe metatietotyöstä onkin menneen maailman retoriikkaa: kirjastojen tulee muistaa jatkossa tämän perusteltävänsä ja sen arvo sekä ihmiskunnan kulttuurille että laadukkaalle tutkimustyölle.

Tietotekniikka muuttaa työnkulkua, tekoäly korvaa ainakin osan ihmistyöstä ja tehokkuuden vaatimukset vaativat uudenlaista työnjakoa. Tek-

niikka antaa vihdoinkin mahdollisuuden tavoitteeseen, jossa muistiorganisaatiot globaalina yhteisönä kuvailevat kunkin dokumentin vain keran ja luovat tähän liittyviä standardeja ja käytänteitä ja myös ottavat ne käyttöön johdonmukaisesti ja laadukkaasti. Tällä tavoin myös ihmiskunnan kulttuurinen muisti säilyy uutta luovana osana digitaalista maailmaa.

Tietoa kirjoittajista:

*Kaisa Hypén, palvelupäällikkö,
Turun kaupunginkirjasto
kaisa.hypen@utu.fi*

*Juha Piukkula, erikoissuunnittelija
Kansallinen audiovisuaalinen instituutti
juha.piukkula@kavi.fi*

*Jarmo Saarti
kirjaston johtaja, Itä-Suomen yliopisto
jarmo.saarti@uef.fi*

*Kirjoittajat ovat Kansalliskirjaston Kuvailu-
standardiryhmän jäseniä*