

Metatietoskeeman valinta – perusteluja

Sisällys

1. Käsitteiden määrittelyä.....	2
2. Miten valita sopiva metadataskeema?	6
Johdanto	6
2.2 Metatiedon funktiot ja muut käyttötarpeet metatietoskeeman valinnan ohjaajina	8
2.3 Standardit, resurssit ja ylläpito metatietoskeeman valinnan perustana	10
2.4 Sisällönhallinta- ja sovellusympäristöt skeeman valinnan näkökulmasta	11
2.5 Käyttöympäristöt ja yhteensopivuus metatietoskeeman valinnassa	12
2.6 Havaittu toimivuus, maine ja suosio	14
3. Miten skeema otetaan käyttöön	15
4. Esimerkkejä erilaisista metatietoskeemoista.....	18
4.1 MARC metatietoskeemana	18
4.2 RDF metatietoskeemana.....	18
4.3 Esimerkkejä muista metatietoskeemoista	20
Lähteet	23

1. Käsitteiden määrittelyä

Metatieto

Metatieto on aineiston kontekstia, sisältöä ja rakennetta sekä niiden hallintaa ja käsittelyä koko elinkaaren ajan kuvaavaa tietoa. Metatietoa voidaan käyttää muun muassa aineiston hakuun, paikallistamiseen, pitkäaikaissäilyttämiseen ja tunnistamiseen (Liu 2008, 8–9 ja 54–56; Foulonneau & Riley 2008, 6–7; Baca toim. 2008, 7–17)

Rakenteellinen metatieto

Metatieto, joka kuvaa teosten tai muiden kuvailtavien aineistojen tai niiden osien tyyppiä, versioita, suhteita tai muita ominaisuuksia. (Baca, M. 2003)

Käyttömetatieto

Metatieto, joka kuvaa teosten tai muiden, fyysisten tai digitaalisten, aineistojen käyttöä, esim. lainaus-tilastot. (Baca, M. 2003)

Metatietokehys (Metadata Framework)

Metatietokehystä voidaan pitää eräänlaisena luurankona, johon erilaiset aineistot kiinnittyvät erilaisilla ratkaisuilla. Kasvava tarve tallentaa digitaalisia aineistoja lisää metatietokehysten tarvetta. Metatietokehysten rakentamiseen on kaksi vaihtoehtoa: 1) kehysten muodostaminen ennen skeemojen ja sovellusten kehittämistä tai 2) niihin perustuen. (Mai Chan, L. & Lei Zeng, M. 2006)

Metatietoskeema

Metatietoskeema on yhtenäinen sääntö rakenne, joka on kehitetty aineistojen dokumentaatioon ja funktionaalisiin toimintoihin (Greenberg 2003, 1878) eli se on eräänlainen standardi, joka kuvaa metatietoa, jolla on tietty tarkoitus ja tietty käyttäjäyhteisö, mutta se ei kuvaa välttämättä metatiedon esitysmuotoa. (Mascaro, M.J. 2004.) Englanninkielisessä kirjallisuudessa käytetään vaihtelevasti termejä "schema", "scheme", ja "element set", kun viitataan metatietoskeemaan. Käytännössä usein "schema" viittaa kokonaisuuteen mukaan luettuina semanttiset ja sisältökomponentit, joihin myös "element set" viittaa. (Zhao, J. 2015, luku 8)

Metatietoskeema on myös ontologia eli käsitelmä, jossa on kuvattu eksplisiittisesti tietty käsitteistö ja näiden käsitteiden väliset suhteet. Käsitelmän perusteella tehdään metatietoskeema (metadata schema), joka muutetaan konekieliseksi. Metatietoskeeman perusteella tietokantaan syötettävät tiedot ovat metatietoa (metadata). Tietojen syöttämisessä on hyvä käyttää yhteisiä sääntöjä eli luettelointisääntöjä (cataloguing rules). Syötettävät metatiedot tulevat osittain auktorisoiduista lähteistä (authority source) eli esim. sovitusta käsitteistöistä, kontrolloiduista asiansastoista tai auktoriteettitietueesta. Auktorisoidulla lähteellä tulee olla yksilöivä tunniste (identifier) eli useimmiten käytetään URI-tunnistetta. (Viljanen, K. 2008.)

Skeema määrittelee metatietojen sisällön ja muodon ja sen, kuinka metatieto pitäisi organisoida. Tässä merkityksessä skeema voi olla esimerkiksi ydinmetatiedon ominaisuuksien luettelo, jonka avulla lähde teos voidaan tarkasti ja yhtenevästi tunnistaa ja hakea viiteitä varten, sekä luettelon käyttöohjeet. Lähde teos voi olla lähes minkäläinen tahansa, mutta se on yleensä tietue. ISO23081-standardin mukaan metatietoskeema on looginen suunnitelma, joka kuvaa metatietoelementtien väliset suhteet tavallisesti määrittelemällä säännöt metatiedon käytölle ja hallinnalle eli semantiikan, syntaksin ja arvojen velvoittavuustason (obligation level of values) (NISO).

Metatietoskeemalla on viisi avainkomponenttia:

- Skeema eli ne tiedot kuvaavat luokat, jotka tallennetaan tietueeseen.
- Sanasto, jonka avulla tieto muotoillaan luokkiin.
- Käsitemalli, joka kuvaa sitä, miten käsitteet ja syötettävä tieto liittyvät toisiinsa.
- Sisältöstandardi, joka kuvaa sitä, miten yksittäinen tieto syötetään metatietoelementteihin.
- Kirjaamisohjeet, jotka määrittelevät tiedon esitysmuotoa.

(Park, J. & Tosaka, Y. 2010)

Skeeman hierakkinen rakenne muodostuu taulukoista (tables) ja niiden sarakkeista (columns) tai kentistä (fields) sekä näytöistä (views). Syntaksi viittaa niihin sääntöihin, jotka määrittävät metatietoelementtien kenttiä. Kentät voidaan tuottaa monilla eri ohjelmointikielillä, esim. metatietoskeema Dublin Core voidaan tuottaa pelkkänä tekstinä, HTML-, XML- tai RDF-muodossa (Open metadata handbook 2014).

Metatietoskeema ohjaa metatietojen luomista, käsittelyä ja tulkintaa. Metatietoskeemojen avulla voidaan yhdenmukaistaa eri toimijoiden tapoja tuottaa, käsitellä ja hyödyntää sisältöyksiköitä kuvailevia metatietoja. Metatietoskeeman avulla voidaan mm. jakaa tietämystä metatiedon rakenteesta ja parantaa metatietojen käytettävyyttä. (Mäkinen, V.-P. 2006.) Metatietoskeema määrittelee, mitkä elementeistä ovat pakollisia, mitkä niistä toistuvat, mitä osallistumisrajoitteita (cardinalities) elementeillä on sekä kuinka tiedot tulee syöttää ja missä muodossa tiedot syötetään tietokantaan (Library UC San Diego 2010).

Metatietoskeemat eroavat toisistaan tiedon rakenteisuuden asteessa (granularity vs. specificity of description metadata) ja tiedon esitysmuodoissa. Skeemat eroavat toisistaan myös eri tiedon kuvailussa silloin, kun aineiston formaatit vaihtelevat (esim. jotkut skeemat sopivat vain sähköisessä muodossa olevan metatiedon kuvailuun, toisissa tiedon formaatilla ei ole väliä). (Beall, J. 2007.) Metatietoskeemat voidaan erotella toisistaan myös a) koon ja tarkoituksen mukaan eli sen mukaan onko skeema kattava vai ydinskeema, onko sen avulla tarkoitus kuvailla sisältöä, hallita sitä vai säilyttää sitä, koskeeko se yksittäisiä kappaleita vai kokoelmaa; b) sisältötyypin mukaan esim. kirja, äänite, video, kuva, paikka jne.; tai c) organisaatiotyyppin mukaan, esim. koulutus-, kirjasto- tai museo-organisaatio. (Park, J. & Tosaka, Y. 2010)

Encoding scheme eli kirjaamisohjeet

Kirjaamisohjeet ovat ylläpidettävä lista niistä arvioista tai syntaksinmukaisista tekstisyötteistä, jotka tarvitaan metatiedon koneelliseen käsittelyyn sisältäen ohjeet ja formaatit, joilla esitetään tietoa, kuten päivämääriä tms. (NISO)

Application profile eli Sovellusohje

Sovellusohjeet antavat mahdollisuuden skeemaa soveltavalle organisaatiolle esittää, kuinka se käyttää skeemaa, sillä sovellus on usein sananmukaisesti sovellus skeemasta. Sovellusohje (application profile) ohjaa metatietoelementtien käyttöä skeeman mukaan. Skeema määrittelee semantiikkaa ja elementtien merkityksiä, mutta sovellusohje määrittelee sääntöjä elementtien käytölle ja elementtien käytön ehtoja (obligations and constraints) sekä tarjoaa esimerkkejä ja kommentteja, joiden avulla niitä voi ymmärtää. Sovellusohjeessa voidaan yhdistää elementtejä eri elementtisarjoista ja täyttää siten funktionaaliset vaatimukset. Sovellusohjeeseen kirjataan toimijalle, tietotyyppille tai jollekin prosessille määritellyt skeeman käyttöohjeet, elementtisarjat, alielementit, skeemoista valitut kirjaamisohjeet (encoding schemes) sekä ohjeita toteuttamiseen. (NISO)

Metadata Framework eli metatietokehys

Metatietokehystä voidaan pitää eräänlaisena luurankona, johon erilaiset aineistot kiinnittyvät erilaisilla ratkaisuilla. Metatietokehysten rakentamiseen on kaksi vaihtoehtoa: 1) kehyksen muodostaminen ennen skeemojen ja sovellusten kehittämistä tai 2) niihin perustuen. (Mai Chan, L. & Lei Zeng, M. 2006.)

(Semanttinen) yhteentoimivuus

Semanttisella yhteentoimivuudella tarkoitetaan mahdollisuutta tehdä saumattomasti tiedonhakuja erilaisiin ja eri periaattein muodostettuihin tietokantoihin ikään kuin ne olisivat yksi virtuaalinen tietovarasto. Semanttisen yhteentoimivuuden voi saavuttaa erilaisten tietokantojen välille vain harmonisaation tai metatietokartoituksen avulla eli käymällä läpi crosswalk-prosessin. (Baca, M. 2003)

Yhteentoimivuus on eri järjestelmien ja alustojen, tietorakenteiden ja käyttöliittymien mahdollisuuksia vaihtaa keskenään dataa/informaatiota ilman suurta sisällön tai toiminnallisuuksien hävikkiä (NISO, 2004) tai erityisjärjestelyjä kummassakaan järjestelmässä (CC:DA, 2000.) tai muokkaamalla dataa tai informaatiota erityisesti. (Mai Chan, L. & Lei Zeng, M. 2006.)

Crosswalk eli yhteen sovittaminen

Crosswalk tarkoittaa metatietomallien, skeemojen tai sovellusohjeiden yhteensovittamista, siten että mahdollistetaan metatiedon elementtien, semantiikan ja muutosääntöjen yhteensopivuus. (NISO.) Tarkka määrittely, joka tehdään konvertoitaessa tietoja metatietostandardista, skeemasta tai sovellusohjeesta toiseen. (IFLA Working Group on the Use of Metadata Schemas 2005).

Crosswalkit ovat yleisin menetelmä, jolla metatietoskeemojen yhteentoimivuus voidaan varmistaa. Se voidaan toteuttaa toisistaan erillisille metatietoskeemoille ja löytää skeemoista vastaavuuksia elementtien tai tulkintojen välille. Se mahdollistaa tehokkaan konversion standardista toiseen. Tällöin samaa hakua voidaan käyttää useiden erilaisten tietokantojen käytössä aivan kuin ne olisivat samassa tietokannassa (Mai Chan, L. & Lei Zeng, M. 2006.) Crosswalk-prosessi mahdollistaa myös tietokantojen konvertoinnin standardista toiseen. Crosswalk on mahdollista vain *harmonisaation* ja/tai *metatietokartoituksen* avulla siten, että saavutetaan *semanttinen yhteentoimivuus* (Baca, M. 2003)

Switching schema eli välittävä skeema

Crosswalk toimii yleensä hyvin, kun se tehdään vain kahden tai muutaman skeeman välillä, mutta useampien skeemojen välillä se voi olla ja työlästä ja vaatii paljon pohdintaa. Sen vuoksi välittävän skeeman (switching schema), uuden tai olemassa olevan, käyttö on hyväksytty tapa helpottaa crosswalkia useamman skeeman välillä. (Mai Chan, L. & Lei Zeng, M. 2006.)

Harmonisointi

Harmonisoinnilla viitataan metatietostandardien välisen konsistenssin tuottamiseen (yhteinen terminologia, menetelmät ja prosessit). Harmonisointia voidaan tehdä muokkaamalla eri standardien mukaan tuotetut metatiedot yhteensopivaan muotoon, josta ne ovat kuitenkin tarvittaessa purettavissa eri standardeiksi. (NISO.) Harmonisoinnin tuloksena syntyy vain yksi metatietosarja, joka voidaan konvertoida toiseen metatietostandardiin. Harmonisaation ansiosta metatiedon ja standardin kehittäminen, tuottaminen ja käyttöönotto helpottuvat, kun niillä on yhteinen terminologia, menetelmät ja prosessit. (IFLA Working Group on the Use of Metadata Schemas 2005.)

Metatietokarttoitus eli metadata mapping

Metatietokarttoituksessa pyritään löytämään vastaava tai lähes vastaava metatietoelementti toisesta metatietoskeemasta ja mahdollistamaan metatietojen yhteentoimivuus erilaisista standardeista riippumatta. (Baca, M. 2003)

Metatietostandardi/sisältöstandardi

Metatietostandardi sisältää periaatteet, käsitteet sekä viitekehyksen, jossa metatietoelementit määritellään (a framework for defining metadata elements). Se sisältää myös perustelut metatiedon käytölle (rationale for use of metadata), ylätasen elementit (high level element sets), yhdistämissäännöt (aggregation layers), ohjeita skeemojen rakentamiseen ja kehittämiseen (guidance on developing and constructing schemas) sekä soveltamiseen ja hallintaan (guidance on implementation and management). (NISO.) Valitsemalla jonkin sisältöstandardin voidaan vaikuttaa siihen, miten tarkasti tietoa kirjataan, millaiset ovat metatiedon tuottamisen käytännöt ja millainen käyttäjäkunta tiedolla on. (Library UC San Diego, 2010)

2. Miten valita sopiva metadataskeema?

Johdanto

Tutkijat ja muut kirjoittajat, jotka ovat tarkastelleet metadataskeemoja ja niiden valintaa ja kirjoittaneet niitä koskevia ohjeita, tutkimusraportteja ja muita tekstejä (ks. lähdeluettelo), painottavat valinnassa eri asioita riippuen näkökulmasta ja aineistoista, joista he kirjoittavat. Useimmat ovat kuitenkin hyvin yhtä mieltä siitä, että käyttäjien ei pitäisi tarvita tietoa tai ymmärrystä siitä, mitä menetelmiä kuvailussa on käytetty (esim. Mai Chan, L. & Lei Zeng, M. 2006). Todellisuudessa kuitenkin eri aineistotyyppit ja erilaiset metatietostandardit, jopa samassa tietokannassa tuovat haasteita käyttäjille ja koelmien ylläpitäjille. Kirjoittajat ovat myös lähes yksimielisiä siitä, että aina kun se on mahdollista, kannattaa valita yksi standardoitu skeema (Esim. Condee-Padunova, K. & Wilson, L. 2015; Kennedy, M. R. 2008; Beall, J. 2007; Rucore 2015.) Vain harvat ovat siltä mieltä, että jos standardoitu skeema ei riitä organisaation tarpeisiin, oman skeeman rakentaminen voi olla perusteltua (esim. Beall, J. 2007).

Yhden standardisoidun metatietoskeeman valinta on tullut trendiksi, koska se usein parantaa käyttäjien ymmärrystä ja aineistojen saavutettavuutta fyysisestä formaatista riippumatta (IFLA Working Group on the Use of Metadata Schemas 2005). Saman skeeman käyttö ei kuitenkaan takaa käyttäjien helppoa pääsyä tietokantaan, vaan tarvitaan huolellisesti suunniteltuja valikoita, joissa on asiasanoja ja luokittelujärjestelmiä. (Baca, M. 2003.) Bowenin (2008) tekemässä tutkimuksessa käykin ilmi, että kirjastot ovat siirtymässä omista skeemoista kohti standardoituja skeemoja. Monet kirjastot suunnittelevat siirtyvänsä lähitulevaisuudessa laajempaan kirjoon skeemoja ja tulevat keskittymään XML-pohjaisiin skeemoihin parantaakseen yhteentoimivuutta ja metatiedon haravointia.

Kirjastomaailmassa on myös tapahtumassa siirtymä yhdestä metatietoskeemasta useampien käyttöön käyttäjien ymmärryksen ja aineistojen saavutettavuuden parantamiseksi, riippumatta niiden fyysisestä muodosta. On otettu käyttöön esim. RDF (Resource Description Framework) ja OAI (Open Archives Initiative) -skeemoja sekä semanttisen verkon työkaluja. XML:n kehittyminen on tarjonnut yhteisen syntaksin, joka parantaa metatietoskeemojen yhteentoimivuutta. Jokaisen skeeman yksittäiset elementit voidaan kartoittaa ja ilmaista XML-kielen avulla, joka helpottaa datan vaihtoa XML-kehyyksessä ja tukee samalla käyttäjiä. Eri tietolähteiden välisen yhdenmukaisuuden vuoksi jotkut kirjastot ovat valinneet useita eri standardeja, kuten MARC + ISBD kirjoille ja fyysisille aineistoille ja Dublin Coren digitaalisille aineistoille yhden metatietoskeeman sijaan. (IFLA Working Group on the Use of Metadata Schemas 2005.) Skeemat eivät siis aina ole toisiaan poissulkevia. Joskus useampia eri menetelmiä ja malleja voidaan käyttää samanaikaisesti (Mai Chan, L. & Lei Zeng, M. 2006).

Jokaisessa yhteisössä, esimerkiksi kirjastoissa, on erilaisia käyttäjävaatimuksia ja erityisiä paikallisia tarpeita. Kaikki skeemat eivät välttämättä kohtaa näitä tarpeita. Kun käytetään kuvailun perustana tiettyä skeemaa tai skeemoja, sen toimivuutta paikallisten tarpeiden tyydyttämiseksi voidaan parantaa sovellusohjeilla (application profile), jotka selvittävät, mitä metatietoon liittyviä termejä ko. organisaatiossa, tietyissä aineistossa, sovelluksessa tai käyttäjäyhteisössä, käytetään (Mai Chan, L. & Lei Zeng, M. 2006).

Metatietoskeemat eroavat toisistaan mm. tiedon rakenteisuuden asteessa (granularity vs. specificity of description metadata) ja tiedon esitysmuodoissa. Skeemat eroavat toisistaan myös eri tiedon kuvailussa silloin, kun aineiston formaatit vaihtelevat (esim. jotkut skeemat sopivat vain sähköisessä muodossa olevan metatiedon kuvailuun, toisissa tiedon formaatilla ei ole väliä). (Beall, J. 2007.)

2.1 Metatietoskeeman valinta käyttäjätarpeiden näkökulmasta

Metatietojen käyttäjät

Metatietoskeemoista ja niiden valinnasta tutkijat ja muut kirjoittajat ovat yksimielisiä siitä, että skeeman valintaa tulisi ohjata ensisijaisesti käyttäjä- ja käyttötarpeet. Käyttäjätarpeilla viitataan siihen, millaisia ihmisiä metatietojen käyttäjät ovat, kuinka homogeenisen joukon he muodostavat jne. (Ks. esim. Kennedy, M. R. 2008; Beall, J. 2007; Rucore 2015.) Tämä edellyttää tietenkin sitä, että käyttäjät, tai ainakin heidän tapansa etsiä tietoa, tunnetaan ainakin jollakin tavalla ja voidaan tehdä johtopäätöksiä siitä, tarvitsevatko käyttäjät tarkkaa tietoa aineistoista, jolloin tarvitaan tarkempaa metatietoa-kin, vai riittääkö yleisempi, esimerkiksi aiheenmukainen tieto aineistosta (Kennedy, M. R. 2008). Jos käyttäjät ovat esimerkiksi tietyn ryhmän, organisaation, tai tieteenalan edustajia ja heitä palvelee joukko metatiedon tuottajia, jotka ovat alan tai aihepiirin asiantuntijoita, voidaan valita skeema, joka tuottaa tarkkaa metatietoa asiantuntijoille (NISO).

Metatietojen tuottajat, kokoelmien ylläpitäjät ja yhteistyö

Kirjaston asiakaskäyttäjien lisäksi metatiedon käyttäjiä ovat a) metatietojen tuottajat, b) kokoelmien ylläpitäjät, c) kokoelman omistajaorganisaatio sekä d) muut (yhteistyö)organisaatiot (IFLA Working Group on the Use of Metadata Schemas 2005). Tämän vuoksi on tärkeää ottaa huomioon metatietoskeeman valinnassa myös aineistoja ja metatietoa tuottavat ja ylläpitävät henkilöt ja tahot ja heidän työprosessinsa: mitkä elementit tärkeimpiä metatiedon tuottajille ja kokoelman ylläpitäjille ja kuinka paljon tarkkuutta ne vaativat sekä mitä työprosessit vaativat metatiedoilta (Rucore 2015). Lisäksi se, mille tasolle metatietoja sovelletaan, tieteenalan, organisaation vai pienemmän ryhmän tasolle, vaikuttaa metatiedolle asetettaviin vaatimuksiin. Lisäksi skeeman valintaan voi vaikuttaa se, ovatko metatiedon tuottajat aineiston spesifin aihepiirin asiantuntijoita ja osaavatko he aihepiiriin liittyvää erityisterminologiaa. Toisin sanoen kysymys on siitä, pystyvätkö kuvailijat käyttämään metatiedon kuvailussa aiheeseen liittyviä erityistermejä vai tapahtuuko kuvailu yleisemmällä tasolla. (Kennedy, M. R. 2008.)

Metatiedon saatavuus ja saavutettavuus

Metatiedon tason (level of community or domain specificity) valinta voi lähteä liikkeelle tietyn organisaation tai tietyn alan omista käyttötarpeista, esim. ONIX-skeema on tehty kirja-alalle ja toimii myös julkaisualustana. Yleisempään tiedonhakuun ja silloin, kun metatiedon tuottajat eivät ole harjaantuneita tehtävässään, voidaan valita yleisempi skeema, joka tuottaa vain rajattuja hakumahdollisuuksia, esimerkiksi tietyn asiasanaston pohjalta. (Beall, J. 2007.) Metatiedon rakenteisuuden aste (degree of the granularity) on siis tärkeää valita käyttäjien ja aineistojen mukaan (Zhao, J. 2015, luku 8).

Jos käyttäjiä ei tunneta, voi olla hyvä aloittaa metatiedon tuottaminen osissa, alkaen niistä elementeistä, jotka ovat tärkeimpiä tiedonhakujen kannalta ja laajentaen sitten muihin elementteihin. Metatiedon laadun kannalta on siis hyvin tärkeää ymmärtää, miten metatietojen avulla voidaan tukea ensisijaisesti tiedon saatavuutta ja saavutettavuutta. (Kennedy, M. R. 2008.) Tästä näkökulmasta tärkeitä voivat myös olla tiedot siitä, milloin käyttäjät todennäköisimmin tarvitsevat tietoa, kuinka usein aineistot päivittyvät, tai onko kokoelma avoimesti käytössä, vai rajoitetaanko siihen pääsyä, ja miten se pitäisi tuoda esiin metatiedoissa (Rucore 2015).

Metatietojen tuottamisessa pitäisi aina pyrkiä validin, yhdenmukaisen, riittävän ja siten luotettavan metatiedon tuottamiseen, jonka avulla kokoelmaa voidaan arvioida oikealla tavalla. Yhdenmukaisuudella tarkoitetaan sitä, että erityistermejä käytetään samalla tavalla kuvailemaan kaikkia aineistoon kuuluvia osia. Riittävyys puolestaan tarkoittaa sitä, että käyttäjät löytävät haluamaansa aineistoa metatietojen avulla, tämä on tärkeää erityisesti silloin, kun käyttäjäryhmiä ei määritellä etukäteen tarkasti. Riittävyys siis riippuu siis käyttäjistä, ja kokoelman kehittäjän on pystyttävä arvioimaan, mitkä metatietoelementit tukevat parhaiten tiedonhakua laajoilla käyttäjäryhmillä. (Greenberg & Robertson 2002, viitattu: Kennedy, M. R. 2008.)

2.2 Metatiedon funktiot ja muut käyttötarpeet metatietoskeeman valinnan ohjaajina

Kun metatietoskeemoista ja niiden valinnoista kirjoittaneet toteavat, että skeeman valintaan pitäisi ohjata käyttäjätarpeiden lisäksi *käyttötarpeilla*, he viittaavat yleensä aineiston tyyppiin tai ns. metatiedon funktioihin eli toiminnallisiin vaatimuksiin. (Ks. esim. Zhao, J. 2015, luku 11; Beall, J. 2007; Rucore 2015; Kelly, B. 2006; Riley, J. 2006). Toisaalta aineiston käyttötarpeet voivat viitata myös siihen, **22, 19, 15** kuinka helposti metatietojen uudelleenkäyttö onnistuu tai kuinka joustava skeema on eli miten helposti skeemaa voidaan laajentaa, muokata tai vaihtaa, jos ja kun käyttäjä- ja/tai käyttötarpeet muuttuvat. (Mascaro, M.J. 2004; Mai Chan, L. & Lei Zeng, M. 2006; Kelly, B. 2006.)

Metatiedon funktioita

Useimmat kirjoittajat kehottavat pohtimaan tarkkaan sitä, millaista kuvailtava aineisto on, missä formaateissa se on, millaisiin aihealueisiin tai tieteenaloihin tai aihepiireihin aineistot liittyvät ja mihin tarkoituksiin aineistoja hankitaan. Tämän perusteella voidaan tarkastella sitä, mitä funktioita erilaisia metatiedolla voi olla. Yleisesti ottaen, kannattaa valita skeema tai profiili, joka sopii kyseiselle aineistolle ja tarvittaviin funktioihin. (Ks. esim. Kennedy, M. R. 2008; Beall, J. 2007.)

Tärkein funktio liittyy sekä käyttäjä- että käyttötarpeisiin: Metatiedon pitäisi tukea aineiston löydettävyyttä (Riley, J. 2006). Kuten edellä on todettu, löydettävyyden tukeminen liittyy ennen kaikkea siihen, mitä käyttäjät metatiedoilta odottavat, miten he ovat tottuneet tietoa hakemaan sekä millaisia tietolähteitä he ovat tottuneet käyttämään. Löydettävyyteen liittyy lisäksi olennaisesti aineiston saavutettavuus: Helppo saavutettavuus parantaa aineiston löydettävyyttä. Yksityiskohtainen kuvailu puolestaan mahdollistaa erilaisten metatietojen, esim. sivunumeroiden tai sisällysluettelon, paremman esille tuomisen. Toisinaan samaa elementtiä käytetään sekä saavutettavuuden että kuvailun vuoksi, kuten esim. nimekettä. (Beall, J. 2007.)

Metatiedot ja tiedonhaut

Näistä syistä skeeman valinnan kannalta on tärkeää selvittää ensinnäkin, löytyvätkö tarvittavat hakujen keskeiset määreet tai elementit skeemasta (IFLA Working Group on the Use of Metadata Schemas 2005). Yhtä tärkeää kuin se, onko skeemassa kaikki tarvittavat elementit, on se, että siinä on vain tarvittavat elementit. Skeeman ylläpito vaatii sekä inhimillisiä että taloudellisia resursseja. Siten laajan metatietoskeeman ylläpito voi kuluttaa niitä aivan turhaan.

Toiseksi hakujen tukeminen voi liittyä aineistotyyppiin: Metatietoskeema määrittelee mitä tietoja kuvailtavasta kohteesta tulee kuvata (voidaan kuvata) ja eri sisältötyypeillä on yleensä hyvin erilaiset mielekkäät tiedot (vrt. esim. runo vs. ilmavalokuva vs. historiallinen tapahtuma). Hyvin yleisen tason metatieto ei riitä sisältöjen kuvailun tavoiteltuun tarkkuustasoon, joten sisällön kuvaamiseen tarvitaan erilaisia metatietoskeemoja, siksi yleisten skeemojen (esim. Dublin Core) lisäksi on kehitetty sisältötyyppikohtaisia metatietoskeemoja, kuten EAD (Encoded Archival Description) arkistomateriaaleille,

MARC kirjoille, CDWA taideteoksille jne. (NISO.) Metatietoskeemaa valittaessa on siis tärkeää selvittää onko skeemassa riittävästi elementtejä keskeisen kuvailutiedon tallentamiseksi, ovatko luokitukset relevantteja ja esittävätkö ne tietoa aineistoista sopivalla tarkkuudella (Kelly, B. 2006). Kolmanneksi hakujen tukeminen liittyy osittain hakujärjestelmiin ja ohjelmistovalintoihin ym. sovellusympäristöön liittyviin seikkoihin, joihin palataan hieman myöhemmin tässä tekstissä.

Aineiston löydettävyyden parantamisen lisäksi metatiedolla voi olla muitakin funktioita, esimerkiksi hallinnollisen, teknisen, oikeudellisen, pitkäaikaissäilytykseen tai muihin tarkoituksiin tarvittavan metatiedon tallentaminen ja käsitteleminen (Beall, J. 2007). Myös metatietojen kielelle, aineiston formaatille tai medialle voidaan asettaa erityisiä vaatimuksia, joita metatietoskeeman on tuettava (IFLA Working Group on the Use of Metadata Schemas 2005). Skeeman sopivuus metatiedon löytämiseen, oikeuksien hallintaan tms. funktioon voi olla tärkeää skeeman valinnan kannalta. Näiden seikkojen vuoksi sen määrittelemisen, mitä funktioita metatiedolle asetetaan, on hyvin tärkeää. Esimerkiksi useista tietokannoista samanaikaisesti tehtävät haut ovat tulleet suosituksi ja siksi metatietojen tuplakontrolli on tullut yhä tärkeämmäksi. (IFLA Working Group on the Use of Metadata Schemas 2005). Kaikkia näitä metatiedon funktioita on selvitettävä ennen skeeman valintaa, sillä kaikki skeemat eivät ole laajennettavissa ja siksi se voi olla skeeman valintaan vaikuttava kriteeri (Beall, J. 2007).

Metatietoskeemojen muokattavuus

Luvun alussa viitattiin siihen, että skeeman valinnassa on tärkeää huomioida se, että tieto ja käyttäjien tietotarpeet muuttuvat ajan kuluessa ja siksi on tärkeää arvioida metatietostandardia ja sen mahdollisuuksia muokkauksiin ja muutoksiin aika ajoin (Rucore 2015). Muokattavuudella viitataan yleensä skeeman sovellettavuuteen paikallisiin tarpeisiin (adaptability of the scheme to local needs) (Beall, J. 2007), laajennettavuuteen (Kelly, B. 2006), joustavuuteen tai mahdollisuuksiin vaihtaa koko skeema toiseen tarvittaessa (Mascaro, M.J. 2004). Sovellettavuus paikallisiin tarpeisiin voi viitata esimerkiksi uusien kenttien tai tagien lisäämiseen, sillä ne voivat lisätä aineiston löydettävyyttä ja kiinnostavuutta paikallisella tasolla sekä parantaa yhteentoimivuutta (Beall, J. 2007). Skeeman laajennettavuus voi tarkoittaa esimerkiksi sitä, että skeemaan lisätään muiden skeemojen elementtejä tai että sitä käytetään eri aineistotyyppien metatietojen tuottamiseen (Kelly, B. 2006).

Metatietoskeemojen joustavuus

Skeeman joustavuudesta on Mascaron tutkimuksen (2008, 33–35) mukaan onkin tullut tärkein skeeman valintaan vaikuttava tekijä. Joustavuutta pidetään ko. tutkimuksen mukaan tärkeimpänä skeeman ominaisuutena. Tutkimuksessa havaittiin, että joustavuus viittasi tällöin erityisesti muokattavuuteen ja eri formaateilla tuotetun metatiedon jakamismahdollisuuksiin. Lisäksi joustavuudella saatettiin tarkoittaa myös sopivuutta eri aineistotyypeille, muokattavuutta eri standardeihin sekä jakamista muiden kanssa. Joustavuutta korostettiin ehkä myös siksi, että eri aineistotyypeille tarkoitettujen formaattien erilaisista kuvailutarpeista oli paljon aiempia huonoja kokemuksia. Skeeman vaihtaminen metatietojen elinkaaren aikana on melko todennäköistä, sillä kaikki skeemat eivät ole muokattavissa tai laajennettavissa paikallisiin tarpeisiin. Aina tulevaa kehitystä ei pystytä ennustamaan, vaan tarve konversioon saattaa tulla yllättäenkin, esimerkiksi kun halutaan yhdistää kokoelmia tai tietokantoja saman organisaation sisällä tai organisaatioiden välillä. (Mascaro, M.J. 2004.) Siksi skeeman valintaan voi vaikuttaa se, kuinka helppoa skeeman vaihtaminen on, jos tarve konversioon ilmenee (IFLA Working Group on the Use of Metadata Schemas 2005).

Metatietoskeemat ja muutokset

Myöhemmin tapahtuviin konversioihin on syytä varautua, useiden skeemojen käyttö samalla syntaksilla on mahdollista. Valmiita crosswalkeja on saatavilla (esim. Dublin Coresta MARC21:een) ja ne helpottavat tietueiden muunnoksia skeemasta toiseen. On myös aivan tavallista, että sähköiset aineistot linkitetään metatietoihin jo olemassa olevissa standardeissa, kuten MARC:issa. Eri ympäristöissä tuotettua metatietoa voidaan siis haravoida ja käyttää toisessa ympäristössä. Tämä on erityisen hyödyllistä, jos tietoja halutaan linkittää eri (tieteen)aloilta, aihealueilta tai sovelluksista, esim. museoista, taidegallerioista, oppilaitoksista, kustantajilta tai hallinnosta. (IFLA Working Group on the Use of Metadata Schemas 2005.)

Kahden kokoelman tai tietokannan yhdistämisessä voidaan ottaa lähtökohdaksi se, että niillä voi olla erilaisia tarpeita ja lähtökohtia ja niin ollen yhteiseksi malliksi, (kirj. huom., nk. välitysskeemaksi) voidaan ottaa MARC, josta muokataan uusia ja yksinkertaisempia skeemoja. Menetelminä voidaan käyttää muokkaamista, muuntamista, laajentamista, osittaista muokkausta, kääntämistä toiselle kielelle jne. Joka tapauksessa uusi skeema on riippuvainen malliskeemasta. Tällainen lähestymistapa varmistaa yhteisen perusrakenteen ja yhteiset elementit ja samalla antaa mahdollisuuden elementtien tarkkuuden ja yksityiskohtien vaihteluille. (Zhao, J. 2015, luku 8.)

2.3 Standardit, resurssit ja ylläpito metatietoskeeman valinnan perustana

Metatietoskeemat ja standardit

Yleensä metatietoskeemaa valittaessa valitaan mieluummin standardoituja skeemoja kuin paikallisia. Standardoiduilla skeemoilla on etunsa, mutta ne voivat myös vaikeuttaa metatietojen käyttöä tai niiden käyttö voi vaatia paljon aikaa. Standardoidun skeeman valinta edellyttää myös organisaatiolta sitä, että standardiin tulevia muutoksia seurataan (Condee-Padunova, K. & Wilson, L. 2015). Jos tarvitaan enemmän kuin yksi skeema, on tärkeää varmistaa, onko skeemojen välisiä valmiita metatietokartoituksia tai crosswalk-prosesseja saatavilla yhteentoimivuuden ylläpitämiseksi ja voidaanko metatietoja vaihtaa yhden skeeman tai standardin avulla yhteistyökumppaneiden kanssa (IFLA Working Group on the Use of Metadata Schemas 2005). Metatietoskeeman valinnassa on tärkeää huomata myös skeeman ylläpitäjäorganisaation merkitys: organisaation vakiintuneisuus ja elävyys ovat ratkaisevan tärkeitä skeeman menestymisen ehtoja. Sovellusten käyttöönottajien kannattaa perehtyä skeeman taustaorganisaatioon, joka tuottaa ja ylläpitää skeemaa, kahesta syystä: varmistaakseen että se pitää skeeman ajan ja käyttäjätarpeiden tasalla sekä sen, että palautetta sovellusten käyttöönottajilta ja käyttäjiltä otetaan vastaan. (Beall, J. 2007.)

Metatietoskeemat ja resurssit

Käyttäjä ja käyttötarpeiden lisäksi skeeman valintaan, kuten muidenkin työtä ohjaavien tekijöiden valintaan voivat vaikuttaa myös hyvin käytännölliset seikat, kuten se kuinka paljon työaikaa ja rahaa uuden skeeman käyttöönottoon on käytettävissä (Kennedy, M. R. 2008). Aikaa ja rahaa ei tarvita vain skeeman käyttöönottoon, teknisiin sovelluksiin tai skeeman lisensointiin, vaan myös metatiedon tuottaminen, skeeman mahdollinen muokkaaminen ja/tai vaihtaminen, ylläpito, käyttö- ja käyttäjätuki sekä henkilöstön mahdolliset koulutustarpeet vaativat panostusta koko metatiedon elinkaaren ajan. (Kelly, B. 2006; IFLA Working Group on the Use of Metadata Schemas 2005.) Muun muassa näistä syistä on hyvä miettiä, onko sopivaa valmista standardoitua skeemaa, joka on avoin ja josta on saatavilla määrittely. Suljettuja tai suojattuja skeemoja, joilla on rajallinen jaettavuus ja aineistojen käytettävyys, kannattaa välttää.

Uuden skeeman käyttöönotto on aina henkilöstöresursseja vaativaa työskentelyä, mutta lisäksi voidaan tarvita henkilöstökoulutusta ja harjoittelua, jos ne eivät ole tuttuja henkilöstölle, sillä jotkut metatietoskeemat vaativat paljon asiantuntemusta (Kelly, B. 2006). Sen vuoksi inhimillisten, teknisten ja taloudellisten resurssien saatavuus ja määrä on hyvä varmistaa etukäteen.

Myös skeeman ylläpito ja laadunvalvonta vaativat työtä ja työaika (Kennedy, M. R. 2008). Sen vuoksi kokoelmaan mahdollisesti tulevat muutokset (esim. järjestelmän vaihto, migraatio eli metatietojen muunnos toiseen formaattiin tai integrointi toiseen tietokantaan) sekä näköpiirissä oleva kasvuodotus vaikuttavat skeeman valintaan siten, että valitaan mahdollisimman yksinkertainen skeema, jotta resursseja vaativilta muutoksilta vältytään. Näistä syistä olisi hyvä kirjata suunnitelma tai strategia miten metatietoa tuotetaan, muokataan ja ylläpidetään sen elinkaaren aikana organisaatiossa, että tarvittavat resurssit voitaisiin esimerkiksi vuosittain varata niitä varten (Kelly, B. 2006; IFLA Working Group on the Use of Metadata Schemas 2005.)

2.4 Sisällönhallinta- ja sovellusympäristöt skeeman valinnan näkökulmasta

Metatiedot ja sisällönhallinta

Metatiedot ovat oleellisessa asemassa sisällönhallintaa kehitettäessä: Metatietoskeemat parantavat yhteensopivuutta – ja täten sisällönhallintaa – eri järjestelmien välillä. Sen sijaan ne eivät yksin takaa yhteensopivuutta eri metatietoskeemojen välillä (Mäkinen, V.-P., 2006). Hajanaisessa sisällönhallintaympäristössä metatietojen käyttämistä täytyy ohjata nimenomaan metatietoskeeman avulla. Esi-tettyjä skeemoja voidaan täten teoriassa käyttää sisällönhallinnan parantamiseksi monimutkaisessa sisällönhallintaympäristössä. Eräs oleellinen metatietoskeeman ominaisuus on sen liittäminen muihin metatietoskeemoihin sovellusten välisen yhteensopivuuden varmistamiseksi. Metatietojen olemassa-olo mahdollistaa sisältöyksiköiden hallintaan ja käsittelyyn liittyvät toimet, mukaan lukien niiden organisoimisen, tuottamisen, hakemisen ja tuhoamisen. (Salminen 2005, viitattu: Mäkinen, V.-P., 2006.) Skeeman valinnan kannalta on siis tärkeää hahmottaa, missä ja miten skeemaa tullaan käyttämään eli käytetäänkö sitä aineiston kuvaamiseen dokumentinhallintajärjestelmässä, verkkosivustolla, myynti-tietojärjestelmässä, ja/tai hakutietokannassa (NISO).

Metatiedot ja hakujärjestelmät

Myös järjestelmän sallimat hakutavat ja kokoelman saavutettavuus vaikuttavat skeeman valintaan. (Kennedy, M. R. 2008). Jos kokoelma on saavutettavissa paikallisen organisaation suunnitteleman verkkokäyttöliittymän välityksellä ja hakuehtoina voidaan käyttää vain aihetta tai tekijää, valittava skeema voi olla hyvin lyhyt ja yksinkertainen. Toisin sanoen, mitä vähäisempiä ovat aineistoyksiköiden ja hakutekijöiden väliset suhteet, sitä yksinkertaisempi skeema tarvitaan. Jos sen sijaan kokoelmaan voidaan tehdä hakuja, joissa hakutermien väliset suhteet eivät ole ennalta määrättyjä, metatieto-skeema voi olla laajempi ja tarjoaa siten paljon erilaisia hakumahdollisuuksia ja mielekkäitä hakutu-loksia.

Lisäksi skeemat linkittyvät sisältöstandardeihin eri tavoin. Jotkut skeemat, kuten MARC, liittyvät hyvin läheisesti sisältöstandardeihin. MARC liitetään usein AACR:iin (Anglo-American Cataloguing rules) ja Library of Congressin asiansastoon. Dublin Core (DC) on puolestaan sisältöstandardeista paljon riip-pumattomampi, joten sen valitseminen skeemaksi johtaa siihen, että sisältöstandardi on valittava erikseen. (Beall, J. 2007.)

Metatietoskeemat ja sovellusympäristöt

Sisällönhallintaympäristön ohella tärkeää on myös ottaa huomioon se, miten metatietoja on tähän saakka tallennettu: olemassa olevat metatietoskeemat ja liittymät näihin sekä metatietoskeeman sovellusympäristö. Se, kuinka laajasti käytetty skeema on ja ovatko sen sovellukset tai käyttöympäristöt olleet samankaltaisia kuin ne, johon skeemaa nyt aiotaan soveltaa, voi vaikuttaa siihen, kuinka helposti sovellukset on otettavissa käyttöön tai kuinka helppoa niihin on löytää valmiita crosswalkeja, kieliver-sioita yms. (IFLA Working Group on the Use of Metadata Schemas 2005).

Myös hakujärjestelmien tai ohjelmistojen valinnalla voi olla merkittävä vaikutus skeeman valintaan. Nk. integroidut kirjastojärjestelmät voivat hakea MARC-muotoista metatietoa, mutta eivät välttämättä muilla standardeilla koodattua metatietoa. Sellaisia hakujärjestelmiä tai alustoja, jotka voivat hyödyntää metatietoa kokonaisuudessaan, voi olla vaikea löytää vähemmän suosituille metatietoskeemoille. Lisäksi järjestelmiä valittaessa olisi hyvä ottaa huomioon se, että metatietoa voidaan tuottaa inhimillisen tuotannon lisäksi koneellisesti. (Beall, J. 2007.) Formaatin pitäisi siis sopia yhteen laajemman teknologisen infrastruktuurin kanssa tai vaihtoehtoisesti organisaation infrastruktuuria saatetaan joutua laajentamaan (Riley, J. 2006).

2.5 Käyttöympäristöt ja yhteensopivuus metatietoskeeman valinnassa

Yhteistyö ja yhdenmukaisuus

Skeeman valintaan vaikuttaa myös suhde muihin kokoelmiin samassa organisaatiossa tai samalla alalla (Kennedy, M. R. 2008). Jos kokoelmalla on yhteyksiä muihin kokoelmiin, niiden väliset suhteet on tärkeää kuvata. Yhteistä kontrolloitua sanastoa käyttäen voidaan esimerkiksi ohjata käyttäjiä käyttämään samoja hakuelementtejä. Skeeman valinnassa ja suunnittelussa voi olla apua siitä, että saman alan tai samankaltaisen kokoelman kuvailuun valitaan samanlainen sanasto. Se voi auttaa havaitsemaan aineistoissa yhteneväisyyksiä ja säästää myös aikaa ja varmistaa yhdenmukaisuuden erityisesti silloin, kun kokoelma on osa laajempaa kokoelmaa. Myös se, haravoidaanko metatietoa muihin organisaatioihin tai muihin organisaation kokoelmiin, voi vaikuttaa skeeman valintaan. Tehokas haravointi edellyttää nimittäin metatietojen laajaa yhdenmukaisuutta ja standardin noudattamista. (Kennedy, M. R. 2008; Kelly, B. 2006). Joskus saatetaan tarvita myös (lisä)metatietoja tai kenttiä verkostoyhteistyön tukemiseksi ja/tai oman organisaation, tieteenalan tms. ryhmän integraatio- ja yhteentoimivuuskehityksen huomioon ottamista koko metatiedon elinkaaren ajan (IFLA Working Group on the Use of Metadata Schemas 2005).

Yhteistyö ja yhteentoimivuus

Yhteentoimivuudesta on tulossa yksi tärkeimmistä metadatasovellusten periaatteista. Muita tärkeitä periaatteita ovat yksinkertaisuus, modulaarisuus, uudelleenkäytettävyys ja laajennettavuus (Duval et al., 2002; Zeng et al., 2003, viitattu: Mai Chan, L. & Lei Zeng, M. 2006). Sen vuoksi ne tulisi ottaa huomioon käyttöönnotossa ja kehityksessä. Jos kokoelman tulee olla yhteentoimiva (interoperable) muiden kokoelmien kanssa, se voi vaikuttaa metatiedon organisointiin (Mai Chan, L. & Lei Zeng, M. 2006). Käytännössä yhteentoimivuus tarkoittaa sitä, että erilaisista metatietokannoista voidaan hakea samalla haulla niin, että tulokset ovat merkityksellisiä. Mascarón (2004) mukaan yhteentoimivuus on suunniteltava skeemaa valittaessa siten, että valitaan joko sama skeema tai käytetään samaa sanastoa tai standardia kaikissa kokoelmissa, joiden halutaan olevan keskenään yhteentoimivia. Yhteentoimivuuden korostaminen kertoo siitä, että metatietoa halutaan jakaa muidenkin kanssa kuin oman orga-

nisaation kanssa, sillä se auttaa tuomaan aineistoja esiin julkisesti. Mascaron tutkimuksessa skeemoihin liittyviä sääntöjä toivottiin noudatettavan tarkasti, että yhteentoimivuus toteutuisi ja metatieto olisi yhdenmukaista organisaatiosta riippumatta. (Mascaro, M.J. 2004, 32–35).

Tähän saakka kirjastoissa yhteentoimivuutta on pidetty lähinnä teknisenä ongelmana, joka vaatii teknisen ratkaisun. Strukturaalista metatietoa on yritetty muokata ja rakentaa profiloitijärjestelmiä siten, että metatietoa voitaisiin tuottaa XML-skeemalla ja mahdollisesti käyttäen validointijärjestelmiä niin, että aineistot saataisiin vastaamaan profiileihin lisättyjä ominaisuuksia. Niiden avulla voidaan kuitenkin tuottaa vain sellaista yhteentoimivuutta sen kapeimmassa merkityksessä. Ne eivät takaa yhteentoimivuutta suurissa, eri organisaatioiden välisissä, digitaalisissa hankkeissa. (McDonough, J. 2008.)

Yhteentoimivuus ja standardit

Ihanteellista olisi, että yhtenäinen standardi, esim. MARC tai DC maksimoisi yhteentoimivuuden ja yhdenmukaisuuden eri kokoelmien välillä (Mai Chan, L. & Lei Zeng, M. 2006). Baca (2003) kuitenkin huomauttaa, että jos tietokantoja ei ole tuotettu samoilla skeemoilla, voidaan niihin kuitenkin tehdä hakuja käyttäen apuna yhteisiä informaatioprotokollia kuten Z39.50 tai yhteistä kuvauskieltä (mark-up language), kuten XML:ää, sekä metadatakartoituksia (metadata mapping).

Eri skeemoja voidaan sovittaa yhteen metatietokartoitusten (metadata mappings) avulla. Tätä kutsutaan joskus myös crosswalkiksi, vaikka crosswalk on vain yksi metatietokartoitukseen tarkoitettu menetelmä. Metatietokartoituksessa luodaan näkyviä eri tarkoituksiin tuotettuihin metatietoihin. Kartoitukset tarkasta formaatista yleisempään ovat hyvin tavallisia. Sen sijaan kartoitukset yleisestä tarkempaan eivät tavallisesti toimi. (Riley, J. 2006.)

Käytännössä yhteentoimivuus (interoperability) tarkoittaa usein skeeman sopivuutta crosswalk-prosessiin tai sitä, onko skeemojen välille tehty metatietokartoituksia (mappings) ja kuinka hyvin ne ovat saatavilla (Beall, J. 2007). Yhteentoimivuus saattaa viitata myös haravoinnin (harvesting) mahdollisuuksiin. Skeema, joka on hyvin yhteentoimiva, sopii metatietojen haravointiin ja metahakuihin metatiedoista. Yhteentoimivuus on siis yleensä hyvä, jos skeema on suosittu ja laajasti käytetty.

Crosswalk-menetelmä

Crosswalk on yleisin menetelmä, jolla metatietoskeemojen yhteentoimivuus voidaan varmistaa. Se voidaan toteuttaa toisistaan erillisille metatietoskeemoille ja löytää skeemoista vastaavuuksia elementtien tai tulkintojen välille. Se mahdollistaa tehokkaan konversion standardista toiseen. Tällöin samaa hakua voidaan käyttää useiden erilaisten tietokantojen käytössä aivan kuin ne olisivat samassa tietokannassa. (Mai Chan, L. & Lei Zeng, M. 2006.)

Suurin ongelma crosswalkissa on metatietojen määrän vaihtelu eri skeemoissa: joitakin elementtejä saattaa olla toisessa skeemassa enemmän kuin toisessa, joitakin vähemmän tai ei lainkaan (Zeng and Xiao, 2001, viitattu: Mai Chan, L. & Lei Zeng, M. 2006). Kuitenkin usein vain elementtien nimet ja määritelmät otetaan huomioon crosswalkia tehtäessä eikä muuta samankaltaisuutta oteta huomioon. Silti monet elementit saattavat olla lähellä ekvivalenttia merkityksensä tai laajuutensa vuoksi. Tästä syystä konversioon perustuvat crosswalkit voivat olla ongelmallisia. (Mai Chan, L. & Lei Zeng, M. 2006.)

Crosswalk toimii yleensä hyvin, kun se tehdään vain kahden tai muutaman skeeman välillä, mutta useampien skeemojen välillä se voi olla ja ikävää ja työlästä ja vaatii paljon pohdintaa. Sen vuoksi välittävän skeeman (switching schema), uuden tai olemassa olevan, käyttö on hyväksytty tapa helpottaa crosswalkia useamman skeeman välillä. (Mai Chan, L. & Lei Zeng, M. 2006.)

2.6 Havaittu toimivuus, maine ja suosio

Viimeisimpänä muttei vähäisimpänä: Havaittu toimivuus, maine ja suosio (proven success, reputation, success and popularity) ovat usein skeemojen tärkeitä valintakriteerejä, sillä ne kertovat usein toimivista ja hyvin dokumentoiduista sovelluksista (Beall, J. 2007). Yleisesti ottaen voidaan sanoa, että kansainväliset vakiintuneet standardit ovat luotettavampia kuin itse kehitetyt tai pienten piirien skeemat (Kelly, B. 2006). Tähän on syynä se, että skeemaan vakiintuneisuus samoin kuin maine liittyvät usein esimerkiksi skeeman laatuun ja ajantasaisuuteen, useampiin eri kieliversioihin sekä dokumentaation laajuuteen ja laatuun (Beall, J. 2007).

Myös havaittu toimivuus sekä saatavilla olevat sovellustuki, käyttäjätuki, esimerkit, ohjelmistotyökäluut, skeeman kieliversiot sekä valmiit sisältöstandardit, jotka voivat helpottaa kehitystä, kustomointia ja päivitystä, ovat usein kehittyneempiä vakiintuneissa standardeissa kuin itse kehitetyissä, pienten piirien skeemoissa (Kelly, B. 2006). Monet asiantuntijat ovatkin yhtä mieltä siitä, että skeemaa valittaessa kannattaa valita tunnettu ja hyvin dokumentoitu skeema, jota käytetään yleisesti valittuun aineistotyyppiin ja joka on käytännössä havaittu toimivaksi (ks. esim. Mäkinen, V.-P., 2006). Valmiit standardit myös varmistavat, että metatiedot ovat jaettavissa standardoitujen käytäntöjen avulla.

3. Miten skeema otetaan käyttöön

Skeeman rakenteen määrittely

Skeeman rakennetta määriteltäessä on valittava joko yksi metatietoelementtisarja tai joukko niitä eli on päätettävä kuvataanko skeeman avulla organisaatiota, toimijoita, auktoriteetteja vai näiden välisiä suhteita (NISO). Eri skeemat tarjoavat erilaisia näkökulmia metatietoelementteihin, esim. kun tehdään tietuetta, voidaan kuvata vain aineistoa, tai aineistoa ja tekijää tai aineistoa, tekijää ja niiden välisiä suhteita. Yleensä ottaen kannattaa harkita kaikkia vaihtoehtoja ja valita formaatti, joka tukee sellaisia hakuja, joita käyttö vaatii (Riley, J. 2006). Näiden tietojen avulla voidaan tuottaa perusmalli, joka kuvastaa aineistojen luonnetta ja käyttäjätarpeita palvelevat aineistot. Perusmalli voi olla yksinkertainen kuvaaja, joka näyttää eri toimijoiden väliset suhteet, tuottajien, ylläpitäjien ja käyttäjien. Tämä on ensimmäinen askel kohti metatietostrategiaa. (Rucore 2015.)

Olemassa olevan metatiedon uudelleenkäyttö on tärkeää, kun vaihdetaan metatietoelementtikokoelmasta toiseen, sillä olemassa olevien termien erottaminen on edellytys skeeman kehittämisen prosessille. Skeemasta olisi siis tärkeää tunnistaa, mitä elementtejä, elementtiryhmiä ja alielementtejä valitussa skeemassa voidaan käyttää muista skeemoista tai mitä niistä voidaan linkittää alkuperäisiin skeemoihin (crosswalk) (NISO). Skeemaa arvioitaessa olisi hyvä pohtia myös sitä, mitä uusia elementtejä tai alielementtejä tarvitaan valitusta skeemasta ja mitä jätetään pois. Metadata voidaan tuottaa myös osissa alkaen niistä elementeistä, jotka ovat tärkeimpiä tiedonhakujen kannalta ja laajentaen sitten muihin elementteihin (Riley, J. 2006). Kaikki metatieto kannattaa myös validoida joko paikallista tai laajempaa skeemamäärittelyä apuna käyttäen (McDonough, J. 2008).

Tietokannan perustaminen

Metatietoskeemaa koskevan data kootaan metatieto-tietokantaan (Metadata Registry) (Mai Chan, L. & Lei Zeng, M. 2006). Metatietoja koskevan tietokannan olemassaolo edistää sen laajempaa käyttöön-ottoa, standardisointia ja yhteentoimivuutta ja uudelleenkäyttöä eri aloilla ja käytännöissä. Valittu uusi metatietoskeema voi vaatia olemassa olevan metatiedon konvertointia toiseen muotoon tai metatiedon harmonisointia (harmonization) – erityisesti silloin kun yhdistetään kaksi tai useampia skeemoja tai tietokantoja ja niiden halutaan olevan yhteentoimivia. Tällöin aloitetaan yleensä metatietokartoitus (metadata mapping) esimerkiksi Crosswalk-prosessilla. Metatietokartoituksessa pyritään löytämään vastaava tai lähes vastaava metatietoelementti toisesta metatietoskeemasta ja mahdollistamaan metatietojen yhteentoimivuus erilaisista standardeista riippumatta (Baca, M. 2003). Myös useiden skeemojen käyttö samalla syntaksilla on mahdollista.

Mahdollinen crosswalk

Crosswalk edellyttää aina jokaisesta metatietoelementin semanttista määrittelyä ja vastaavuuksien etsimistä skeemojen metatietoelementeille. Jos vastaavuutta ei löydy, crosswalk ei ole mahdollista. Jos tiedolle ei voida löytää vastaavuutta, se jätetään pois. Näin käy usein, kun yksityiskohtaisesta konvertoidaan yleisempään skeemaan. Ongelman kiertämiseksi voidaan tehdä suhteellinen crosswalk (relative crosswalk) yhdistämällä kaikki elementit johonkin toiseen elementtiin riippumatta semanttisesta yhdenmukaisuudesta. Lisäksi crosswalk-prosessissa tarvitaan tietoa siitä, onko elementti pakollinen vai ei tai onko pakollisuudelle jokin ehto ja voiko metatietoelementti toistua samassa tietueessa. (Mai Chan, L. & Lei Zeng, M. 2006.) Lisäksi tarvitaan tietoa niistä skeeman ominaisuuksista, kuten ontologioiden hierarkiat, tai elementtien sisältöihin liittyvistä ominaisuuksista, esim. vapaateksti, pvm tai kontrolloitu sanasto, jotka voivat vaikeuttaa metatietokartoitusta.

Crosswalk toimii yleensä hyvin, kun se tehdään vain kahden tai muutaman skeeman välillä, mutta useampien skeemojen välillä se voi olla työlästä ja vaatia paljon pohdintaa. Sen vuoksi välittävän skeeman (switching schema), uuden tai olemassa olevan, käyttö on hyväksytty tapa helpottaa crosswalkia useamman skeeman välillä. Myös valmiita crosswalkeja on saatavilla (esim. Dublin Coresta MARC21:een) ja ne voivat helpottaa tietueiden muunnoksia skeemasta toiseen. (Mai Chan, L. & Lei Zeng, M. 2006.)

Metatietoskeemasta sisältöstandardiin

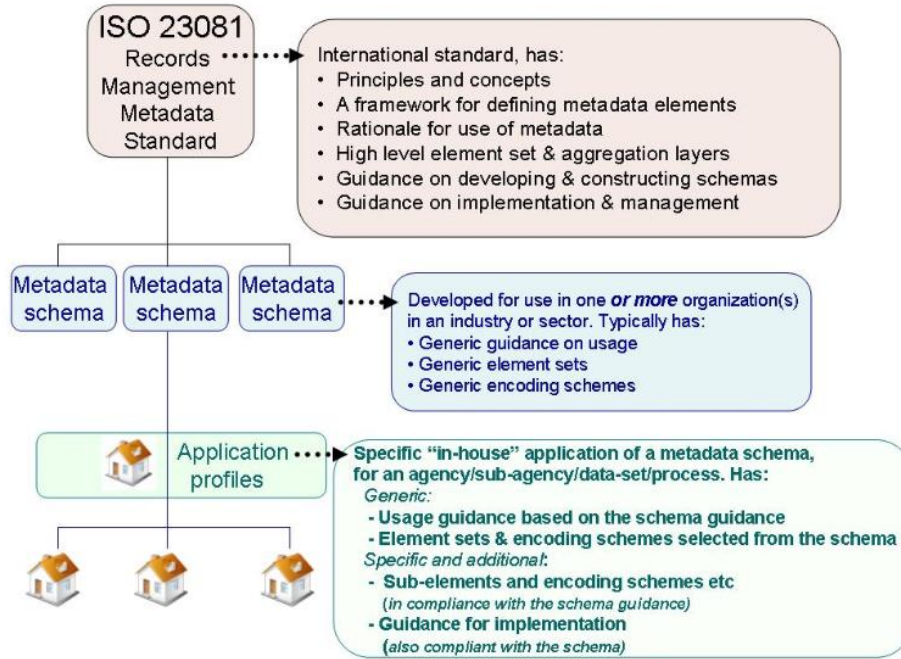
Skeeman valinnan jälkeen on valittava sisältöstandardi. Se kuvaa sitä, miten yksittäinen tieto syötetään metatietoelementteihin. Skeemat linkittyvät sisältöstandardeihin eri tavoin. Jotkut skeemat, kuten MARC, liittyvät hyvin läheisesti sisältöstandardeihin. MARC liitetään usein AACR:iin (Anglo-American Cataloguing rules) ja Library of Congressin asiansastoon. Dublin Core (DC) on puolestaan sisältöstandardeista paljon riippumattomampi, joten sen valitseminen skeemaksi johtaa siihen, että sisältöstandardi on valittava erikseen. (Beall, J. 2007.) Sisältöstandardeihin voimakkaasti liittyvän skeeman valinta johtaa yleensä myös siihen, että valitaan ko. sisältöstandardi ja siihen liittyvä ontologia. Sisältöstandardin valinta vaikuttaa yhteentoimivuuden toteuttamismahdollisuuksiin (crosswalk) eri tietokantojen välillä. Lisäksi skeemat voivat erota toisistaan myös siinä, millaisella merkistöllä tietoja voidaan tallentaa.

Metatietostandardi sisältää periaatteet, käsitteet sekä viitekehiksen, jossa metatietoelementit määritellään (a framework for defining metadata elements) Se sisältää myös perustelut metatiedon käytölle (rationale for use of metadata), ylätasen elementit (high level element sets), yhdistämissäännöt (aggregation layers), ohjeita skeemojen rakentamiseen ja kehittämiseen (guidance on developing and constructing schemas) sekä soveltamiseen ja hallintaan (guidance on implementation and management). (NISO.) Valitsemalla jonkin sisältöstandardin voidaan vaikuttaa siihen, miten tarkasti tietoa kirjataan, millaiset ovat metatiedon tuottamisen käytännöt ja millainen käyttäjäkunta tiedolla on (Library UC San Diego, 2010). Sisältöstandardin käyttöönottoon liittyvä tuki ja dokumentaatio vaihtelevat standardin mukaan – tavallisesti tunnetut kansainväliset standardit ovat hyvin dokumentoituja ja niiden käyttöönottoon löytyy paljon tukea. Dokumentaatio ja sovelluksen tuki vaihtelevat myös standardin taustaorganisaation mukaan, joten siitä kannattaa ottaa selvää ennen käyttöönottoa.

Ohjeet käyttöönoton tukena

Skeeman kirjalliset sovellusohjeet (application profile) ovat tärkeitä, sillä korostavat paikallisia sopimuksia ja käytäntöjä (IFLA Working Group on the Use of Metadata Schemas 2005). Sovellusohjeet voivat sisältää ohjeita siitä, mitä elementtejä käytetään ja millä tavalla sisältöjä kuvaillaan elementtien avulla, mitä termejä ja sanastoja käytetään sekä mitä tunnisteita käytetään. Jos metatietoa tuotetaan automaattisen prosessin avulla, sovellusohjeet olisi hyvä olla kaikkien metatietoa tallentavien saatavilla.

Myös työprosessi voi tarvita ohjeistusta, esimerkiksi siitä, miten tuotetaan uusia tietueita ja miten niitä muokataan (NISO). Hyvät kirjaamissäännöt (encoding scheme) sisältävät hyödyllisen ja yleisen ohjeistuksen, jota voidaan muokata, sekä kuvausta siitä, mitä organisaation aiempaan käytössä olleeseen skeemaan verrattuna uudessa skeemassa on korvattu/muutettu tai perustelut sille, miksi uusi skeema poikkeaa aiemmasta skeemasta.



Kuva: NISO ISO/TC46/SC11N8001R

4. Esimerkkejä erilaisista metatietoskeemoista

4.1 MARC metatietoskeemana

MARC on vakiintunut standardi koneellisesti luettavien tietueiden tuottamiseen ja se on lähes jokaisen kirjaston verkkoluettelon pohjana (Huber, A. 2005). MARC:in etuina pidetään usein sitä, että se on asiantuntijoiden kehittämä ja vakiintunut standardi, joka mahdollistaa yhdisteltävyyden hakukäyttöliittymässä sekä tarkan kuvailun (Riley, J. 2006). Toisin sanoen se on hyvin rakenteistettu ja tukee siten hyvin perinteisiä metatiedon funktioita kirjastoympäristössä ja joka sopii monille erilaisille aineistotyypeille (Zhao, J. 2015, luku 8). MARC:iin on tehty metatietokartoituksia (mappings) monista metatietojärjestelmistä siten, että MARC-tietueet voidaan luoda helposti ja niistä voidaan linkittää kirjastoluetteloihin (Huber, A. 2005).

MARC:in heikkoudet liitetään osaltaan samoihin seikkoihin: koska se on asiantuntijoiden kehittämä ja kuvailu on (ehkä liiankin) tarkkaa, sen käyttö edellyttää koulutusta ja osaamista sekä aikaa ja siten myös resursseja. Lisäksi hakukäyttöliittymäkään eivät hyödynnä kaikkia sen tarjoamia mahdollisuuksia. Myös sovellusten kirjavuutta ja kirjastomaisuutta pidetään jossain määrin ongelmina. (Riley, J. 2006.)

Suurimmat puutteet liittyvät kuitenkin siihen, että se on rajoittunut digitaalisten aineistojen ja esim. käsikirjoitusten kuvailussa sekä rakenteelliseen ja hallinnolliseen metatiedon kuvailussa, joiden merkitys korostuu erityisesti digitaalisten aineistojen kohdalla (vrt. digitaalisten aineistojen käytön tekniset edellytykset sekä aineistojen säilyttäminen). Lisäksi MARC:ia on kritisoitu siitä, että siitä puuttuu joustavuus ja yhteentoimivuus. Tällöin viitataan yleensä skeemojen muokattavuuteen ja eri formaateilla tuotetun metatiedon jakamismahdollisuuksiin, sekä metatiedon jakamiseen muiden organisaatioiden – tai jopa eri sektoreiden – kanssa. (Riley, J. 2006.)

4.2 RDF metatietoskeemana

RDF on malli metatietojen esittämiseksi internetissä tai oletusalusta konekäsiteltävälle semantiikalle. RDF tietomalli ei ole oikea metatietoskeema, vaan tarjoaa vain abstraktin, käsitteellisen kehysten metatiedon määrittelemiselle ja käytölle tai muille metatietomalleille. Sitä voidaan käyttää uusien mallien (aineistot/ ominaisuudet) käsitteellisen tiedon kuvailemiseen tai tuottamiseen tai informaation mallintamiseen, jota voi myös soveltaa verkkoaineistoihin ja eri syntaksiformaatteihin. (Open Metadata Handbook 2014.)

RDF:n avulla voidaan kuvata resursseja ja niihin liittyviä ominaisuuksia. Siitä on pyritty tekemään mahdollisimman joustava, jotta sitä voitaisiin käyttää moninaisissa yhteyksissä. RDF:n tietomalli on avoin, minkä ansiosta RDF-muodossa olevaa tietoa voidaan helposti muokata ja prosessoida. Lisäksi RDF sisältää joitakin tarkasti määriteltyjä tapoja tiedon esittämiseksi, mikä entisestään tehostaa sovellusten välistä yhteentoimivuutta. Tällaisten yhteisesti sovittujen esitystapojen avulla sovellukset tietävät heti, mikä on tietyn tiedon merkitys, ja osaavat hyödyntää tätä tietoa asiaan kuuluvalla tavalla. (Klyne & Carroll 2004, viitattu Mäkinen, V.-P., 2006.)

RDF:llä ei sinänsä ole tiettyä sovellusaluetta tai aineistolajia. Se on yleinen kehys, jota pitää laajentaa sanastolla/ontologialla, niin että sen avulla voidaan kuvailla jotakin. Aineistojen kuvailu perustuu objekteihin ja niiden ominaisuuksiin, jotka on kuvailtu RDF:llä. Sen avulla on siis mahdollista ku-

vailla/tuottaa uusia sanastoja aineistojen tai asioiden kuvailuun, jotka voivat edelleen muodostaa sanaston. RDF on tarkoitettu käsitteiden ja ontologioiden erittelyyn, joista ennen pitkää syntyy uusia käsitteitä. (Open Metadata Handbook 2014.)

RDF on hyvin joustava tapa kuvata aineistoja ja niiden välisiä suhteita. Sitä voidaan käyttää myös metatietosanastojen tuottamiseen. Sen vuoksi metatietosanasto voidaan tehdä niin, että se kuvaa aineistojen välisiä suhteita, organisaatioiden käytäntöjä ja aineistojen elinkaareen liittyviä asioita. (Miller, L. 2001.)

RDF ei ole sinänsä metatietoskeema vaan järjestelmä, jonka avulla voidaan koodata sellaisia skeemoja standardoiduissa rakenteissa. Koska se on kehitetty aineistojen kuvaamiseen verkossa, se tarjoaa standardisoidun tavan kuvata elementtien nimiä, niiden sisältöjä ja niiden välisiä suhteita, että aineistojen löytäminen ja tiedon jakaminen niistä olisi helpompaa. (Huber, A. 2005.) RDF yleensä koodataan XML:llä ja sitä voidaan käyttää kehyksenä lähes mille tahansa metatietoskeemalle (Badhusha, K.N. 2013).

RDF Schema puolestaan on RDF-tiedon kuvailuun kehitetty kieli. Sen avulla voidaan määritellä sanasto, jota sitten voidaan käyttää RDF-muotoisen tiedon esittämiseen. RDF Scheman avulla voidaan siis tehdä RDF-muotoista metatietoa kuvaileva metatietoskeema. (Brickley & Guha 2004, viitattu Mäkinen, V.-P., 2006).

RDF on käytetyin skeema ja se tarjoaa jo suuren valikoiman ontologioita/sanastoja, joita on toteutettu ja joita ylläpidetään W3C:n ja monien yliopistojen tukemana jatkuvasti. Mitä tahansa ontologiaa ja sanastoa, joka on kunnolla toteutettu ja ylläpidetty RDF:llä, voidaan soveltaa muihin avoimen bibliografian hankkeisiin, jos ne on julkaistu avoimella lisenssillä. (Open Metadata Handbook 2014)

RDF:n vahvuudet ja heikkoudet

RDF:n etuina pidetään sitä, että koneiden on helppo ymmärtää sitä ja sitä on hyvin helppo muokata, käsitellä ja laajentaa (Badhusha, K.N. 2013). Se on myös joustava tapa kuvata aineistoja, niiden välisiä suhteita ja ominaisuuksia – lähes mitä tahansa, ei vain yhtä sovellusaluetta tai aineistolajia – hyvin monenlaisten sanastojen/ontologioiden avulla. (Open Metadata Handbook 2014.) RDF tukee erilaisia skeemasisältöjä ja useita kieliä, ja elementtejä voidaan ryhmitellä siten kuin halutaan. RDF mahdollistaa kuvailustandardien kehittymisen. (Badhusha, K.N. 2013.)

Parasta RDF:ssä on kuitenkin sen avoimuus, joustavuus ja yhteentoimivuus: RDF tukee hakuja useista eri tietokannoista samalla haulla ja RDF-dataa voidaan jakaa avointen standardien avulla ja avoimena linkitettyinä datana. Toisaalta RDF erottaa tiedon hallinnan tiedon esittämisestä, mikä tekee samalla molemmista tehokkaampia. (Badhusha, K.N. 2013; Open Metadata Handbook 2014.) Lisäksi SPARQL: on hyvin toimiva hakujärjestelmä, jota voidaan käyttää hakujen tekemiseen tietokannasta, johon on tallennettu RDF-muotoista metatietoa (Open Metadata Handbook 2014).

RDF:N heikkouksina nähdään yleensä se, että XML:n tuottamiseen on hyvin vähän käyttäjäystävällisiä työkaluja. Sen käyttö myös edellyttää yhdenmukaista määrittelyä, tietojen vaihtamista ja metatiedon prosessointia. Lisäksi Jos haut eivät ole täysin ennustettavia, SPARQL-hakujärjestelmä saattaa tuottaa nollatuloksen. (Open Metadata Handbook 2014.)

RDF:n avulla tehtävä kuvailu myös riippuu yhdestä tai useammasta ulkoisesta lähteestä: RDF-tietueessa metatieto ei ole itsenäinen kokonaisuus vaan enemmänkin osa globaalia tietoverkkoa, jossa keillä tahansa on mahdollisuus lisätä mihin tahansa aineistoon mitä tahansa metatietoa. Metatieto ei

myöskään ole yhdessä RDF-dokumentissa vaan esimerkiksi metatiedon luokittelut voivat olla eri paikoissa ja käännöksiä voidaan tehdä missä tahansa. RDF ei ole metatietoskeema vaan käsitemalli. (Open Metadata Handbook 2014.)

RDF:n käyttö vaatii suuria tallennusresursseja ja infrastruktuuria. RDF:ää onkin syytetty skaalattavuuden puutteesta – yksinkertaistakin tehtävää varten, kuten muutaman miljoonan bibliografisen tietueen hallintaa, täytyy tehdä suuri urakka. (Open Metadata Handbook 2014.)

Metatiedon avoimuuskin asetetaan joskus kyseenalaiseksi: "RDF voi olla hyvä abstrakti malli, mutta sen käytännön toteutuksia ja tukea bibliografisen tiedon avaamiseksi tarvitaan vielä. Vain suuret toimijat voivat hallita RDF:n tarvitsemaa infrastruktuuria, eikä avoimuutta voida laskea niiden varaan. (Open Metadata Handbook 2014.)

4.3 Esimerkkejä muista metatietoskeemoista

MARCXML

Kuten MARC, XML on tiedonsiirtoformaatti. XML on kuitenkin metakieli eikä sillä ole tarkkaan määriteltyä joukkoa kenttiä tai syntaksia, mutta se tarjoaa käyttäjille joustavuutta muodostaa ne itse (Harold & Means, 2001, s. 3–4, viitattu: Mascaró, M.J. 2004). XML:n kannattajat väittävät, että tämä joustavuus mahdollistaa helpon muokkaamisen ja kuvailustandardien kehittymisen (Fiander, 2001, viitattu: Mascaró, M.J. 2004, 16)

MARC:in kritisoinnista on syntynyt XML:ään liittyvä esitys: MARCXML, joka on XML-skeema, jonka avulla MARC-tietue voidaan muokata XML:ksi. Sillä on rakenteellisia etuja kuvailumenetelmänä ja siihen on myös lisätty XML:n muokkaamisen mukanaan tuoma joustavuus. (Mascaró, M.J. 2004.)

MARCXML:n vahvuudet ja heikkoudet

MARCXML:n vahvuuksina pidetään sitä, että se kopioi MARC21:n rakenteen XML-syntaksiksi. Siihen sisältyy myös piilo-oletus, että sisällöt/ skeemaan syötettävät arvot ovat samat kuin MARC:issa. Kirjastoille tyypilliseen tapaan se tarjoaa kuvailun tarkkuutta ja siitä voidaan tehdä hävikitön konversio MARC:iin. Lisäksi sen avulla on mahdollista tallentaa XML-kuvailua myös laajoihin digitaalisiin aineistoihin ja se voi toimia välittävänä skeemana MARC:in ja toisen, XML-koodatun, metatietoformaatin välillä. (Riley, J. 2006.)

MARCXML:n heikkouksina puolestaan pidetään sitä, että sen syntaksi on hyvin moniulotteinen. Lisäksi MARCXML:n kentät ovat numeerisia eikä suora tietojen syöttäminen ilman XML-koodia ei ole mahdollista. Täydellinen sisällöllinen vastaavuus edellyttää lisätyökaluja, että metatiedot saadaan vastamaan XML-skeemaa. (Riley, J. 2006.)

MODS

MODS on MARC-pohjainen skeema, joka tarkastelee MARC elementtisarjaa XML:llä koodattuna. MODS on kuvailevan metatiedon standardi, eli XML skeema, joka on suunniteltu digitaalisten aineistojen metatiedon kuvailemiseen kirjastoissa. (Alemneh, D.G. 2007.)

MODS-skeema on tarkoitettu bibliografisten elementtien kuvailuun, ja sitä voidaan käyttää erilaisiin tarkoituksiin, erityisesti kirjastosovelluksiin. MODS on MARC:in kaltainen, mutta tarkoitettu yksinkertaisemmaksi. MARC:iin verrattuna siihen on lisätty joitakin elementtejä ja poistettu joitakin. Lisäksi elementtejä on ryhmitelty uudella tavalla. (Riley, J. 2006.)

MODS:n vahvuudet ja heikkoudet

MODS:n etuina pidetään sen kuvaustarkkuutta ja kirjastomaista kuvailua verrattuna esim. Dublin Coreen. Se on sopiva skeema, erityisesti silloin, kun halutaan tallentaa XML-kuvailua laajoihin digitaalisiin aineistoihin ja tarjota hävikitön konversio MARC:iin, sillä sen elementit ovat semanttisesti samankaltaisia kuin MARC:issa ja se on myös hierarkkinen. MODS ei kuitenkaan sisällä taustaoletuksia kuvailusäännöistä ja elementtikuvauksia voidaan käyttää yhä uudelleen. Sen tagit ovat myös kielipohjaisia, eivätkä numeerisia, toisin kuin MARC:in kentät, ja siten helpompia käyttää. XML-skeeman käyttö tarjoaa siis kaivattua joustavuutta kuvailuun. (Alemneh, D.G. 2007.)

Heikkouksina pidetään usein sitä, että huolimatta samankaltaisuudesta MARC:in kanssa, MODS:ia ei voida konvertoida MARC:iin ja takaisin häviöttömästi vaan yksi MODS-elementti saattaa sisältää useita MARC-elementtejä. Tällöin jonkin verran MARC:in semantiikasta häviää konversioissa. Lisäksi MODS on pääasiassa kirjastoyhteisön käytössä eikä sen avulla tuotettu metatieto ole siten helposti jaettavissa. (Alemneh, D.G. 2007.)

XML

XML-on paitsi dokumenttityypin määrittely, joka kuvaa sitä, millaisia tageja ja niiden välisiä suhteita XML-dokumentissa voidaan käyttää. XML on myös XML-skeema, joka on uudempi tapa tallentaa tietoa XML-järjestelmässä. Sen avulla voidaan määritellä tarkemmin säännöt erillisessä XML-dokumentissa, joita XML-dokumentin pitää noudattaa. Siten XML-skeema ja siihen liittyvä skeemakieli on paljon enemmän kuin dokumenttityypin määrittely, ja mahdollistaa helpommin XML-datan kehittämisen ja ylläpidon. (Huber, A. 2005.) XML metatietotietue on itsenäinen kokonaisuus, jolla on hyvin määritelty rakenne (Kunze, T. ym. 2002).

XML on metatietojärjestelmien kieli. XML-kieli suunniteltiin alun perin elektronisten tekstien merkintäkieleksi, mutta sitä on käytetty hyvin moniin eri metatietosovelluksiin myös (Huber, A. 2005). XML:ää pidetään välitysformaattina kommunikoinnille ja sen menestyksen uskotaan takaavan mahdollisuudet tukea useita erilaisia sovelluksia (McDonough, J. 2008).

XML:n vahvuudet ja heikkoudet

XML:llä on paljon etuja metatietojen tallennuksessa: vahvuus, ohjelmistoriippumattomuus ja sen vaihdettavuus järjestelmästä toiseen sekä rakenteen sopivuus digitaalisten aineistojen kuvaukseen (Huber, A. 2005). XML lisää joustavuutta, koska uusia tageja voidaan lisätä tarpeen mukaan ja rakenne voi olla kuinka monimutkainen tahansa. Se myös takaa datan rakenteen validoinnin ja mediariippumattomuuden, kun sisältöä voidaan tuottaa eri formaateissa. (McDonough, J. 2008). XML on lisäksi myyjä- ja alustariippumatonta: metatietoa voidaan tuottaa kaupallisilla ohjelmistoilla tai jopa yksikertaisilla tekstieditoreilla. (Bosak, 1998, viitattu: McDonough, J. 2008.) Näistä syistä monet katsovat, että se tukee aineistojen ja tiedon yhteentoimivuutta globaalisti.

Lisäksi XML:n etuina pidetään sitä, että yhteentoimivuus sisällön sekä sovellusten ja alustojen suhteen on hyvä. Keskeiset kehitystä kuvaavat sanat ovat vapaus ja riippumattomuus: XML:n soveltajat ja sisällöntuottajat vapautuvat työkalujen myyjien kontrollista ja käyttäjät voivat valita omia näkökulmia sisältöön ja rakentaa erilaisia hyviä massatyökaluja sisällönhallintaan. XML:n soveltajilla on myös mahdollisuus käyttää itsenäisten ohjelmistotuottajien ohjelmia ja julkaista kansainvälisesti mediasta riippumatta. Myös käyttäjillä on vapautua ohjelmistojen rajoituksista ja vapaus luoda omia tageja (eli omia näkökulmia sisältöihin) sekä määritellä omia rakenteita aineistoille ja tiedoille käyttäen omaa semantiikkaa. (McDonough, J. 2008.)

XML:n keskeisenä heikkoutena pidetään sitä, että XML:n tuottamiseen on hyvin vähän käyttäjäystävällisiä työkaluja eikä sen käyttö mahdollista muiden metatietostandardien käyttöä. Lisäksi, menestyksestään huolimatta, XML ei ole täyttänyt lupauksia yhteentoimivuudessa, yllättävää kyllä juuri syntaksin tasolla (McDonough, J. 2008). Kirjastot ovat olettaneet, että XML-standardin käyttö sisältää ja rakenteita kuvaavaa metatietoa tallennettaessa takaisi yhteentoimivuuden ja auttaisi erottamaan selkeästi sisällönkuvailun ja korkeamman tason työkalut sekä palvelut, joiden avulla voidaan tuottaa standardeja sisältökoodeja (Hurley et al., 1999, viitattu: McDonough, J. 2008). XML-skeemojen suunnittelijat ja soveltajat usein etsivät hyviä käytäntöjä, joiden tavoitteena on tuottaa joustavia, laajennettavia, modulaarisia ja yleisiä välineitä metatiedon tuottamiseen, mutta ikävä kyllä se johtaa vain yhteentoimivuuden murenemiseen. Aineistojen yhteentoimivuus järjestelmien välillä edellyttää samanlaisten skeemojen valintaa, ei niiden soveltamista. (McDonough, J. 2008.)

ONIX

ONIX on XML-sovellus, joka on suunniteltu kirja-alan metatiedon välittämiseen kustantajilta myyjille. Sen seurauksena on syntynyt hyvä mahdollisuus kuvata bibliografinen avaindata, hinta- ja saatavuustiedot (Huber, A. 2005).

ONIX:n vahvuudet ja heikkoudet

ONIX sopii kuvailevan metatiedon välittämiseen, mutta rakenteellisen ja hallinnollisen metatiedon kuvaamisessa sekä muiden kuin kirja-aineistojen kuvaamisessa ONIX on hyvin rajoittunut (Huber, A. 2005).

Lähteet

Alemneh, D.G. 2007. An Introduction to MODS: The Metadata Object Description Schema. <http://www.library.unt.edu/sites/default/files/documents/digital-libraries-uploads/tech-talks/MODS.ppt>

Ashraf, T. & Gulati, P.A. 2013. Design, Development, and Management of Resources for Digital Library Services.

Baca, M. 2003. Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information. In: Cataloging & Classification Quarterly Vol. 36, No. 3/4, 2003. http://polaris.gseis.ucla.edu/gleazer/260_readings/Baca.pdf

Baca, M. (toim.) Gill, T. & Gilliland, A.J. & Whalen, M. & Woodley, M.S. 2008. Introduction to Metadata. Getty Research Institute, Los Angeles, 2. painos.

Badhusha, K.N. 2013. Significance of Metadata, teoksessa: **Ashraf, T. & Gulati, P.A. 2013.** Design, Development, and Management of Resources for Digital Library Services.

Beall, J. 2007. Discrete Criteria for Selecting and Comparing Metadata Schemes. In: Against the Grain, Vol. 19, No 1, 2007, article 7. <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=5228&context=atg>

Bowen, J. 2008. Metadata to Support Next-Generation Library Resource Discovery: Lessons from the eXtensible Catalog, Phase 1. In: Information technology and libraries, Vol 27, No 2, 2008. <https://ejournals.bc.edu/ojs/index.php/ital/article/view/3253/2866>

Condee-Padunova, K. & Wilson, L. 2015. More Metadata, Less Process? University of Pittsburgh <http://metaware.buzz/2015/09/07/more-metadata-less-process/>

Foulonneau, M. & Riley, J. 2008. Metadata for Digital Resources, Implementation, Systems Design and Interoperability, Chandos Information Professional Series. Chandos Publishing, Cambridge.

Greenberg, J. 2003. Metadata and the World Wide Web. In Encyclopedia of Library and Information Science. (pp. 1876-1888). New York: Marcel Dekker, Inc. Viitattu: Mascaro, M.J. The Value of Flexibility in Metadata Schemas. Pro gradu, the School of Information and Library Science of the University of North Carolina. <https://ils.unc.edu/MSpapers/2955.pdf>

Greene, M. A., and Meissner, D, More Product, Less Process: Revamping Traditional Archival Processing. The American Archivist 68, no. 2 (2005): 208-63. Accessed July 10, 2015. <http://www.archivists.org/prof-education/pre-readings/IMPLP/AA68.2.MeissnerGreene.pdf>

Honma, T. 2015. Organizing Existing Metadata Terms and Structural Constraints to Support Metadata Schema Creation, University of Tsukuba, Japan. <http://jenserikmai.info/glocalko/Honma.pdf>

Huber, A. 2005. Metadata in the Oxford Digital Library. <http://www.odl.ox.ac.uk/metadata.htm#refs>

IFLA Working Group on the Use of Metadata Schemas 2005. Guidance on the Nature, Implementation, and Evaluation of Metadata Schemas in Libraries. Final Report of the IFLA Cataloguing Section For the Review and Approval of the IFLA Cataloguing Section

http://www.ifla.org/files/assets/cataloguing/pubs/metadata_schemas-20050731.pdf

Kelly B. 2006. Choosing a Metadata Standard For Resource Discovery. UKOLN, The UK Office of Library Networking.

<http://www.ukoln.ac.uk/ga-focus/documents/briefings/briefing-63/html/>

Kennedy, M. R. 2008. Nine questions to guide you in choosing a metadata schema. Journal of Digital Information Vol 9, No 1, 2008, ISSN: 1368-7506.

<https://journals.tdl.org/jodi/index.php/jodi/article/view/226/205>

Kunze, T. & Brase, J. & Nejdil, W. 2002. Editing Learning Object Metadata: Schema Driven Input of RDF Metadata with the OLR3-Editor.

https://www.researchgate.net/publication/2546410_Editing_Learning_Object_Metadata_Schema_Driven_Input_of_RDF_Metadata_with_the_OLR3-Editor

Library UC San Diego 2010. Metadata Schemas and Standards.

<http://libraries.ucsd.edu/services/data-curation/sharing-discovery/describe-your-data/metadata-schemas.html>

Liu, J. 2007. Metadata and Its Applications in the Digital Library Approaches and Practices, Libraries Unlimited, Westport, Connecticut.

Mai Chan, L. & Lei Zeng, M. 2006. Metadata Interoperability and Standardization – A Study of Methodology Part I. In: D-Lib Magazine 2006, Vol. 12 No. 6, ISSN 1082-9873.

<http://www.dlib.org/dlib/june06/chan/06chan.html>

Mascaro, M.J. 2004. The Value of Flexibility in Metadata Schemas. Pro gradu, the School of Information and Library Science of the University of North Carolina.

<https://ils.unc.edu/MSpapers/2955.pdf>

McDonough, J. 2008. Structural Metadata and the Social Limitation of Interoperability: A Sociotechnical View of XML and Digital Library Standards Development, The Markup Conference 2008, Balisage, Aug. 12 - 15, 2008.

<http://www.balisage.net/Proceedings/vol1/html/McDonough01/BalisageVol1-McDonough01.html>

Miller, L. 2001. Using RDF query to manage metadata vocabularies.

<http://ilrt.org/discovery/2001/06/process/>

Mäkinen, V.-P. 2006. Metatietoskeema lainsäädäntöprosessin asiakirjoille. Pro gradu, tietojärjestelmätiede, Tietojenkäsittelytieteiden laitos, Jyväskylän yliopisto. http://www.it.jyu.fi/raske/julkaisut/Gradu_Makinen20060320.pdf

Mölsä, H. 2011. Astuvansalmen kallion ontologisointi - prosessikuvaus. Opinnäytetyö, Sähköisen asiainnin ja arkistoinnin koulutusohjelma, Mikkelin ammattikorkeakoulu.

https://www.theseus.fi/bitstream/handle/10024/27078/Molsa_Henna_2011.pdf?sequence=1

NISO National Information Standards Organization Where to Start Advice On Creating a Metadata Schema ISO/TC46/SC

http://www.niso.org/apps/group_public/download.php/5271/N800R1_Where_to_start_advice_on_creating_a_metadata_schema.pdf

Open Metadata Handbook 2014. Metadata Standards.

https://en.wikibooks.org/wiki/Open_Metadata_Handbook/Metadata_Standards#RDF.2FOWL

Park, J. & Tosaka, Y. 2010. Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability. In: Information technology and libraries, Vol 29, No 3. 2010. <https://ejournals.bc.edu/ojs/index.php/ital/article/download/3136/2750>

Riley, J. 2006. An introduction to metadata in digital projects. <http://slideplayer.com/slide/7541359/>

RuCore Rutgers University Libraries Community Repository 2015. Understanding Metadata

<https://rucore.libraries.rutgers.edu/cyber/metadata.php>

Viljanen, K. 2008. Metatietoskeemat. Semantic Computing Research Group (SeCo), Helsinki University of Technology (TKK), Laboratory of Media Technology ja University of Helsinki, Department of Computer Science.

<http://seco.cs.aalto.fi/events/2008/2008-01-25-finnonto/esitykset/viljanen-hajautettu-sisallontuotanto-ja-metatietoskeemat-2008-01-25.pdf>

Youngs, K. & Hull, S. & Duffy, T. & Faber, M. 2015. JISC Digital Media Putting Things in Order: a Directory of Metadata Schemas and Related Standards.

<http://www.jiscdigitalmedia.ac.uk/guide/putting-things-in-order-links-to-metadata-schemas-and-related-standards>

Zhao, J. 2015. Best Practices For Structural Metadata, luvut 8 ja 11

http://www.library.illinois.edu/dcc/bestpractices/chapter_11_structuralmetadata.html

http://www.library.illinois.edu/dcc/bestpractices/chapter_08_descriptive_metadata.html#7.2_Metadata_Recommendations