

STANDARDIEN TOTEUTTAMISESTA

Juha Hakala

Kaksiosainen Signum-artikkelini kuvasi standardien kehittämistä ([Hakala 2023A](#)) ja siinä ilmeneviä haasteita ([Hakala 2023B](#)). Mutta standardeilla on arvoa vain silloin, kun niitä käytetään. Tässä artikkelissa kerron käytännön kokemuksiin perustuen, miten kirjastojärjestelmätoimittajilta sujuu standardien toteuttaminen. Toimiiko niiden yhteistyö kirjastojen kanssa? Saavatko kirjastot sitä mitä tilaavat, ja olemmeko me vaatineet oikeita asioita?

Yleisesti ottaen kirjastojärjestelmien standardituki on hyvällä tolalla, mutta poikkeuksiakin löytyy. Vanhentuneet tai liian monimutkaiset standardit ovat ongelma, koska kirjastojenkin voi silloin olla vaikea esittää tarpeensa riittävän täsmällisesti, eivätkä järjestelmätoimittajan resurssit ja osaaminen riitä haluttujen ominaisuuksien toteuttamiseen. Kaukopalvelun ja tiedonhaun standardit ovat tästä hyviä esimerkkejä.

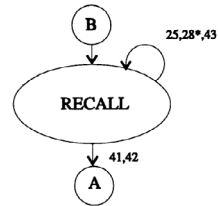
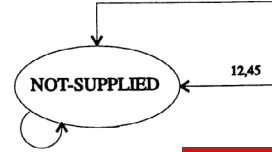
Kaukopalvelu

Yksi molempia Linnea-hankkeita ja niiden jälkeenkin toteutettuja projekteja yhdistävä tekijä on kaukopalveluun liittyvien vaatimusten kariutuminen.

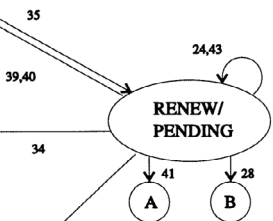
Ensimmäisessä Linnea-hankkeessa kaukopalvelun toteutusta koskevat vaatimukset olivat yleisiä, ja yleiseksi jäi myös järjestelmätoimittajan vastaus: kaukopalvelun toteutuksen periaatteita selvitetään. Kaukopalvelustandar-

di [ISO 10160/10161](#):n ilmestyttyä vuonna 1993 aloitettiin keskustelu sitä noudattavan kaukopalvelutoiminnon rakentamisesta VTLS-ohjelmistoon, mutta käytäntöön asti ei päästy. Tässä suhteessa VTLS, Inc. ei ollut poikkeus, sillä vain harvoihin kirjastojärjestelmiin rakennettiin tämän alkuperäisen kaukopalvelustandardin tuki. Esimerkiksi British Libraryn BLDSC-yksikkö epäonnistui kehittämishankkeessa.

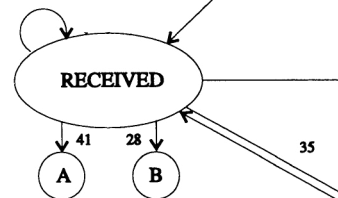
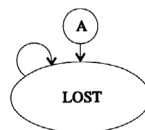
ISO 10160/10161 -standardin implementoinnin hankaloitti sen määrittelemän asiakas- ja palvelinohjelmiston välisen protokollan yhteydellisyys. Standardi määrittelee joukon tiloja, joiden kautta kaukopalveluprosessin on edettävä, tilauksen teosta lainatun niteen palauttamiseen lainaavaan kirjastoon. Ohjelmiston pitää tietää, missä tilassa kukin kaukolainaprosessi on. Jos kirjastojärjestelmä ei jakanut standardin käsitystä kaukopalveluprosessin kulusta, ISO 10160/10161 -standardin tukeminen kyseisessä ohjelmistossa



6,13,14,24,
28**,29**,32,
34** 39,40,43,44



6,13,14,24,
28**,34**,
39,40,
41**,42



25

oli vähintään hankalaa ja kenties jopa mahdotonta.

Standardin kaukopalveluprosessi perustui siihen, miten suuret tieteelliset kirjastot toimivat 1970/1980-luvuilla erityisesti Kanadassa, missä suuri osa standardin kehittämistyöstä tehtiin. Noista ajoista kirjastojen kaukopalvelukäytänteet olivat muuttuneet, joten standardi oli 2000-luvulle tullessa tältä osin eilispäivää. Standardi oli vanhentunut myös teknisesti, koska siinä määritellyt asiakas- ja palvelinsovelluksen väliset viestit on koodattava [Abstract Syntax Notation One](#)-standardin ja sen [Basic Encoding Rules](#)-sääntöjen mukaisesti. Nämä 30 vuotta vanhat määrittelyt on jo kauan sitten korvannut XML ja siihen perustuvat tekniikat kuten JSON, ja ASN.1/BER-taitoisten ohjelmoijien löytäminen on yhä vaikeampaa.

Voyager-ohjelmistoa käyttöön otettaessa kaukopalvelustandardi oli ollut jo pitkään valmis, mutta sitä tukevia ohjelmistoja ei vielä vuosia standardin ilmestymisen jälkeenkään ollut montaa. Kun esimerkiksi BLDSC-tuki puuttui, kaukopalvelustandardin toteutuksesta saatava hyöty olisi jäänyt niukaksi. Linnea2-vaatimusmäärittelyyn standardiin pohjautuva kaukopalvelumoduli kirjattiin, ja Endeavor vastasi asian olevan selvitettävänä, ja sinne se myös jäi. Tämä ei välttämättä

ollut huono asia, koska alkuperäinen kaukopalvelustandardi on poistumassa käytöstä.

Kaukopalvelun standardisoinnista vastaa ISO TC 46/SC 4 Technical interoperability. Eri vaihtoehtoja selvitettyään se päätti kehittää kokonaan uuden ja ajantasaisen standardin. [ISO 18626 Interlibrary Loan Transactions](#) julkaistiin vuonna 2014, ja vastuun kantaja oli Kanadan sijaan Tanska. Siellä haluttiin vahvistaa kansallisen ja erittäin hyvin toimivan (ja rahoitetun) kaukopalveluverkon standardiperustaa. Uusi standardi on yhteydetön eli ei tee mitään

Uutta ja vanhaa standardia noudattavat sovellukset eivät pysty kommunikoimaan keskenään.

oletuksia prosessin tilasta taustajärjestelmissä. Se ei myöskään sovelle protokollaviestien koodauksessa 30 vuotta vanhaa teknologiaa, vaan Web-tekniikoita. Suomeksi tämä merkitsee että uutta ja vanhaa standardia noudattavat sovellukset eivät pysty kommunikoimaan keskenään.

ISO 18626 -standardin viimeisin versio on vuodelta 2021, mutta kehitystyö jatkuu. ISO TC 46/SC 4 käynnisti 2023 vuosikokouksessaan selvityksen, jossa arvioidaan tarve ja mahdollisuudet lisätä standardiin mm. e-aineistojen kaukolainaukseen tarvittavia ominaisuuksia.

Uuden kaukopalvelustandardin käyttöönotto on edennyt hitaasti mutta varmasti. Teknisesti se on osoittautunut odotusten mukaisesti edeltäjään-

sä helpommin implementoitavaksi. Tätä kirjoitettaessa standardia tukee Tanskan ansiosta Ex Libris Alma, ja moni muukin kirjastojärjestelmä – toivottavasti ennen pitkää myös VuFind, KOHA ja kaikki muutkin Suomessa laajalti käytetyt ohjelmistot.

Tiedonhaku

A lun perin vuonna 1988 julkaistu [ANSI/NISO Z39.50](#) on Kongressin kirjaston kehittämä amerikkalainen standardi, joka soveltuu erityisesti tiedonhakuun MARC-formaattia soveltavista bibliografisista tietokannoista. Kirjastoleiman vuoksi sen käyttö ei laajentunut muun tyyppisiin tietokantoihin, vaikka periaatteessa se olisi ollut mahdollista.

Standardia käytettäessä tiedonhakijan ei tarvitse tietää haun kohteena olevasta tietokannan asiakasliittymästä mitään. Tämä oli erittäin kätevää silloin, kun järjestelmät olivat komentopohjaisia. Nykyään etäkäytön haasteena ei enää ole hakuliittymän komentojen opiskelu, vaan semantiikka. Tehokas tiedonhaku edellyttää etätietokannan sallimien hakutermin tuntemusta. Tätä ongelmaa Z39.50 ei poista, mutta tarjoaa siihen tukea. Z39.50-asiakasohjelmisto voi testihakujen avulla selvittää, mitä hakutermejä etätietokanta tukee, ja missä muodossa tietueet voidaan siirtää.

Z39.50-asiakasohjelmiston käyttäjä voi ennalta selvittää sallitut hakutermit ja formaatit. Jos hän tekee ISBN-haun ja tilaa tietueet MARC-for-

maatissa, Z39.50-protokollaa käyttävä sovellus muuttaa nämä toiveet ohjelmistoriippumattomaan muotoon. Jokaiselle hakutermitte, metadataformaateille jne. on määritelty Z39.50-standardissa OID-standardin mukainen tunnus. Jos haluaa tietueet MARC-muodossa, asiakasohjelma pyytää niitä palvelimelta OID-tunnuksella 1.2.840.10003.5.10.

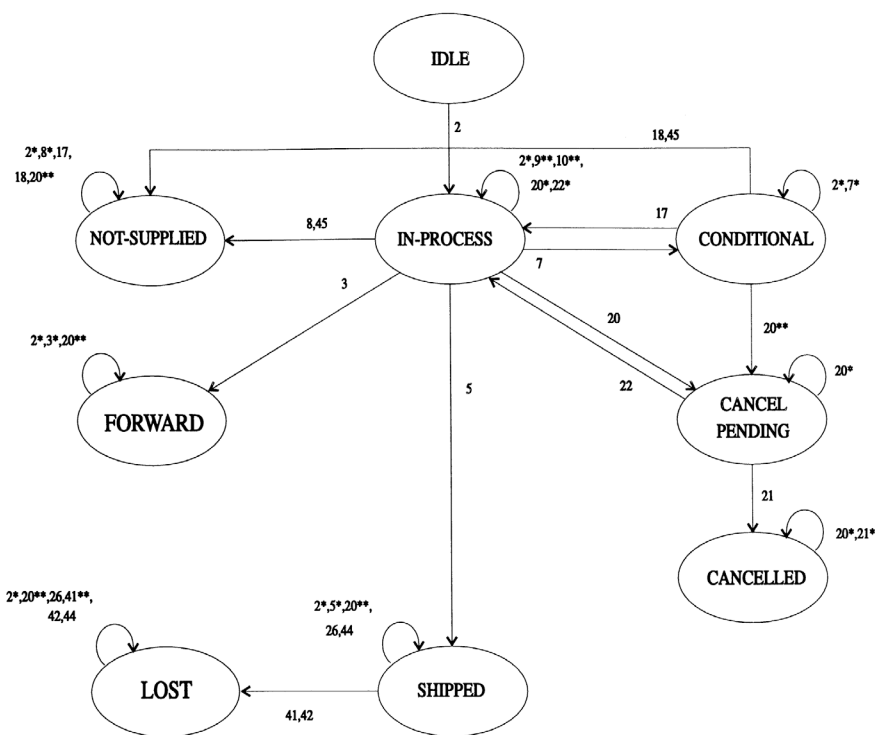
Z39.50-tiedonhaku ei onnistu, jos tiedonhakijan valitsema hakutermi puuttuu kohdetietokannasta, tai jos sitä ei saa käyttää sanahausa. Myös standardi itsessään voi rajoittaa hakumahdollisuuksia. Esimerkiksi pysyvillä tunneilla ei ole omia OID-tunnuksia, vaan niihin kohdistuvissa hauissa asiakasohjelmisto joutuu käyttämään yleistä standard identifier -hakutermiä, jonka OID on 1007.

Z39.50-pohjainen tiedonhaku toimii parhaiten silloin, kun kohdejärjestelmät ovat riittävässä määrin sekä teknisesti että semanttisesti yhteensopivia. Tekninen yhteensopivuus perustuu esimerkiksi samojen Z39.50-palvelujen tukemiseen. Vähimmäisvaatimus on standardissa määriteltyjen Init, Search ja Present-palvelujen tuki, mutta ne ovat vain murto-osa koko standardista. Tietueiden siirrossa on tuettava vähintään MARC-tietueiden lähettämistä karsimattomina sekä ISO 2709 -vaihtomuotostandardia.

Eri kirjastojärjestelmien Z39.50-toe-
teutuksien semanttista yhteismitalli-

Tehokas tiedonhaku edellyttää etätietokannan sallimien hakutermin tuntemusta.





Kaukolainan käsittelyprosessi tilaajan näkökulmasta, kun kaukolainaa ei tarvitse palauttaa. ISO 10160 -standardin mukaan.

suutta tutkittiin EU:n ONE- eli [OPAC Network in Europe](#) hankkeessa vuosina 1995–1997. Hämmäntävä ja virtuaalisten yhteisluetteloiden kannalta tuhoisa tulos oli, että ainoa hakutermi, jota oli mahdollista käyttää kaikissa testiin osallistuneissa tietokannoissa, oli nimekkeen yksittäinen sana.

Kirjastojen tietokantojen semanttinen yhteismitallisuus ei ollut läheskään niin hyvällä tasolla kuin oli etukäteen luultu. Ongelmalle löytyi useita syitä. 1990-luvulla käytettiin kansallisia MARC-formaatteja ja kuvailusääntöjä, ja niissä oli merkittäviä eroja. Kirjastot tekivät sisällönkuvailua eri luokituksilla ja sanastoilla, mikä teki aiheenmukaisen tiedonhaun useista tietokannoista yhtä aikaa käytännössä mahdottomak-

si. Auktoriteettitietokannat eivät olleet nykyisellä tasolla, ja haku rajautui usein toimijoiden auktorisoituun nimenmuotoon, joissa oli paljon hajontaa maiden välillä. Esimerkiksi Tshaikovskille löytyi yli 40 varianttinameä. 30 vuotta ei voinut käyttää ISNI- tai ORCID-tunnusta, jolla haku olisi onnistunut kohdetietokannoissa käytetyistä nimen muodoista riippumatta.

Metatietoihin liittyvien haasteiden lisäksi ongelmia aiheuttivat kirjastojärjestelmien Z39.50-toteutuksien puutteet. Niistä osa oli ilmeisiä, kuten joidenkin standardissa määriteltyjen palvelujen puuttuminen kokonaan. Käytännön tiedonhaun kannalta ongelmallisempia olivat tuettujen hakutermin tietokantakohtaiset rajoitukset.

Esimerkiksi Voyagerin Web-haussa oli mahdollista rajata hakua aineistotyypillä, mutta Z39.50-haussa aineistotyyppiä ei voinut käyttää hakuterminä, koska Boolean haussa käsiteltävien tulosjoukkojen maksimikoko oli Voyagerissa rajattu suorituskykyistä liian pieneksi.

Kaikista ongelmistaan huolimatta Z39.50 oli ja on Linnea-kirjastoille tärkeä standardi poimintaluetteloinnin ja Z39.50-hakuun tarjotun LINDA/MELINDA-yhteisluettelon vuoksi. Ohjelmiston hankintasopimuksessa Endeavor sitoutui lisäämään indeksilistauksen (Scan), tulosjoukon lajittelun (Sort) ja lukuisia muita uusia piirteitä Voyagerin Z39.50-palvelimeen. Nämä uudistukset luvattiin ohjelmistoversioon Voyager 2001.

Käytännössä mitään näistä parannuksista ei saatu, eikä Endeavorin Z39.50-osaamisessa muutenkaan ollut kehumista. Kun Kongressin kirjasto otti Voyagerin käyttöön, sen Z39.50-asiantuntija Larry Dixson havaitsi ettei Voyagerin Z39.50-palvelimen Search-toiminto ollut luotettava. Tietyt tuiki tavalliset ja täysin luvalliset Z39.50 haut kaatoivat palvelimen. Se oli vakava ongelma Kongressin kirjastolle, koska sen tietokantaa käytettiin ja käytetään edelleen paljon Z39.50-poimintaluettelointiin. Kongressin kirjaston havaitsemat ongelmat olivat kriittisiä myös Linnea2-hankkeelle, koska ne riivaisivat myös LINDA-tietokannan käyttöä.

Kongressin kirjaston tarjoamas-

Voyagerin Z39.50-palvelimen Search-toiminto ei ollut luotettava. Tiedyt tuiki tavalliset ja täysin luvalliset haut kaatoivat palvelimen.

ta erittäin yksityiskohtaisesta dokumentaatiosta huolimatta Endeavor oli kykenemätön ja ehkä halutonkin Kongressin kirjaston ongelmia ratkaisemaan. Selvää ohjelmistovirhettä voi varsin pitkälle väittää ohjelmiston ominaisuudeksi, ja välttyä remontilta. Kongressin kirjaston avuksi tuli tanskalainen Index Data -yritys, jonka Z39.50-osaaminen oli toista luokkaa kuin Endeavorin.

Index Data kehitti Voyagerin Z39.50-palvelimen tueksi apuohjelman, jonka turvin palvelin pysyi tollillaan. Kaikki Z39.50-haut ohjattiin palvelimelle tämän apuohjelman kautta. Kongressin kirjaston Z39.50-asiantuntija Larry Dixson keräsi Voyagerin Z39.50-palvelimen lokista listan ”tappavista” kyselyistä, jonka hän lähetti Index Datalle. Apuohjelma tarkasti kaikki palvelimelle saapuvat haut ja torjui ongelmatapaukset joko kokonaan tai muunsi ne sellaisiksi, että Voyager selvisi niistä. Sama apuohjelma turvasi sittemmin myös Linnea-verkon yhteisjärjestelmän tietokantojen Z39.50-palvelimet. Mutta meidänkin piti hoitaa asia Index Datan kanssa, suoraan Endeavorilta korjausohjelmaa ei saanut.

Edellä kerrotusta voi päätellä, ettei Endeavorilla ollut osaamista eikä edes kiinnostusta toimivan Z39.50-palvelinohjelmiston tarjoamiseen, eikä siten myöskään ohjelmiston kehittämiseen Linnea2-vaatimusmäärittelyn edellyttämällä tavalla. Ehkä Z39.50-palvelimen rakentaminen ja integrointi kirjastojärjestelmään vaatii taitoja, joita tyyppillisellä kirjastojärjestelmän kehittäjällä ei ole. Osaamista olisi ollut helppo hankkia, sillä Ex Libris ja monet muut ohjelmistotalot ovat tehneet Z39.50-kehitystyössä yhteistyötä Index Datan kanssa. Endeavor otti mieluummin kantaakseen heikosti toimivan Z39.50-palvelimen aiheuttaman mainehaitan.

Endeavorin ongelmista voi päätellä, ettei Z39.50 ole helppo standardi toteutettavaksi. Haasteellinen se on myös kirjaston näkökulmasta. Kirjaston pitää dokumentoida muun muassa Z39.50-palvelinrajapintansa. Mallia voi ottaa Kongressin kirjastosta, joka on tehnyt [rajapinnan kuvauksen](#) esimerkillisesti. Kuvauksesta selviää kaikki tarvittava, kuten sallitut hakutermit.

Z39.50-palvelimilta voi kerätä konfiguraatietietoja myös automaattisesti. Index Data on koonnut [IRSpy](#)-ohjelmallaan [tietämyskannan](#) tarjolla olevista Z39.50- ja SRU/SRW-palvelimista. Samankaltainen BookWhere-ohjelmiston tietämyskanta on helpottanut vuosien ajan Linnea-kirjastojen luetteloiden tietuepöytäkirjoitusta etätietokannoista.

Index Datan kokoamien tietojen

perusteella erilaisia Z39.50-palvelinohjelmistoja on noin kymmenen. Arviointia vaikeuttaa se, että moni ohjelmistotalo käyttää Index Datan avoimen lähdekoodin YAZ-sovellusta joko sellaisenaan tai muokattuna.

Index Data on koonnut haravoimistaan tiedoista 1831 Z39.50-tietokantaa kattavan tilaston, joka antaa melko surullisen kuvan Z39.50-tiedonhaun tilanteesta, sillä semanttisessa ja teknisessä yhteismitallisuudessa on paljon parantamisen varaa. Reilut 20% kohdetietokannoista tuke edes Search ja Present-palveluja. Toisin sanoen niistä ei voi tehdä hakuja. Indeksien selauspalvelu, Scan, on toteutettu kahdessa palvelimessa kolmesta, tulosjoukon lajittelu (Sort) kahdessa viidestä. Yleisimmin tuettu hakuterminä on Nimeke, mutta sitäkin tukee vain vajaat 60% kaikista palvelimista. ISBN- ja ISSN-tunnushaut, joita voisi kuvitella jokaisen palvelimen tukevan, on mahdollinen vain puolessa niistä.

Metahaku

Index Datan kokoama tilasto osoittaa metahakusovelluksien rakentamisen haasteet. Koska ne eivät voineet ennakkoon rajata asiakkaiden käyttämiä hakutermejä ja kohdejärjestelmiä, saman haun tulosjoukko saattoi kohdetietokannan mukaan olla tyhjä, vajaa, liian suuri, epärelevantti ja hyvällä onnella täysin validi. Jos kohdejärjestelmiä oli enemmän kuin yksi ja yksittäiset tulosjoukot suuria, hakuun kuluva aika saattoi olla pitkä ja koko

haun tulosjoukko vaikea käsitellä.

Oleellinen osa metahakujärjestelmää eli portaalia on kohdetietokantojen hakuliittymien tiedot sisältävä tietämyskanta. Valitettavasti metahakujärjestelmiä kehittäneet yritykset pitivät ohjelmistojensa tietämyskantoja liikesalaisuuksina, joiden sisältämiin metatietoihin ei ollut oikeuksia edes portaalaja ostaneilla organisaatioilla.

Jos metahaun kohteena olevan tietokannan hakuliittymä oli epästandardi, sen käyttöä koskeva tieto piti kerätä ”screen scraping”-menetelmällä. Ohjelman vastaukset ihmisen tekemiin hakuihin kirjattiin, mikä mahdollisti metahaun, mutta vain siksi kunnes käyttöliittymää muutettiin. On ymmärrettävää, että tätä työläästi kerättävää ja ylläpidettävää tietoa ei haluttu jakaa. Mutta jos käyttöliittymä oli standardisoitu, sen kuvaus oli mahdollista kerätä ja pitää yllä ohjelmallisesti, ja jakaa sovelluskehittäjien kesken. Yhteistyö näiden tietojen ylläpidossa olisi ollut eduksi sekä järjestelmien kehittäjille että niiden käyttäjille.

Vuonna 2003 NISO käynnisti [Metaresearch Initiativen](#), jonka tavoitteena oli metahakuun liittyvien standardien kehittäminen esimerkiksi tietämyskantojen sisältämän tiedon standardisoitua vaihtoa varten. Jotkin hankkeeseen osallistuneet yritykset suhtautuivat sii-

hen kielteisesti, ja projekti hermostutti NISO:n johtokunnassa edustettuina olleet ohjelmistotalot ja kustantajat. Oli tuskin vain sattumaa, että hanketta tukenut NISO:n Chief Executive Officer sai potkut, eikä hankkeen tuottamia standardiluonnoksia saatu valmiiksi.

Järjestelmätoimittajille metahakujärjestelmä tarjosi mahdollisuuden myydä kirjastoille uusi tuote. Se oli tarpeen, koska 2000-luvulla teollistuneissa maissa useimmilla kirjastoilla oli jo integroitu kirjastojärjestelmä. Useimmille ohjel-

mistotaloille lopputulos oli pettymys, koska Ex Libris puhdisti pöydän. Sen yhteistyö Index Datan kanssa Z39.50-kehityksessä tarjosi vahvan teknisen perustan MetaLib-sovellukselle. Eikä siinä kaikki: Ex Libris oli lisensoinut belgialaisen Herbert Van de Sompelin kehittämän SFX-linkityspalvelun, joka toi yritykselle ainutkertaisen kilpailuedun. MetaLib/SFX-yhdistelmä oli kilpailijoihin nähden ylivoimainen tuote, mutta MetaLibin menestys jäi silti vain muutaman vuoden mittaiseksi, koska portaalien hakutulokset eivät tyydyttäneet käyttäjiä. SFX sitä vastoin elää tätä kirjoitettaessa edelleen; sillä on [Wikipedian mukaan 2 400 käyttäjäkirjastoa](#).

Linnea-verkossa virtuaalisten yhteisluetteloiden rooli jäi mar-

**Valitettavasti
metahakujärjestelmiä
kehittäneet yritykset
pitivät ohjelmistojensa
tietämyskantoja
liikesalaisuuksina.**

**Riippuvuus
kaupallisten
yritysten fyysisistä
yhteisluetteloista
rajoittaa valinnan
mahdollisuuksia.**

ginaaliseksi. Maakuntakirjastojen MANDA-yhteisluettelo korvattiin [Nelli-portaalilla](#). Se oli Kansalliskirjaston 2004 perustettu palvelu, jossa käytettiin MetaLib- ja SFX-tuotteita. Kansalliskirjasto esitti 2007 yleisten kirjastojen neuvostolle tuolloin jo neljä vuotta jäädytettynä olleen MANDA-tietokannan korvaamista Nelli-portaalia hyödyntävällä Virtuaali-MANDA:lla, jonka monihaku kohdistuisi kaikkiin maakuntakirjastokoelmiin ja Varastokirjastoon. Yleisten kirjastojen neuvosto hyväksyi Kansalliskirjaston ehdotuksen ja katsoi, että MANDA-yhteisluettelo voidaan lakauttaa.

Virtuaali-MANDA:n elinkaari jäi alkuperäistä fyysistä yhteisluetteloakin lyhyemmäksi. Kansalliskirjaston kirjastoverkkopalveluiden ylläpitämä Finna-asiakasliittymä korvasi sen vuonna 2016. Metahakusovellusten tähdenlento oli silloin jo kaukana takana. Ex Libris tarjosi asiakkaille jo 2011 MetaLib+ tuotetta, johon sisältyi Primo-asiakasliittymä, fyysinen yhteisluettelo Primo Central Index (PCI) ja

niitä tukeva metahakutoiminnallisuus:

MetaLib ei ole enää vuosiin kuulunut Ex Libriksen tuotevalikoimaan, eikä paluuta portaaleihin ole näköpiirissä. Ne eivät toimineet asiakaskäytössä riittävän hyvin, ja ongelmia oli käytännössä mahdotonta ratkoa, koska ne eivät liittyneet vain portaalisovel-
luksiin sinänsä, vaan myös portaalien tietämyskantojen laatuongelmiin ja kohdetietokantojen riittämättömään yhteismitallisuuteen. Mutta vaikka portaaleilla ei näillä näkymin ole tulevaisuutta, tarvitsemme rakenteista ja standardisoitua tietoa Internet-tietokannoista.

Riippuvuus kaupallisten yritysten fyysisistä yhteisluetteloista rajoittaa valinnan mahdollisuuksia esimerkiksi ohjelmistovalinnoissa. Jos joudumme käyttämään PCI:n tai WorldCatin sijaan alkuperäisiä lähteitä kuten Kongressin kirjaston tietokantaa, näiden etätietokantojen hakuliittymiä koskeva metatieto on tarpeen. Tätä kirjoitettaessa sitä tarjoaa luetteloiden käyttämisen BookWhere-ohjelmiston tietämyskanta, tulevaisuudessa toivottavasti jokin standardisoidumpi ja avoimempi ratkaisu.

Kirjoittaja:

JUHA HAKALA

jehakala@gmail.com

