

Luetteloinnin tulevaisuus On the Record – raportin valossa

**Linda-päivä 2008-05-06
Juha Hakala & Eeva Murtomaa,
Kansalliskirjasto**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Sisältö

- **On the Record –raportti: tausta ja vastaanotto**
- **Raportin suositukset ja niiden toteuttamisen haasteet**
- **Suosittelujen soveltaminen Suomessa**
 - Luetteloinnin tehostaminen
 - Sanastot, ontologiat ja auktoriteettityö

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

On the Record -raportti

•Marraskuussa 2006 Kongressin kirjasto asetti ulkoisen asiantuntijatyöryhmän pohtimaan luetteloinnin tulevaisuutta; tarkempi toimenkuva oli:

- Present findings on how bibliographic control can effectively support management of and access to library materials in the evolving information and technology environment
- Recommend ways in which the library community can collectively move toward achieving this vision
- Advise the Library of Congress on its role and priorities

•Ryhmässä edustettuna kirjastojen lisäksi mm. Google, Microsoft ja OCLC; LC:stä vain yksi henkilö

- <http://www.loc.gov/bibliographic-future/>

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Lähtökohdat

•Bibliografinen valvonta on ymmärrettävä perinteistä luettelointia laajemmin niin, että se kattaa kaiken kirjastojen kautta haettavan aineiston, erilaiset käyttäjäyhteisöt ja eri paikat, joista tietoa etsitään

•Bibliografinen universumi laajenee kirjastojen, kustantajien ja tietokantatuottajien ohi tekijöihin, välittäjiin, kirjakauppoihin ja käyttäjäyhteisöihin, perinteisistä sektori- ja maantieteellisistä rajoista piittaamatta

•Kongressin kirjaston tehtävät muuttuvat toimintaympäristön mukana; joistakin vanhoista (kansalliskirjasto)vastuista voidaan luopua samalla kun otetaan/määrätään uusia

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Visio

- **”The Working Group envisions a future for bibliographic control that will be collaborative, decentralized, international in scope, and Web-based.**
- **The realization of this future will occur in cooperation with private sector and with the active collaboration of library users.**
- **Data will be collected from different sources, change will happen quickly, and bibliographic control will be dynamic, not static”.**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Raportin vastaanotosta

- **Työryhmän näkemykset ovat saaneet voimakasta kritiikkiä osakseen Kongressin kirjaston sisältä:**
 - Thomas J. Mann: On the Record but off the Track (<http://www.guild2910.org/WorkingGrpResponse2008.pdf>)
- **”The report shows a naive faith in interactive Web applications and pushes librarians to spend more time with technology at the expense of developing and managing catalogs for scholarship.”**
- **”Managing research materials requires high-level professional thinking, not just computer algorithms and nonstandardized tags contributed by anyone with access to the Internet.”**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Raportin vastaanotosta (2)

- **Keskustelu jatkuu vilkkaana, Thomas J. Mannin vastineelle kirjoitetaan parhaillaan vastineita**
- **Pohjimmiltaan on kyse siitä, jatkammeko luettelointia vanhaan malliin, vai muutammeko kurssia enemmän tai vähemmän radikaalisti**
- **Toteutuessaan työryhmän suositukset olisivat merkittävä linjanmuutos**
 - Toteutumisen todennäköisyydestä ei tietoa; suositukset eivät sido LC:n käsiä

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksia: yleistä

1. Lisätään luetteloinnin tehokkuutta
2. Parannetaan erikoiskokoelmien käytettävyyttä
3. Kehitetään tekniikkaa (Web-pohjaista) tulevaisuutta silmällä pitäen
4. Kehitetään palveluita (Web-pohjaista) tulevaisuutta silmällä pitäen
5. Vahvistetaan kirjastoalan tutkimusta ja koulutusta
 - Jokaiselle suositukselle on listattu joukko konkreettisia esityksiä toivottuine tuloksineen, minkä lisäksi on arvioitu status quon seuraukset

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Luetteloinnin tehostaminen

- **Päällekkäistyön eliminointi**
 - Kustantajien tuottaman kuvailutiedon tehokkaampi hyväksikäyttö ONIXia hyödyntämällä
 - Käytetään hyväksi Internetissä olevia kuvailutietoja (esim. Amazon, Internet Movie Database)
 - Osoitetaan metatiedon arvo tiedonhaussa
- **Luettelointivastuun hajauttaminen**
- **Yhteistyö auktoriteettitietojen tallennuksessa ja ylläpidossa**
 - Kansalliskirjastojen roolia korostetaan, entä muut tahot, kuten tekijänoikeusjärjestöt?

Haasteita: päällekkäistyön eliminointi

1. **Kustantajayhteistyö**
 - ONIXia käytetään vain vaihtoformaattina, eivätkä kustantajien omat järjestelmät (luultavasti) ole yhteismitallisia sen kanssa
 - Kustantajilla ei ole (yhteisiä) luettelointisääntöjä, joten datan laadusta ei ole takeita vaikka sen syntaksi noudattaisi ONIXia
2. **Internet-metadatan käyttö**
 - Internetissä ei ole yhteisiä luettelointisääntöjä eikä metadataformaattia; muualla tuotettujen kuvailutietojen soveltaminen on tehtävä järjestelmäkohtaisesti (esim. julkaisuarkistot) ja varoen, ”omia” metatietoja sotkematta
3. **Metadatan arvon osoittaminen tiedonhaussa**
 - Sekä informaatiotutkijat että hakukoneiden rakentajat skeptisiä

Erikoiskokoelmien käytettävyyden parantaminen

1. Priorisoidaan erikoiskokoelmien kuvailu
2. Tehostetaan näiden kokoelmien kuvailua esim. soveltamalla eri kuvailutasoja
3. Liitetään erikoiskokoelmat samoihin hakujärjestelmiin muiden aineistojen kanssa
4. Lisätään digitointia käytön helpottamiseksi
5. Helpotetaan erikoiskokoelmia koskevan bibliografisen ja auktoriteettitiedon kopioluettelointia

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Erikoiskokoelmien haasteita

- Miten valitaan kuvailtavat kokoelmat?
- Miten organisoida kokoelmatason kuvailutietojen tehokas (asiakas)käyttö?
- Miten saadaan kaikki metadata mielekkäästi haettavaksi samaan järjestelmään?
 - Kun kaikki muistiorganisaatiot ovat mukana, haettavaksi voi tulla periaatteessa mitä vain
 - Kaikkea kirjastoaineistoakaan ei vielä saa yhdestä tuutista
- Mitä muuta jätetään tekemättä, jotta erikoiskokoelmat voidaan digitoida ja/tai kuvailla?

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Tekniikan kehittäminen

- **Web järjestelmien infrastruktuuriksi**
 - Kehitetään joustavampi ja helpommin laajennettava metadata/vaihtoformaatti MARCin ja ISO2709:n tilalle
 - Kirjastojen ja kirja-alan yhteistyö
 - Integroidaan kirjastoalan standardit Web-ympäristöön
 - (Meta)datan pitää olla koneluettavaa ja –ymmärrettävää
 - Sovelletaan Web-pohjaisia (URI)-tunnisteita metadataformaattien tietoelementeille ja sanastojen termeille
 - Tunnisteilla luodaan linkki kentästä / termistä sen määritykseen tai metadataformaatin skeemaan

Tekniikan kehittäminen (2)

- **Standardit**
 - Kehitetään moderni perusta bibliografiselle tietojärjestelmälle, joka sisältää uuden vaihtoformaatin (nyt ISO 2709), luettelointiformaatin (MARC), luettelointisäännöt (AACR), sanastot, auktoriteetit sekä FRBR-mallin käyttöönoton
 - Ovatko MARCXML, MODS ja RDA OK?
 - Tehostetaan standardien kehittämisprosessia
 - Metadatastandardien tulee edistää kuvailutietojen automaattista käsittelyä
 - Parannetaan tiedonkulkua käyttäjiltä standardien kehittäjille
 - Keskeytetään RDA-luettelointisääntöjen kehittäminen kunnes RDA:n edut on vakuuttavasti osoitettu

Tekniikan haasteita

- Tarvitsemme kokonaisen läjän uusia standardeja; kuka ne kehittää, ja miten urakka määritellään ja priorisoidaan?**

- NISO:lla & ISO TC 46:lla on jo nyt täysi agenda
- Standardoinnissa on mukana eri tahoja joilla on osin ristiriitaiset ja päällekkäiset intressit; kirjastojen aktiivisuus riittämätöntä
 - Sovelletaanko ODF:ää, OOXML:ää vai PDF:ää, jotka ovat kaikki ISO-standardeja?
 - Valitaanko URN, DOI, PURL vai jokin muu URI?

- Pystymmekö kehittämään entistä yksinkertaisemmat kuvailusäännöt nykyistä monimuotoisemmille aineistoille?**

- Mitä metadataformaattia / -formaatteja meidän on tuettava MARCin ohella tai asemesta?**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Miksi ISO2709 + MARC on työryhmän mielestä vaihdettava?

- ISO 2709 + MARC ovat 40 vuotta vanhaa tekniikkaa eikä niitä käytetä muualla kuin kirjastoissa**

- Näyttöluetteloissa oleva kuvailutieto ei "näy" Web-sovelluksille, vaikka OPACissa olisikin Web-asiakasliittymä

- Formaatin on erotettava toisistaan ammattilaisten, amatöörien ja automaattien (ohjelmistojen) tuottama metadata**

- Annotointi (arvostelut, kommentit), käyttötiedot

- Formaatin on oltava joustava siten, että metadataa voidaan hyödyntää valikoiden käyttöympäristön tarpeiden mukaisesti**

- Muutosvaatimukset ovat niin suuret, ettei ISO 2709 + MARC enää josta riittävästi**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Käytännön esimerkki a'la uusi Libris

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix libris: <http://libris.kb.se/experimental/> .
@prefix annotea: <http://www.w3.org/2000/10/annotation-ns#> .
# RDF in Turtle/N3 created for the bibliographic record 5059476

<http://libris.kb.se/bib/5059476>
  # short bibliographic description
  dc:title "The difference engine" ;
  dc:creator "Gibson, William, 1948-" ;
  dc:identifier <URN:ISBN:0-575-04762-3> ;
  # record truncated for brevity

  # links to authors with authority records
  dc:creator <http://libris.kb.se/auth/220040> ;

  # links to subjects with authority records
  dc:subject <http://libris.kb.se/auth/12345> ;

  # links to other editions of the same work
  libris:frbr_related <http://libris.kb.se/bib/5060570> ;

  # user annotations
  annotea:hasAnnotation <http://libris.kb.se/annotation/123>
```

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Annotointi: Libriksen ratkaisu

•The Annotea Annotation schema [AnnotationNS] defines properties for:

- identifying the document being annotated,
- a specific context within that document to which the body of the annotation refers,
- the author of the annotation,
- and more.

•Annotea clients can provide a variety of presentations of these annotations. As the Annotea framework is designed on top of RDF, it can be easily extended to other shared metadata applications.

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Palveluiden kehittäminen

- **Suunnitellaan palvelut ja järjestelmät myös tulevaisuuden käyttäjille (sekä ihmisille että ohjelmistoille)**
 - Linkitetään kirjastoluetteloihin relevanttia ulkopuolista informaatiota
 - Libris-käyttäjien ykköstoive; kaivattiin linkejä kirjallisuusarvosteluihin, kansikuviin, Wikipediaan etc.
 - Sovelletaan loppukäyttäjien tuottamaa kuvailutietoa (sotkematta sitä muuhun metadataan)
 - Selvitetään mahdollisuudet metadatan automaattiseen tuottamiseen

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Palveluiden kehittäminen (2)

- **FRBR-mallin aito implementointi**
 - Järjestelmien pitäisi sallia teostason tietojen tallennus ja vaihto perinteisen bibliografisen tiedon tapaan
- **Asiasanojen käytön tehostaminen**
 - Sanastojen väliset linkit
 - Sanastojen maksuttomuus (konelukuisessa muodossa)
 - Huom! Ei suositusta LCSH:n ontologisoimisesta, mm. koska LCSH soveltuisi tähän rakenteensa vuoksi huonosti
- **Automaattisen indeksoinnin edellytysten selvittäminen**
 - Koneellisin menetelmin saatuja asiasanoja ja luokituksia pitäisi voida hyödyntää perinteisen sisällönkuvailun tapaan

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Palveluiden kehittämisen haasteita

- **Perinteisten kirjastojärjestelmien FRBR-muunnos on osoittautunut teknisesti vaikeaksi, useimmat toteutukset tehty käyttöliittymän tasolla**

- Poikkeuksia WorldCat & VTLS:n Virtua, jossa kaikki bibliografinen data tallennetaan sisäisesti XML-muodossa

- **Sanastojen (ja ontologioiden) väliset käsitetason linkit on luotava käsityönä, mikä vie aikaa ja rahaa**

- **Automaattista indeksointia on tutkittu noin 30 vuotta, laihahkoin tuloksin**

- **Muiden tuottaman metadatan laadusta ei ole takeita**

- Luotettavien yhteistyökumppaneiden valinta tärkeää

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Tutkimuksen ja koulutuksen vahvistaminen

- **Bibliografisen valvonnan tehokkuuden osoittaminen**

- Kehitetään mittarit luettelointiin pohjautuvien palveluiden tehokkuuden mittaamiseen
- Tuetaan alan tutkimusta

- **Kirjastoalan koulutuksen kehittäminen**

- Ammattitaitoisista luetteloojista ja sisällönkuvailijoista on jo nyt pulaa, ja tilanne tulee pahenemaan
- Tarvitaan yhteistyötä informaatiotutkimuksen laitosten kanssa, täydennyskoulutusta sekä laadukkaiden oppimateriaalien jakelua verkossa

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Johtopäätelmiä

•Suositukset osin suht' radikaaleja

- Paljon kohua herätti mm. esitys RDA-kehitystyön jäädyttämisestä; minusta vahva Web 2.0 –painotus on suurempi politiikkamuutos
- Thomas J. Mannin vastareaktio on ymmärrettävä, mutta yliampuva ja perustuu osin Web 2.0 –pohjaisten palvelujen luonteen ja toteutustapojen suurpiirteiseen sisäistämiseen

•Toiveiden tynnyri on iso; osa esityksistä saataneen toteutetuksi, mutta tuskin kaikkia

- Esteinä sekä raha että tekniset ongelmat (standardien puute)

•Esitetyt tavoitteet ovat relevantteja Suomessakin, mutta meillä prioriteetit voivat olla erilaiset

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksen soveltaminen Suomessa: luetteloinnin tehostaminen

•Suomessa on kolme ”virallista” metadataformaattia:

- MARC21 mahdollisimman vähin lisämaustein
- Dublin Core kansallisin ominaispiirtein (perusversion lisäksi mm. asiakirjoille, opinnäytteille ja kokoelmille)
- Kirja-ONIX suomalaisen kirja-alan tarpeisiin räätälöitynä

•Käyttäjätukea tarvitaan lisää etenkin Dublin Corelle

•Muistiorganisaatioiden yhteistyötä varten tehty yhteinen Dublin Core –pohjainen metadatatamääritys KAMUT-hankkeessa

- Määrityksen toteuttava ohjelma (”asiakasliittymäsovellus”) on periaatteessa helposti toteutettavissa, käytäntö on kiemuraisempi koska KAMUT-periaatteiden käytännön toteutus lienee kesken

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksen soveltaminen Suomessa: luetteloinnin tehostaminen (2)

- **Kuvailutiedot voidaan jo nyt tallentaa ja jakaa eri formateissa, riippuen vastaanottajan tarpeista**

- Oleellinen pullonkaula kuvailun taso; esim. kustantajat ja tutkijat tarvitsevat lisää opastusta

- **Kansallisbibliografialuettelointia on hajautettu, mutta parantamiseen varaa on edelleen (mm. kustantajat ja korkeakoulut olisi hyvä saada laajemmin mukaan)**

- **Kopioiluettelointi on tehokasta; USA:n kaltaista riippuvuutta yhdestä palvelusta (WorldCat) ei ole eikä sellaista pidä rakentaakaan**

- **Keinot ulkopuolisen metadatan linkitykseen vähissä; ongelma helpottanee uusien järjestelmien myötä**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksen soveltaminen Suomessa: luetteloinnin tehostaminen (3)

- **Auktoriteettitietokannan puuttuminen on Suomen todellinen Akilleen kantapää**

- Nykytilanteesta kärsivät sekä tiedonhakijat että esim. yhteistyö semanttisen Webin tutkijoiden kanssa
- Erillinen a-tietokanta pitäisi saada aikaan, vaikka kirjastojen ja tekijänoikeusjärjestöjen yhteistyönä
- Kun tietokanta on luotu, sen käytöstä pitäisi saada aikaan järkevä (lue: riittävän joustava) periaatesopimus

- **Yhdysvaltoihin verrattuna myös kokoelmien kuvailu on vasta lähtökuopissa, vaikka Kokoelmakartta onkin lupaava alku asian organisoimiseksi**

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksen soveltaminen Suomessa: sanastot ja ontologiat

- Yleinen suomalainen asiasanasto (YSA) on saavuttanut Suomessa samanlaisen de facto – standardin aseman kuin LCSH USA:ssa, ja:
- LCSH:sta ei ole edes suunnitteilla yleistä ontologiaa, YSA:sta se on jo tehty, ks. <http://www.yso.fi/>
 - <http://www.yso.fi/onto/yso/p17153> (YSO-termin “aakkostus” URI-tunniste FinnONTO:n malliin)
- YSA:n ontologisointi on tehnyt siitä Suomen idullaan olevan semanttisen Webin ytimen; järjestelmän soveltaminen ja ylläpito kiinnostaa monia
 - Mikä on Kansalliskirjaston ja muiden kirjastojen rooli jatkossa YSO:n ja erikoisalojen ontologioiden ylläpidossa?

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosituksen soveltaminen Suomessa: sanastot ja ontologiat (2)

- Suomessa lähtötilanne on USA:an verrattuna varsin hyvä, mutta:
 - YSA/YSO on yhteismitallistettava
 - Allärs & Musa/Cilla pitää ontologisoida
 - YSA/YSO on käännettävä englanniksi ja linkitettävä relevantteihin englanninkielisiin sanastoihin
- YSO:n ja Allso:n ylläpidosta pitää tehdä ratkaisu (kuka vastaa, millä resursseilla)
- Ontologioiden ylläpitoon ja käyttöön tarvittavien sovellusten tuki pitää taata
 - Tällä hetkellä Semantic Web 2.0 –hankkeen vastuulla, mutta mikä on tilanne kun projekti päättyy?

KANSALLISKIRJASTO



NATIONALBIBLIOTEKET

Suosittelujen soveltaminen Suomessa: tutkimus ja koulutus

- Ammattitaitoisista luetteloijista ja sisällönkuvailijoista tulee pulaa Suomessakin; täydennys- ja sisäinen koulutus lienevät Suomessa tehokkaimmat tavat vähentää henkilöstöpulaa
- Suomessa on korkeatasoista alamme tutkimusta (esim. semanttisen laskennan tutkimusryhmässä Otaniemessä sekä TaY:n informaatiotieteiden tiedekunnassa) jota kirjastojen on käytettävä aktiivisemmin hyväksi järjestelmäkehityksessä
 - Yhteisiä tutkimushankkeita tarvitaan lisää