

Kansainvälisestä formaattikehityksestä –  
LoC:n Bibframe- ja NISO:n Roadmap-hankkeet

2013-03-21  
Juha Hakala  
Kansalliskirjasto

## Sisältö

- MARC-formaatin ja erityisesti MARC 21:n ongelmista ja niiden ratkaisemisen periaatteista
- Kongressin kirjaston BIBFRAME-hanke
- NISO:n Tiekarttaprojekti
- Hankkeiden mahdollisista vaikutuksista Suomessa

## Taustaa

- MARC:in kuolemaa on peräänkuulutettu jo ainakin 2000-luvun alkupuolelta lähtien
  - Roy Tennant: MARC must die (2002),  
<http://www.libraryjournal.com/article/CA250046.html>
- Mutta Tennant ei puhu vain formaatista:
  - "When I refer to MARC,... I am conflating several interrelated things. There are the MARC syntax, the MARC data elements, and the AACR"
  - "MARC and AACR2 are focused on capturing the paper catalog card in computer form"

## Sitten vuoden 2002 moni asia on jo muuttunut:

- AACR2 –säännöt korvataan RDA:lla
  - MARC21:een on lisätty RDA:n edellyttämät tietoelementit - edellytys sille että RDA saadaan tuotantoon esim. USA:ssa ja Englannissa kuluvan vuoden aikana
- Tennantin kolmesta ongelmakohdasta kaksi (kuvailun säännöt ja formaatin tietoelementit) on korjattu, mutta syntaksin modernisointi puuttuu

## Mikä MARC:in syntaksissa on vialla?

- Omintakeisuus
  - Metatietomme ovat vankasti siilossa, ja siilo on pieni: kirjastojärjestelmätoimittajia on vähän
- Korttiluettelon henkinen perintö näkyy tietomallissa
  - RDA-tietoelementit on lisätty, mutta FRBR-tuen toteuttaminen näyttää olevan nykyisissä järjestelmissä iso kynnys
- XML-toteutus (MARCXML) kömpelö
- MARC 21 –spesifit ongelmat, kuten
  - Suku- ja etunimi samassa kentässä
  - Välimerkit datassa / tietoelementtien erottimina

## Mitä MARCin tilalle?

- Uuden formaatin pitäisi ratkaista joukko hyvin eritasoisia ja erilaisia ongelmia
  - Toiveiden tynnyriä on koottu vuosikausia; odotukset voivat olla epärealistisen korkeat
- Tavoite voidaan formuloida eri tavoin; tässä Roy Tennantin näkemys vuodelta 2002:
  - "we must assimilate MARC into a broader, richer, more diverse set of tools, standards, and protocols"
  - -> MARCin kuolemasta sen reinkarnaatioon – tavoitteeksi kuvailevien metatietojen esittäminen semanttisen Webin vaihtomuodossa eli RDF-muodossa ISO 2709:n sijaan

# Resource Description Framework (RDF)

- *RDF is a standard model for data interchange on the Web.*
- *RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.*
- *RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”).*
- *Using this simple model, RDF allows structured and semi-structured data to be mixed, exposed, and shared across different applications.*
- 1000 \$:n kysymys: miten yhdistämme RDA:n ja RDF:n?

## RDA:sta ja RDF:stä

- Sekä RDA:sta että RDF:stä puhutaan usein kovin teoreettisesti; voi olla vaikeaa nähdä miten teoriasta päästään käytäntöön
  - Sekä RDA:han että RDF:n on suhtauduttu varauksella, ja niiden käytännön toteuttaminen on vienyt aikaa
- MARC-formaatin seuraajaa kehitettäessä teorioista on pystyttävä johtamaan toimivaa käytäntöä; tämä ei välttämättä ole helppoa
  - Uuden formaatin on tarjottava merkittäviä käytännön parannuksia vanhaan verrattuna; pelkkä puhe siilojen rikkomisesta ei vielä avaa kukkaron nyörejä



# BIBLIOGRAPHIC FRAMEWORK TRANSITION INITIATIVE

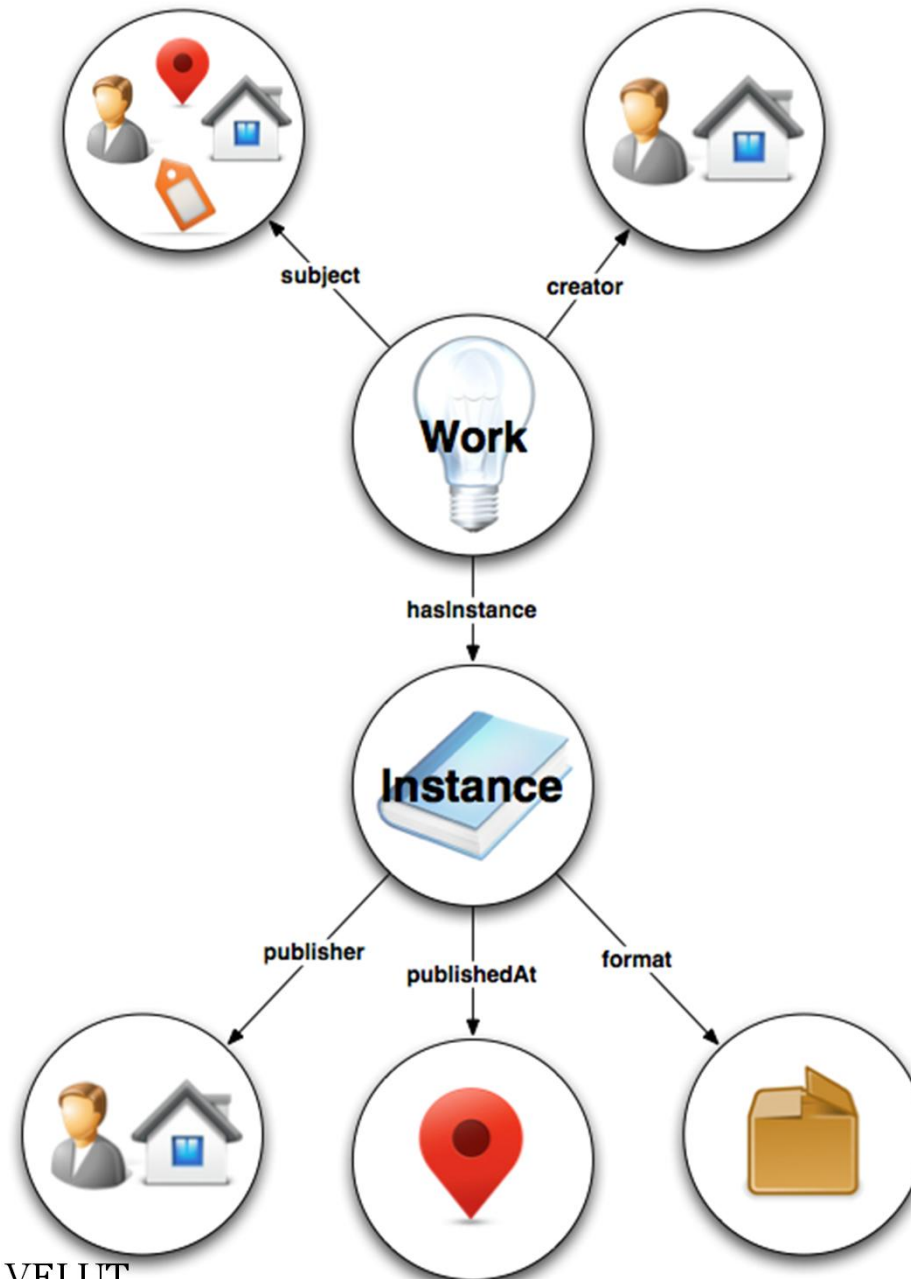
- Kongressin kirjaston hanke, 2011-
  - Toistaiseksi mukana vain pieni joukko edelläkävijäkirjastoja, mutta ryhmää on tarkoitus kasvattaa työn edetessä
    - Euroopasta British Library ja Saksan kansalliskirjasto
  - Projekti on rajattu MARC 21:een
  - Formaali standardointi on tarkoitus aloittaa vasta sen jälkeen kun formaatti on valmis (vrt. MARC-formaatin kehittäminen)
- Projektista vastaa Network Development and Standards Office, jolla on myös MARC 21 –vastuu (ja lukuisia muita formaattivastuita)
  - Myös MARC 21:stä kehitetään edelleen; sitä ei ole tarkoitus “jäädyttää”

## BIBFRAME ja FRBR

- Uuden formaatin pitää olla aidosti FRBR-yhteensopiva (jota MARC ei ole)
  - BIBFRAME ei käytä FRBR-terminologiaa, koska malli otettaneen käyttöön vain kirjastoissa
    - Neutraalilla terminologialla houkutellaan muita RDA:n soveltajia BIBFRAME-käyttäjiksi
  - "[BIBFRAME] is designed to integrate with and engage in the wider information community while also serving the specific needs of its maintenance community"

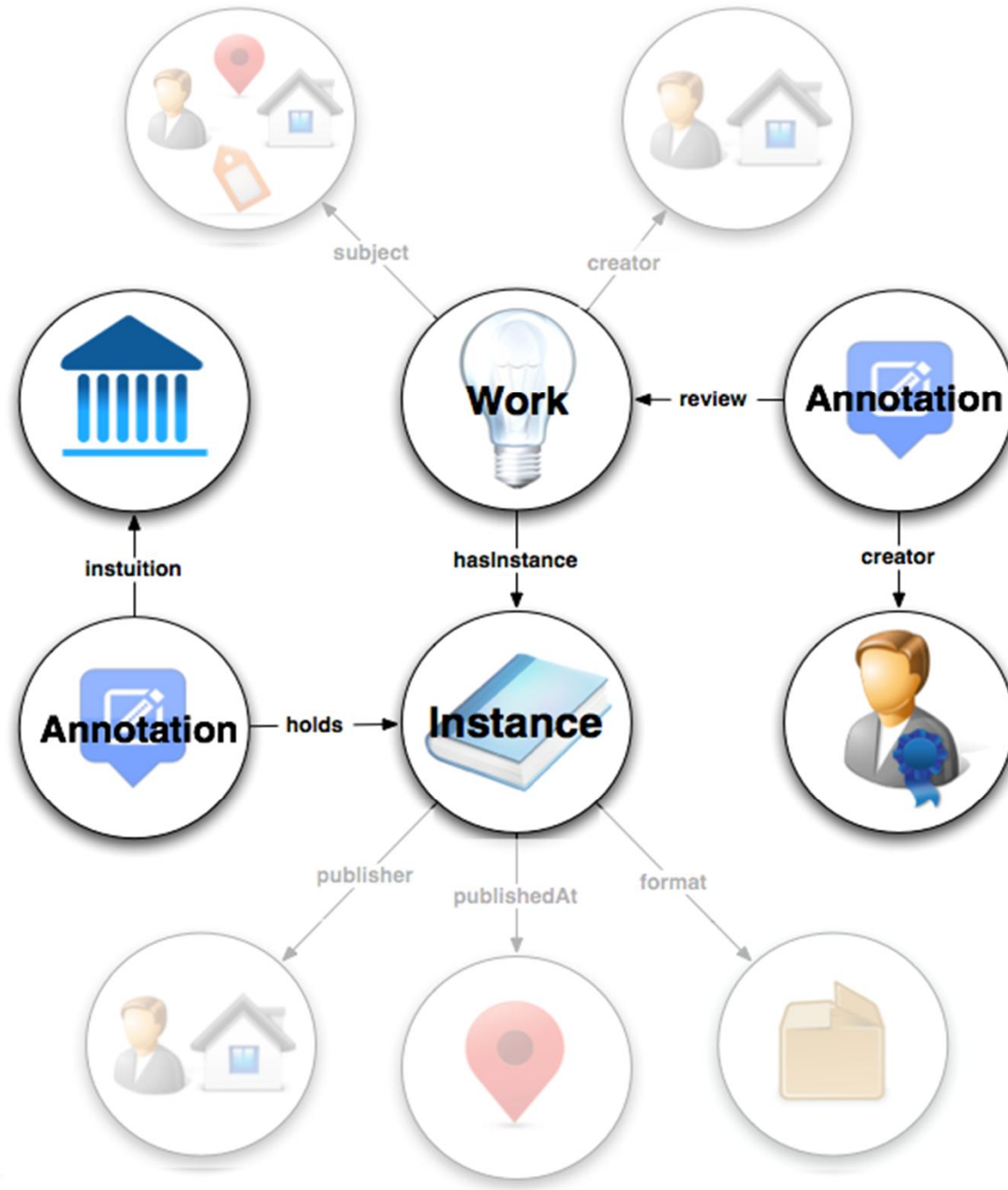
## BIBFRAME-tietomalli

- Teos (Work)
  - Suunnilleen sama kuin FRBR:n teos + ekspressio
- Manifestaatio (Instance)
  - Teoksen yksittäinen aineellinen ilmentymä
- Auktoriteetti (Authority)
  - Henkilö, organisaatio, paikka, asiasana tms. jolla on selkeästi määritelty suhde teokseen tai sen manifestaatioon
- Liitännäiset (Annotation)
  - Esim. kirja-arvostelu, kirjan kansikuva
- Sekä malli että sen suomennos ovat luonnoksia



## Liitännäiset

- Teoksen metatietoihin voidaan linkittää arvosteluja, sisällysluetteloita, otteita yms.
- Manifestaation tietoja voidaan rikastaa kansikuvalla, varastotiedoilla, Web-osoitteella yms.
- Erillisiin auktoriteettitietueisiin tallennetaan lisätietoja teoksesta itsestään, teoksen tai esim. sen arvostelun tekijöistä, asiasanoista, jne.
  - RDF:n edellyttämä linkitys muuanne semanttiseen Webiin rikastaa erityisesti auktoriteettitietoja



## Tavoitteita

1. Teoksen ja sen eri manifestaatioiden metatietojen selkeä erottaminen toisistaan
2. Toimijoiden tunnistaminen
  - Toimijoiden nimille tarvitaan pysyvät tunnisteet (ISNI, ORCID) jotka helpottavat identifiointia ja pysyvien linkkien rakentamista
3. Roolien selkeä määrittely
  - A:n arvostelu B:n teoksen X suomenkielisestä käännöksestä, jonka on tehnyt C

## MARC 21 ja BIBFRAME: linjauksia

- MARC 21 –tietue pitää voida konvertoida häviöttömästi BIBFRAME-tietueeksi, ja
- MARC 21 –tietoelementit pitää voida konvertoida häviöttömästi BIBFRAME-tietueesta takaisin MARC 21-tietueeksi
- Välimerkeistä datassa hankkiudutaan eroon
  - Myös silloin kun välimerkit erottavat tietoelementtejä (esim. sukunimi, etunimi)
- -> BIBFRAME-formaatista tulee MARC 21:stä hienojakoisempi (esim. suku- ja etunimelle omat tietoelementit)



# MARC 21 ja BIBFRAME: avoimia kysymyksiä

- Osakohteiden käsittelystä ei ole vielä puhuttu
  - Oletan että osakohteita tuetaan jatkossakin, mutta (XML/RDF-pohjainen) tekninen ratkaisu voi poiketa siitä mihin olemme tottuneet
  - Tasojen määrää ei ole pakko rajata kahteen
- BIBFRAME-formaattien määrästä ei ole päätöstä
  - Mahdollisesti vain yksi (joka kattaisi kaikki MARC 21 -formaattit)
- Hallinnollisesta metadatatista ei ole päätöstä
  - Tietojen päällekkäisestä tallennuksesta eri formaatteihin (MARC 21, teknisen metadatan MIX, textMD etc., PREMIS) pitäisi päästä eroon, mutta toistaiseksi emme tiedä miten suuri osa hallinnollisen metadatan formaattien tietoelementeistä lisätään BIBFRAME:n RDF-skeemaan

## BIBFRAME:n XML/RDF-ratkaisu

- BIBFRAME ei tule käyttämään hyväksi olemassa olevia RDF-skeemoja, vaan sille rakennetaan oma skeema hallittavuuden maksimoimiseksi (esim. laajentaminen on helppoa)
- BIBFRAME-skeeman ensimmäinen luonnos on julkaistaan lähitulevaisuudessa osoitteessa <http://loc.gov/bibframe/vocab>
- Mappaukset BIBFRAME-skeemasta esim. FOAF-, SKOS- tai Dublin Coreen RDF-skeemoihin laaditaan myöhemmin
  - Kirjastojen pilottihankkeissa on yleensä käytetty valmiita skeemoja kuten FOAF:ia, koska se luo pohjan datan yhteiskäytölle
  - Mappausten avulla päästään samaan tulokseen, mutta "mutkan kautta"

## Linkitetty bibliografinen tieto

- RDF-syntaksissa tietueen kaikki tietoelementit ovat linkitettävissä / linkitetään muualle verkkoon; kun tietoja päivitetään, muutokset näkyvät kirjaston käyttäjille
  - Miten linkit valitaan, ja miten ne pidetään "hengissä"?
  - Kuka takaa linkityskohteiden laadun säilymisen?
- Kun metatietomme eivät enää ole "siilossa", emme pysty kontrolloimaan, mitä tietoa käyttäjämme saavat
  - "Vastuu siirtyy kuulijalle"
- BIBFRAME-formaattiin siirtymisen suurin käytännön vaikutus on siirtyminen linkitettyyn ja entistä avoimempaan dataan; tähän liittyvät poliittiset yms. linjaukset tarvitaan pian

## Täsmennettävää tietomallista

- Onko BIBFRAME:n teos yhtä kuin FRBR:n teos + ekspressio?
  - Jos on, miten BIBFRAME-tietueessa erotetaan nämä FRBR:n tasoa?
- Ovatko FRBR:n "manifestation" ja BIBFRAME:n "instance" sama asia (kuten suomennosluonnos antaa ymmärtää)?
  - Jos ei ole, miten osoitetaan milloin on kyse aidosta manifestaatiosta?
- Millainen suhde teoksella ja sen liitännäisillä (jotka usein ovat teoksia nekin) on?
- Millaisia ovat itsenäisten teosten väliset suhteet ja miten ne esitetään BIBFRAME:ssa?

## Avoimia kysymyksiä

- BIBFRAME:n uudet tietoelementit (jotka eivät ole konvertoitavissa takaisin MARC 21:een)?
- BIBFRAME:n MARC 21:stä hienojakoisemmat tietoelementit?
- Miten otetaan huomioon muut MARC-formaatit?
  - Muissa formaateissa on tietoelementtejä joita MARC 21:ssä ei ole ja jotka tulisi lisätä BIBFRAME:en
- Miten otetaan huomioon muut keskeiset metatietoformaatit kuten Dublin Core?
- Miten otetaan huomioon Kongressin kirjaston tukemat hallinnollisen metatiedon formaatit?

## Nykytilanne

- BIBFRAME-hanke on ollut käynnissä toista vuotta, mutta tekemistä on jäljellä paljon
  - Skeeman ohella moni muukin asia on vielä kesken, minkä takia projektiryhmää ole laajennettu pilottikirjastoista
  - Ei ole varmaa valmistuuko uusi formaatti vuoden 2013 lopussa, kuten on ollut tarkoitus
- Järjestelmätoimittajien kiinnostus BIBFRAME:en on ollut varsin vähäistä (poikkeuksina OCLC ja VTLS)
  - Meillä ei ole arvioita siitä, miten paljon työtä uuden formaatin toteuttaminen vaatii (eikä sellaista voi vielä tehdä)
- Kirjastot ovat ns. odottavalla kannalla

## BIBFRAME:n soveltamisesta

- MARC-pohjaisen järjestelmän modernisointi BIBFRAME:lle sopivaksi voi olla iso työ, riippuen siitä miten sovellus on rakennettu
  - Häviötön konversio MARC 21 – BIBFRAME - MARC 21 antaa toimittajille armonaikaa; siirtymäkausi voi olla pitkä
  - Mutta vaikka BIBFRAME – MARC 21 –konversiossa ei häviäisi dataa, avoimuus ja linkit voivat kärsiä
- Formaattivaihdoksen maksavat viime kädessä kirjastot; antaako BIBFRAME niin paljon uutta, että muutos on perusteltavissa?
  - Miten korkealle priorisoimme aidosti avoimen ja linkitetyn datan tuottamisen?

## BIBFRAME ja kirjastot

- Kirjastoille BIBFRAME on tärkeä uudistus; sen avulla pääsemme osaksi (semanttista) Webiä
  - Mutta kirjastojen tuottaman metatiedon semantiikan tulkinta voi edelleen olla haasteellista, johtuen tietojen rikkaudesta
- Järjestelmävalinnoissa tulisi ottaa huomioon, että saamme jatkossa metatietoa eri formaateissa eri tahoilta (esim. sosiaalista metadata Dublin Corena)
  - Pitäisikö uusien järjestelmien tukea useita metatietoformaatteja rinnan sekä tiedon tallennuksessa että haussa?
  - Finnassa käytettyyn VuFind-ohjelmiston versioon voidaan lisätä BIBFRAME-tuki kun formaatti on "riittävän vakaa"



## NISO:n tiekarttahanke

- BIBFRAME on rajattu MARC 21:een, se ei ota huomioon edes muita MARC-formaatteja, saati sitten muita metatietoformaatteja
  - BIBFRAME:en päästäkseen muiden MARC-formaattien käyttäjien on ensin konvertoitava metatietonsa MARC 21:een voidakseen soveltaa Kongressin kirjaston BIBFRAME-konversiotyökaluja
- NISO:n hankkeessa selvitetään muista formaateista johtuvat tarpeet laajentaa BIBFRAME:a (skeema; muut laajennukset)
- Tuloksena on (vain) raportti

## NISO-hankkeen haasteita

- Mitkä muut formaatit otetaan tutkittavaksi?
  - Pelkästään MARC-formaatteja on paljon, eikä ryhmään ole saatu edes niiden kaikkien edustajia
- Millä tavoin eroavuudet otetaan huomioon?
  - MARC 21:een konvertoitavissa olevat tietoelementit ovat "piece of cake" ja lisättävissä olevat luultavasti OK, mutta ristiriitatilanteiden käsittely on vaikeaa ja "maailmankuvaerojen" (ONIX, arkistojen ja museoiden formaatit) yhteensovittaminen tosi vaikeaa
- Kongressin kirjasto kutsuttiin mukaan, mutta se ei osallistune hankkeeseen ainakaan aktiivisesti
  - Otetaanko Tiekartta-hankkeen suosituksia huomioon BIBFRAME:n kehittämisessä, ja jos, niin miten?

## BIBFRAME ja Suomi

- Kansalliskirjasto osallistuu Tiekarttahankkeeseen, ja alustavasti meitä on pyydetty mukaan myös BIBFRAME:en siinä vaiheessa kun projektiryhmä laajenee
- Tavoitteenamme on mm. varmistaa se, että uusi formaatti on suomalaisten kirjastojen kannalta toimiva ja että Kongressin kirjaston BIBFRAME -konversio-ohjelmat toimivat kunnolla myös meidän MARC 21 –datallamme
  - FINMARC – BIBFRAME –konversiota emme tule tutkimaan / tukemaan
- Tulemme informoimaan kirjastoja näiden hankkeiden edistymisestä

## Lisätietoja

- BIBFRAME-hanke + keskustelulista:
  - <http://bibframe.org/>
  - <http://listserv.loc.gov/listarch/bibframe.html>
- Esimerkkejä BIBFRAME-tietueista ja konversio-ohjelma:
  - <http://bibframe.org/demos/>
  - <http://bibframe.org/tools/>
- Tiekarttahanke:  
<http://www.niso.org/topics/tl/BibliographicRoadmap/>