



Open > Link > GO

Avointa linkitettyä dataa kirjastomaailmassa

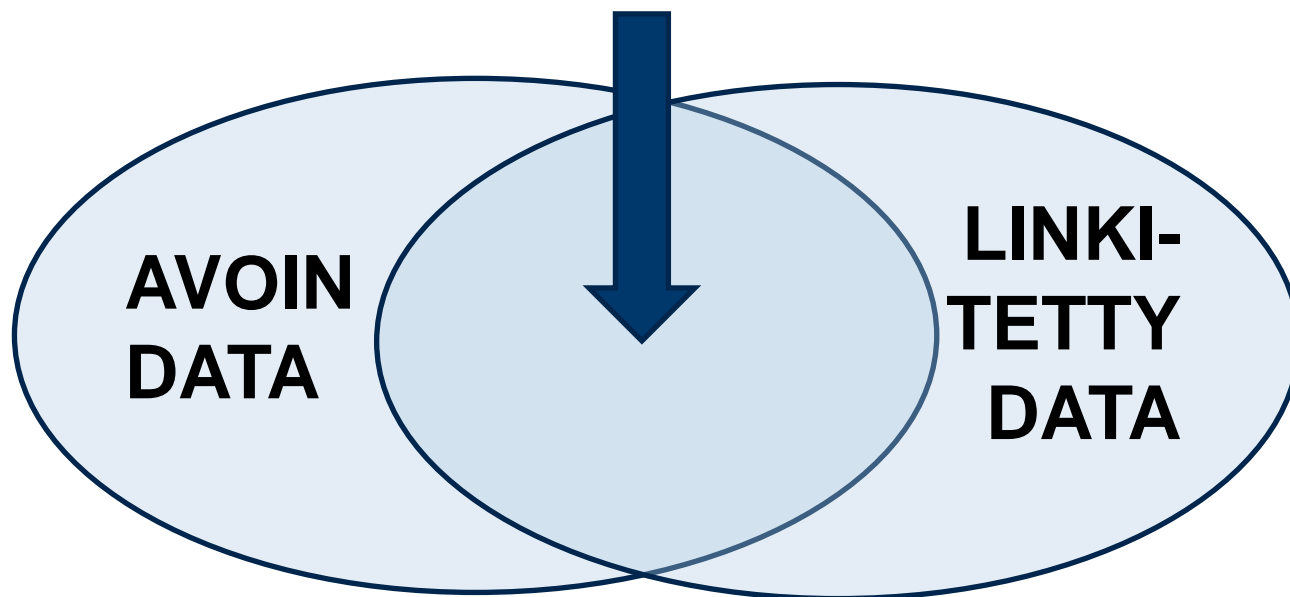
Kirjastoverkkopäivät, 25.10.2012

Laila Heinemann
Tietojärjestelmäasiantuntija, Kansalliskirjasto/Kirjastoverkkopalvelut
laila.heinemann (at) helsinki.fi

Kirjastoissa avattavaa dataa?

- Osoite- ja aukiolotiedot
- Palvelutiedot (*lukusali, satutunnit, musiikkiosasto, kotipalvelu jne.*)
- Tapahtumatiedot
- Bibliografinen metatieto
- Auktoriteettitiedot (*asiasanastot, tekijät jne.*)
- Digitoidut aineistot (HUOM! tekijänoikeus)
- Lainaustiedot (HUOM! anonyymit)

Avoim linkitetty data



Avoin data = Open data

- tarkoittaa yksinkertaisesti dataa, joka on **vapaasti käytettävissä ja kierrätettävissä avoimella lisenssillä**
 - Creative Commons 0, CC Attribution 3.0, ODC-BY tms.
- **periaatteellinen/poliittinen kysymys**
- Valtioneuvoston suositus
Julkishallinnon tietoluovutusten periaatteet ja käytännöt:
http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/076_ict/20120208Julkis/name.jsp

Linkitetty data = Linked data

- tarkoittaa **tietoelementtejä**, jotka on linkitetty toisiinsa
- on seuraava askel internetin loogisessa jatkumossa
 - linkitetyt palvelimet > linkitetyt dokumentit > linkitetyt tiedot
- tietomalli ja sen tekniset toteutukset
- **tekninen kysymys**

Avoin linkitetty data = Linked open data (LOD)

- dataa, joka on avattu sellaisessa muodossa, että myös **muut järjestelmät** pystyvät hyödyntämään sitä
- sitä voidaan laajentaa, kierrättää ja yhdistellä edelleen eri sovelluksissa
- **toiminnallinen** kysymys



”Viiden tähden” avoin data



saatavilla verkossa **avoimella lisenssillä**



saatavilla **koneluettavassa rakenteisessa muodossa**
(esim. excel-taulukkona eikä vain kuvana siitä)



saatavilla koneluettavassa rakenteisessa ja
järjestelmäriippumattomassa formaatissa
(esim. CSVnä eikä excelinä)

MARC



edellisen lisäksi on käytetty avoimia standardeja (RDF ja SPARQL) identifioimaan tietoja, niin että julkaistuun **dataan voidaan viitata**



Kaiken edellisen lisäksi: data on **linkitetty muiden dataan** jotta se liittyy oikeaan kontekstiin

Datan avaamisen esteitä?

Oikeudet



- Tekijänoikeudet
 - Liittyy kokoteksteihin ja digitaalisiin objekteihin
 - Metatietoon ei Suomen lainsäädännön mukaan ole tekijänoikeuksia
- Omistusoikeudet
 - Kokonaiseen tietokantaan voi olla tietokantaoikeus > ei saa imuroida/jakaa edelleen ”huomattavaa osaa”
 - Metatieto-oikeuksia voi olla rajoitettu ulkopuolisen tiedontuottajan kanssa solmituissa sopimuksissa (esim. vain käyttöoikeus, mutta ei luovutusoikeutta)

Raha

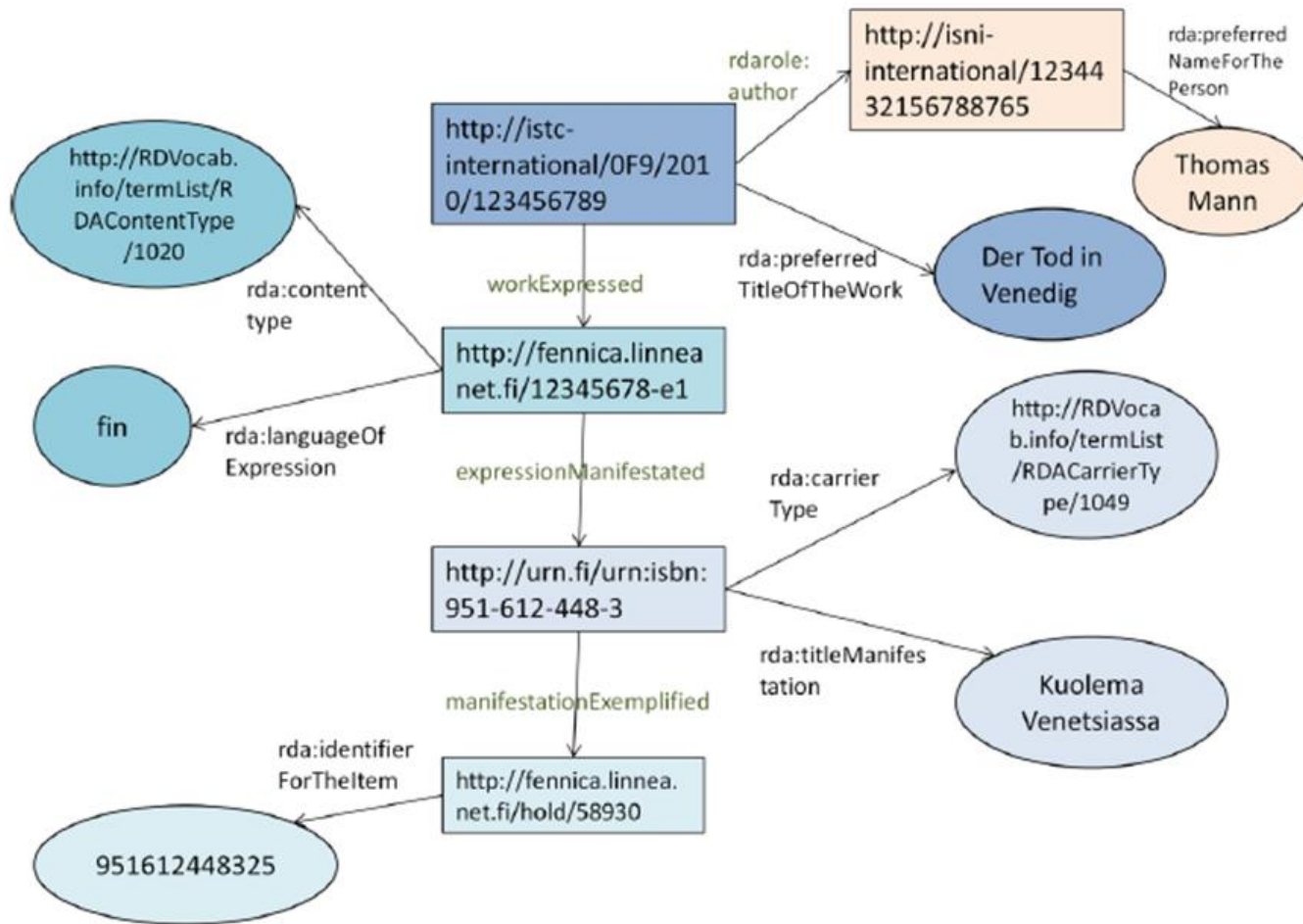


- (Meta)tieto ei synny itsestään
– kuka kustantaa sen tuottamisen?
- Valtionhallinnossa **tahtotila**, että valtion olisi katettava budjettivaroin julkisen avoimen tiedon tuotantokustannukset (ei vielä käytännössä toteutunut kattavasti)
- Ei ota kantaa kaupallisen toimijan tuottaman datan ostokustannuksiin

RDA ja BIBFRAME : 2 askelta kohti linkitettävyyttä

RDA-tietomalli

- Perustuu tietoelementtien **linkitykseen**
= sama logiikka kuin RDF-mallinnuksessa
- Otetaan tuotantokäyttöön pilottikirjastoissa jo 2013
 - Kongressin kirjasto 31.3.2013
 - British Library, Deutsche Bibliothek
- Hanke käynnistynyt Kansalliskirjastossa > käyttöön 2015?
- Ongelmana vanhentuneet kirjastojärjestelmät, jotka eivät pysty hyödyntämään täysimittaisesti
- Sanastoa ja tunnisteet: <http://rdvocab.info/>




Esimerkki RDF-mallinnuksesta (Osa URI-tunnisteista kuvitteellisia)

FINMARC > MARC 21 > BIBFRAME??

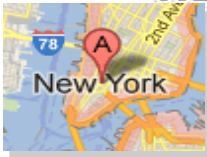


- Nykyinen MARC-formaatti ei sellaisenaan sovellu linkittämiseen, mm.
 - Liian karkea rakenne
 - Kielisidonnaisia määreitä
 - Kaikki tieto yhdessä tietueessa
- **Bibliographic Framework Transition Initiative = BIBFRAME**
 - Pyrkimys luoda FRBR/RDA-pohjainen formaatti
 - RDF-yhteensopiva
 - Tiedot voidaan konvertoida vanhasta MARCista ilman hävikkiä
 - <http://www.loc.gov/marc/transition/>

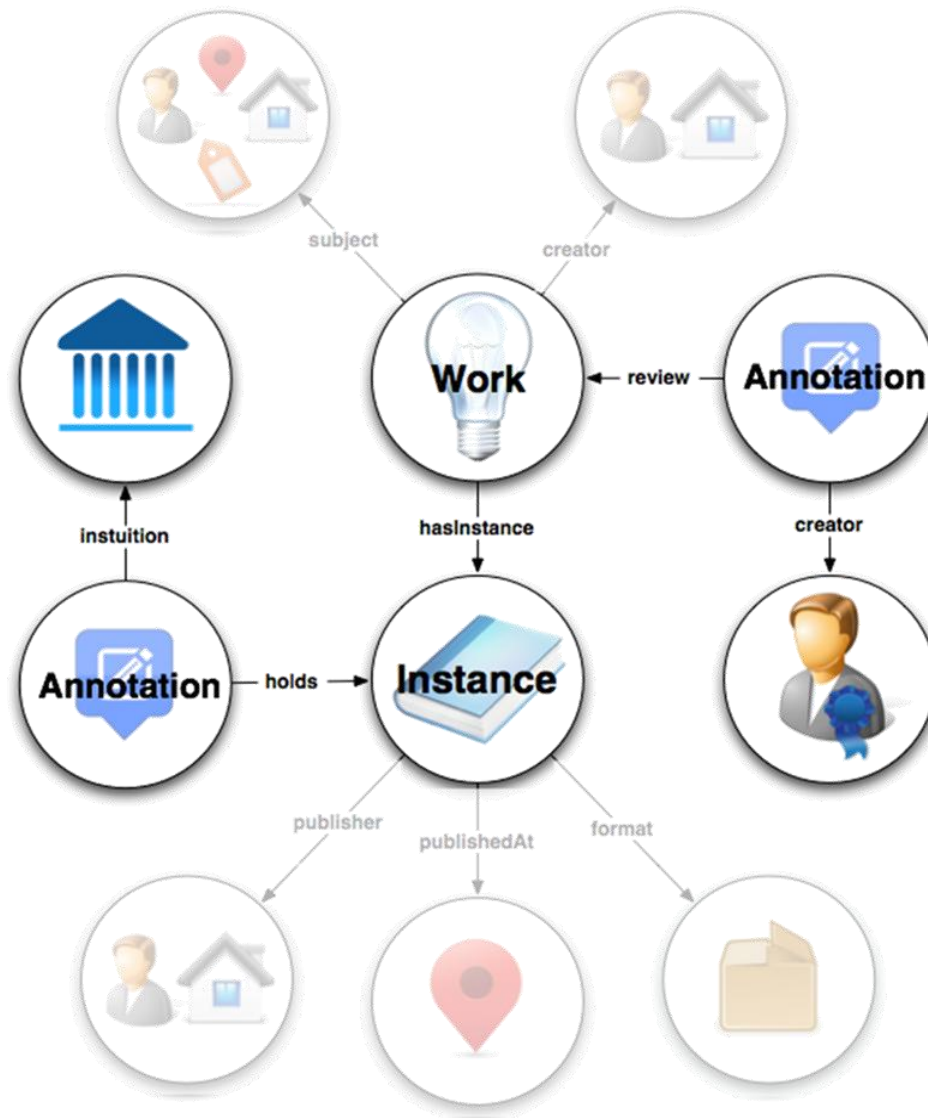
Deconstructing MARC – Identifying MARC Resources (MARCR)



```

001      26132811
008      920608s1992    nyu
010      $a92021087
020      $a1555701205
040      $aDLC$cDLC$dAGL
049      $aHNKA
050      00 $aZ678.9$b.C58 1992
082      00 $a025/.00285$220
1        $aCohn, John M
10       $aPlanning for automation :$ba how-to-do it manual for
         librarians /$cJohn M. Cohn, Ann L. Kelsey, Keith Michael
         Fiels
         $aNew York $bNeal Schuman Publishers,$cc1992
300      $avi, 116 p. ;$c20 cm
440      0 $aHow-to-do it manuals for libraries ;$uno. 25
504      $aIncludes bibliographical references and index
650      0 $aLibraries$xAutomation$xManagement$xHandbooks, manuals, etc
650      0 $aLibrary planning$xHandbooks manuals, etc
700      10 $aKelsey, Ann L
700      10 $aFiels, Keith Michael
    
```



Kuva:
Sally McCallum,
IGeLU 2012

Muna vai kana?

- Kannattaisiko odottaa ensin uutta luettelointiformaattia?
- Toistaiseksi (lähes) kaikki ovat ”keksineet oman pyöränsä”
- Uutta luettelointiformaattia testaavat ensimmäiseksi ne kirjastot, jotka ovat jo julkaisseet dataansa linkitettyinä...

Avoimesti ja linkitettävästi julkaistua (kirjasto)maailmalla

Europeana

- Yhteiseurooppalainen kulttuuriaineistojen portaali
- Mukana satoja kirjastoja, museoita, arkistoja ja muita kulttuurialan toimijoita
 - Suomessa KDK-hanke välittää sisältöjä myös Europeanaan
- Data on julkaistu kahden RDF-syntaksin mukaisesti: RDF/XML ja N-Triples. Europeana-portaalissa se on myös RDFa-muodossa
- <http://www.pro.europeana.eu/web/guest/re-use-data>



OCLC

- Julkaistu **WorldCat**in bibliografisista tietueista yleisimmät (= teos vähintään 250 kirjaston kokoelmassa)
 - Yhteensä 1,2 miljoonaa tietuetta > n. 80 miljoonaa triplettiä
- Seuraavassa vaiheessa julkaistaan **Dewey Decimal Classification System**, **FAST** (Faceted Application of Subject Terminology) ja **VIAF** (Virtual International Authority File)
- Data on julkaistu RDF-muodossa käyttäen Schema.org ontologiaa sekä sen laajennuksia, jotka on rakennettu itse yhdessä yhteistyökumppanien kanssa
- <http://www.oclc.org/us/en/news/releases/2012/201252.htm>

Library of Congress

- Tarjoaa Kongressin kirjaston tuottamia standardeja ja sanastoja, jotka ovat jo olleet avoimesti selattavissa
 - auktoriteettitietoja: subject headings, nimet, luokitukset, maantieteelliset alueet jne.
- Seuraavassa vaiheessa tulossa lisää sanastoja jotka liittyvät PREMIS-, MARC-, MODS-, METS-, ja MIX-standardeihin.
- <http://id.loc.gov/>

LC:n palvelu tarjoaa käyttäjille

(sekä ihmisille että koneellisille)

- Ilmaisen pääsyn dataan
- Rakenteisen pääsyn yksittäisiin tietoelementteihin
- Mahdollisuuden ladata kokonaisia sanastoja ja niiden arvojoukkoja erilaisissa formaateissa
- Mahdollisuuden linkittää omia metatietoja Kongressin kirjaston tietoelementteihin
- Yksinkertaisen REST-rajapinnan, jonka kautta aineistoja voi hakea verkon yli
- Selkeät lisenssi- ja käyttöehdot jokaiselle sanastolle
- Yhteisen kanavan, jonka kautta pääsee käsiksi kaikkien Kongressin kirjaston osastojen dataan
- Mahdollisuuden visualisoida käsitteiden ja arvojoukkojen välisiä suhteita

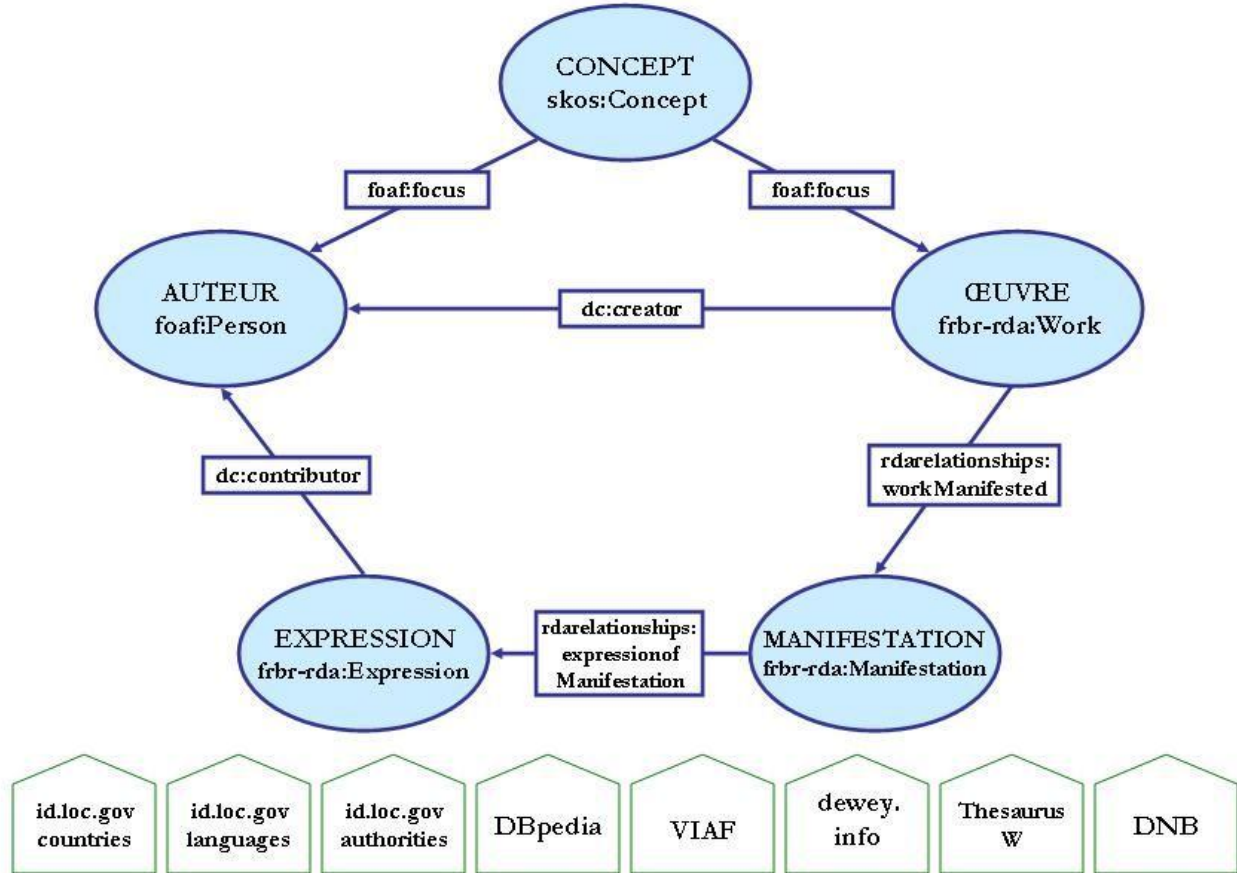
Bibliothèque National de France

- Tarjoaa valikoituja FRBR entiteettejä: tekijät, teokset ja aiheet
- Palvelu oli alunperin tarkoitettu “ihmisluettavia” selaimia varten, mutta samalla toimii myös koneluettavasti
- Kaikki datasetit ovat ladattavissa RDF-dumppeina
- Linkitetty sekä BNF:n omiin aineistoihin (yleiskokoelman luettelo, arkistoaineistojen ja käsikirjoitusten luettelo, Gallica jne.) että ulkopuolisiin (Ranskan tieteellisten kirjastojen yhteisluettelo *Sudoc*, WorldCat, Wikipedia jne.)
- <http://data.bnf.fr/semanticweb-en>



BNF:n tietomalli yksinkertaisesti

Alignements externes

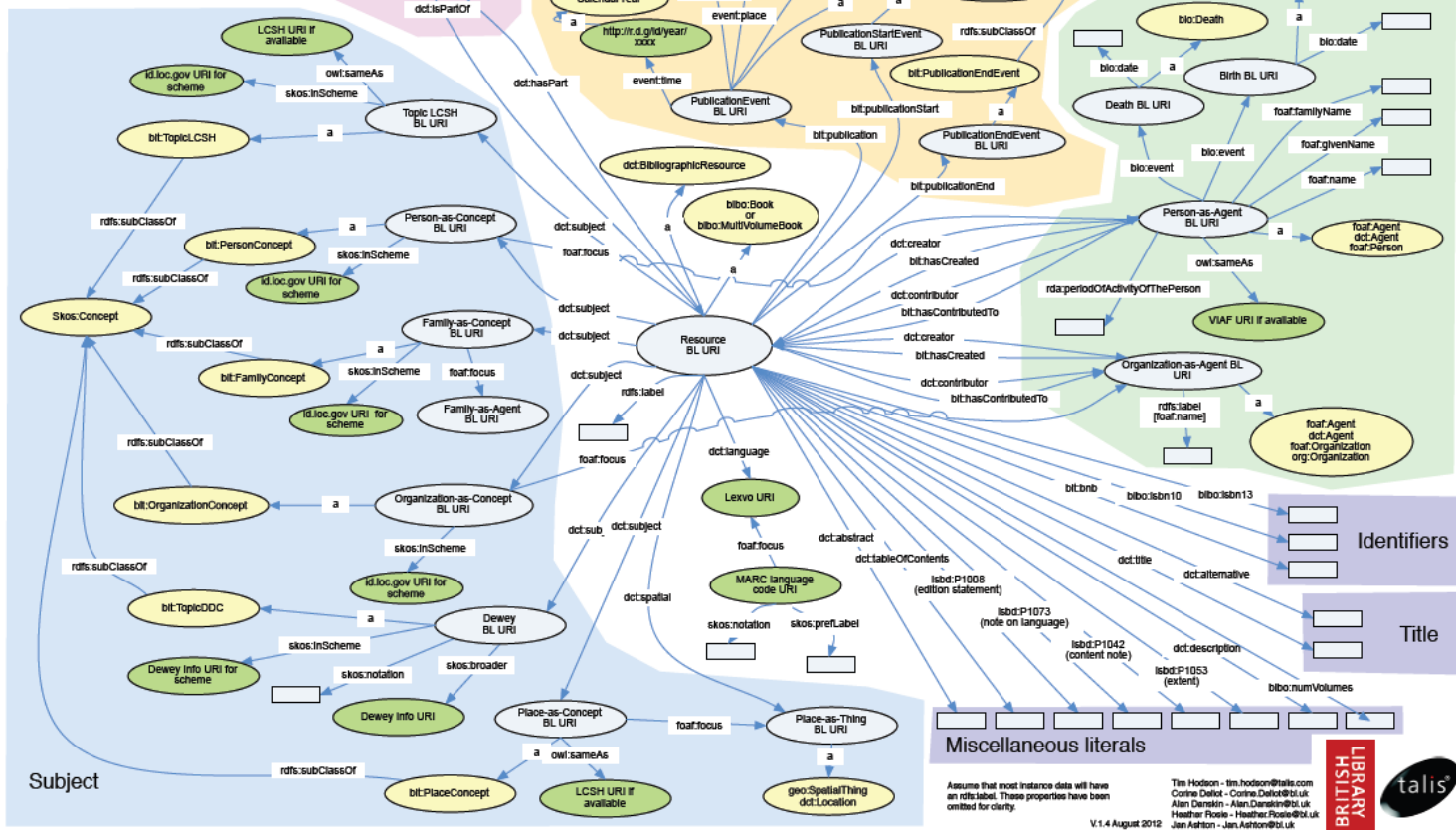


British Library

- Tarjoaa version kansallisbibliografiasta
- Toistaiseksi julkaistut monografiat ja kausijulkaisut
- Tulevaisuudessa kate laajennetaan myös muihin teoksiin ja julkaisujen ennakkotietoihin
- Noin **2,8 miljoonaa tietuetta**, joista on generoitu noin **87.700.000 triplettiä**
- Julkaistu Talisin alustalla
- <http://www.bl.uk/bibliographic/datafree.html>

```

@prefix bit: <http://www.bl.uk/chanas/bibliographic/bldatamodel>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix dct: <http://purl.org/dc/terms#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix isbd: <http://ifastandarda.info/elements/isbd/>.
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix bibo: <http://purl.org/net/bibo/0.1/#>.
@prefix rda: <http://rdvocab.info/elements/crit2/>.
@prefix bibo: <http://purl.org/net/bibo/0.1/#>.
@prefix foaf: <http://xmlns.com/foaf/0.1/#>.
@prefix event: <http://www.w3.org/2004/02/event.owl#>.
@prefix geo: <http://www.w3.org/2003/01/geo/wgsg4_pos#>.
    
```



British
Library
malli mono-
grafioille

Muita (esim.)

- **Deutsche Bibliothek**

- Gemeinsame Normdatei (= Auktoriteettietokanta) ja bibliografinen metadata
- http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html

- **Dansk Bibliotekscenter**

- Tekninen ratkaisu sama kuin British Librarylla

Kansalliskirjaston avoin data

- Asiasanastot (YSA, Allärs, MUSA/Cilla)
 - Saatavana RDF- ja SKOS-muodossa
 - ONKI-palvelun kautta <http://onki.fi/fi/browser/>
- Tulossa
 - Nimiauktoriteetit (*kunhan ensin saadaan ylipäättään tuotantoon*)
- Selvitettävänä
 - Metatietovarannon bibliografinen data

Lisätietoa

- Open Bibliographic Data Guide:
<http://obd.jisc.ac.uk/>
- Open Knowledge Foundation (OKF) Working Group on Open Bibliographic Data:
<http://wiki.okfn.org/wg/bibliography>
- W3C Library Linked Data Incubator Group:
<http://www.w3.org/2005/Incubator/lld/>



16 40

KANSALLIS
KIRJASTO