



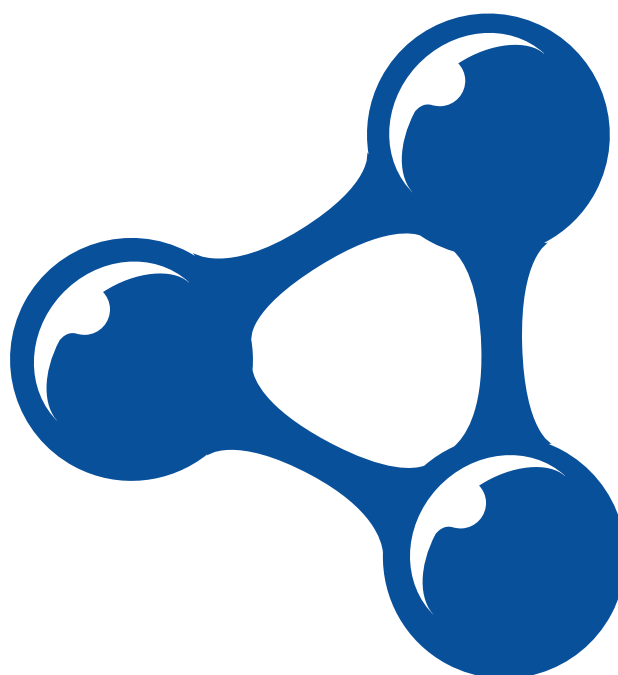
LINK BIBLIOTEKET ÅBENT!

Er Linked Open Data noget, bibliotekerne kan bruge i deres informationsarbejde? Hvilke kilder, værktøjer og metoder er brugbare – og giver det fordele i arbejde med registrering og formidling? Svaret søges i et igangværende DEFF-projekt.

KNUT ANTON BØCKMAN, KNAB@KB.DK
SYSTEMKONSULENT, DET KONGELIGE BIBLIOTEK

Linked Open Data (LOD) handler om at gøre struktureret information mere anvendelig på nettet, og metoden har efterhånden fået en række succesfulde anvendelsesmuligheder. Biblioteksverdenen har også fået øje på potentialet. En række nationalbibliografier og emneordsvokabularer er publiceret som LOD, og det samme gælder dataset fra kulturarvsportaler som Europeana og DPLA. BIBFRAME er et forsøg på at finde et katalogformat, der håndterer denne tilgang til data, men allerede nu udvikles også LOD-baserede bibliotekskataloger af Libris (Sverige) og Deichmanske (Oslo). Formidlingsgrænseflader bygget på LOD-publicerede ressourcer kan blandt andet ses i de tyske, franske og spanske nationalbiblioteker. I projektet Linked, Open & Social, hvor SDUB, SB og KB er aktive deltagere i spor 3, undersøger vi, om LOD også kan udnyttes i mindre skala på de enkelte biblioteker.

Hvilke er de ti største byer, hvor borgmesteren er en kvinde? Oplysningerne til at besvare spørgsmålet kan givetvis findes i Wikipedia, men at finde svaret ved at læse artiklerne er omstændeligt arbejde. Fra en liste over verdens største byer kan man følge links til artiklerne om hver enkelt by, finde ud af, hvem der er borgmester, og fastslå, om vedkommende er kvinde. Derefter videre til den næste by på listen. Der findes en bedre metode; at benytte Wikidata, der er Wikipedias strukturerede bagside. Wikidata indeholder de samme data, men som LOD, det vil sige strukturerede beskrivelser af, hvad der er byer, hvilken størrelse de har, hvem der er borgmester i dem, og hvilket køn disse personer har. Svaret findes ved, at oplysningerne er maskinlæsbare og kan kombineres, og data er dermed blevet mere anvendelige.



Et web af data

Lad forskellen mellem Wikipedia og Wikidata illustrere skellet mellem et web af dokumenter og et web af data. Distinktionen laves af Tim Berners-Lee, når han i 2006 fremsætter principperne for linked data. I Wikipedia linkes der mellem dokumenter, der indeholder information om ting (byer, personer), i Wikidata linkes der mellem tingene selv. Tingene forbindes gennem beskrivelser i form af *prædikative udsagn om entydigt definerede enheder*. De udtrykkes i et Resource Description Framework (RDF) i tripler med subjekt, prædikat, objekt. Mennesker forstår komplekse tekster ud fra den kontekst, de indgår i, men informationen skal forenkles og defineres entydigt for at blive maskinlæsbare. Fx kan sætningen ”Terry Eagletons ’Ideology’ er fra 1991” brydes ned til:

1. Terry Eagleton har skrevet ’Ideology’
2. Terry Eagleton er en person
3. ’Ideology’ er en bog
4. ’Ideology’ blev udgivet i 1991

Dette lyder omstændeligt, men maskiners manglende kontekstuelle forståelse opvejes af, at de er rigtig gode til gentagende handlinger. Derved håndteres det store antal tripler, der gør information fra et komplekst dokument (som en nationalbibliografi) maskinlæsbart. Subjekt, objekt og prædikat identificeres entydigt med Unique Resource Identifiers (URI) tilgængelige via http-protokollen.

Det er afgørende for LOD, at selve linket, forbindelsen mellem de to 'ting' selv, er defineret – til forskel fra almindelige html-links, hvis betydning forstås ud fra kontekst.

Hvor kommer URI'erne fra? Man kan publicere dem efter behov, men idéen i LOD er, at man benytter anerkendte og udbredte vokabularer, enten direkte eller via linkning. Dermed kan andres arbejde genanvendes og ens egne data forbindes med andres data og bliver mere anvendelige. Data, der publiceres og anvendes som LOD, skal være offentlige og til fri brug.

I bibliotekerne

Hvis LOD er etablerede vokabularer, identifikatorer, strukturerede data og eksponering på nettet – har biblioteker så ikke længe drevet med dette i form af klassifikationssystemer, autoritetsposter, katalogposter og web-opacs? Desværre ikke. Vores opacs og discovery-systemer publicerer katalogposter, der i bund og grund er dokumenter. Der er links i dem, men de leder typisk til andre dokumenter, og links er ikke tilstrækkeligt defineret til automatisk forståelse. Mange biblioteker holdt for 20 år siden op med at benytte autoritetsdata. Andre benytter dem kun i lokalt vedligeholdte baser, uden at linke ud til almene kilder, der kan forbinde ens data med andre data – i og udenfor bibliotekssektoren. At der i så lille grad linkes ind til biblioteksdata er en af grundene til, at biblioteksressourcer opnår en dårlig eksponering i søgemaskinerne.

LOD i registrering og formidling

Det er ikke på nogen måde givet, at det er muligt eller ønskeligt at publicere alle katalogdata, (endsige alle biblioteksdata) som LOD. Det aktuelle projekt vil heller ikke afgøre dette. Men det undersøger, hvordan vi kan arbejde med LOD i registrering og formidling. Det er ingen tvivl om, at behovet for URI'er aktualiserer spørgsmålet om brug af autoritetsdata. Centralt vil være at finde metoder til at knytte lokale registreringer af fx personer, emner og stedsnavne til anerkendte kilder. I tilfælde, hvor man har lokale autoritetsdata, er planen at skabe forbindelser fx til VIAF og ORCID for personer, til Getty's AAT for emner og til GeoNames for stedsnavne. Benyttes

ikke lokale autoritetsdata, vil der først være en udfordring med at normalisere og deduplikere variant-indførsler. Til begge aktiviteter findes der værktøjer som OpenRefine eller Catmandu, og projektet skal indhøste erfaringer med brugen af disse.

Det andet undersøgelsesområde er, hvordan LOD kan bidrage til bedre formidling. Med URI'er i posterne tilbydes et sikkert punkt ("Det er denne person"; "Det er dette sted"), der ikke hviler på match af tekststrengene. Fra disse URI'er kan vi bygge services, der beriger de lokale data ved at hente informationer ind fra andre kilder, der benytter eller henviser til de samme URI'er. De færreste brugere vil have glæde af at blive ledt videre til fx VIAF, men via VIAFs URI'er vil man kunne hente data ind fra andre kilder og præsentere beriget information til brugerne.

Linked, Open & Social løber til 2017, og projektledelsen ligger hos DTIC. 🌟



FAKTA

En af de mest kendte succesfulde anvendelser af linked data er BBCs strukturering af nyhedsindhold.

Læs mere her: <http://bbc.in/1FjEz3e>

Linked Jazz er et andet meget vellykket projekt, der også har været stilskabende ikke kun i brug af LOD, men også fx crowdsourcing.

Se det live her: <https://linkedjazz.org/>

Det franske nationalbiblioteks LOD-baserede brugergrænseflade: <http://data.bnf.fr/>

Og den tilsvarende fra det spanske nationalbibliotek: <http://datos.bne.es/> samt Deutsche Digitale Bibliotheks portal <https://www.deutsche-digitale-bibliothek.de/>

Bibliotekers brug af LOD er genstand for en årlig konference, SWIB, som i år holdes i Hamburg 23.-25. november (<http://swib.org/swib15/>).

Det er normalt også et stort emne på ELAG (European Library Automation Group), hvis næste konference afholdes på Det Kongelige Bibliotek 6.-9. Juni 2016 (<http://elag2016.org/>).