

# Med stormskridt mod et nyt webparadigme

Den digitale tidsalder hvor bibliotekerne kunne udvælge, strukturere og topstyre flowet af verdens informationer er endegyldigt forbi. Fremtiden kan være Linked Data, hvor potentialet er enormt. Hvis bibliotekerne åbner deres kataloger.

Esben Fjord, Gladsaxe Bibliotekerne [esbenf@gladsaxe.dk](mailto:esbenf@gladsaxe.dk)

**D**en semantiske web har ført en usikker tilværelse, siden Tim Berners-Lee introducerede sin vision i slutningen af det forrige årtusinde. Men på den nyligt afholdte Online-konference i London, var der megen snak om den semantiske webs snarlige gennembrud som model for fremtidens internet. Det er introduktionen af en ny måde at tænke internettet på, hvor man i modsætning til tidligere ikke fokuserer på forbindelsen mellem dokumenter, men i stedet taler om forbindelser mellem data, såkaldte Linked Data i en slags Web of Data. Men hvad er Linked Data og The Web of Data? Og hvad bliver bibliotekernes rolle i det nye webparadigme?

## En lang historie – kort

Visionen om den semantiske web, der blev skabt af Berners-Lee, handler om, hvordan man kan udtrykke information i et sprog, der kan forstås af computere. I praksis er den semantiske web et lag af metadata, der lægges ovenpå den nuværende webs dokumenter, der beskriver mening og relationer ud fra strikse logiske regler. Disse semantiske metalag gør computere i stand til at 'forstå' dokumenterne på samme måde som mennesker. Et af de første eksempler på et sådan lag af metadata er Dublin Core-standarden, der bruges til at beskrive visse typer af dokumenter ud fra standardiserede felter, der i høj grad minder om dem, man kender fra bibliotekerne.

## Svær begyndelse

Som udgangspunkt var den akademiske verden motoren i udviklingen af den semantiske web. Udgangspunktet var forskningsrelateret, og det handlede i høj grad om at skabe et begrebsapparat, der kunne danne grundlag for implementeringen af visionen, der så i praksis skulle udføres decentralt af webbens indholdsleverandører.

Det viste sig dog hurtigt, at det i praksis var svært for indholdsleverandørerne at gøre den semantiske web til virkelighed. Det var kun ganske få aktører (bl.a. mange biblioteker), der understøttede og værdisatte deres informationer og dokumenter med de nødvendige semantiske lag af metadata. Begrebsapparatet var komplekst og svært at forstå, og der var meget ekstra arbejde forbundet med processen, som det umiddelbart var svært at se nytten af. Derfor forblev visionen om den semantiske web i mange år en drøm.

## Linked Data og The Web of Data

I sommeren 2006 skrev Tim Berners-Lee et notat, hvor han for første gang introducerede begrebet Linked Data. Ifølge Berners-Lee var Linked Data den sidste byggesten i den semantiske web.

Linked Data handler om at finde og strukturere de data, der findes i de ustrukturerede dokumenter og deres relationer til andre data. Hvis man f.eks. tager en Wikipedia-artikel om København, handler Linked Data om at finde alle de fakta og relationer, der findes i artiklen og beskrive dem strukturelt. Eksempelvis at København er hovedstad, ligger i Danmark, er placeret på en bestemt geografisk lokation m.m. Linked Data handler med andre ord om at skabe katalogiserede metadata ud fra flade ustrukturerede dokumenter og linke dem til hinanden.

Linked Data kaldes også for The Web of Data og handler om at forbinde data med hinanden i stedet for dokumenter, som på internettet i dag. Det er et forsøg på at samle strukturerede datamønstre, som f.eks. bibliografiske data om bøger, musik, geografiske data m.fl. og forbinde dem til hinanden. Og dermed skabe et grundlag for at bygge intelligente applikationer ovenpå, der kan løse den udfordring, som den semantiske web i sin tid beskrev.

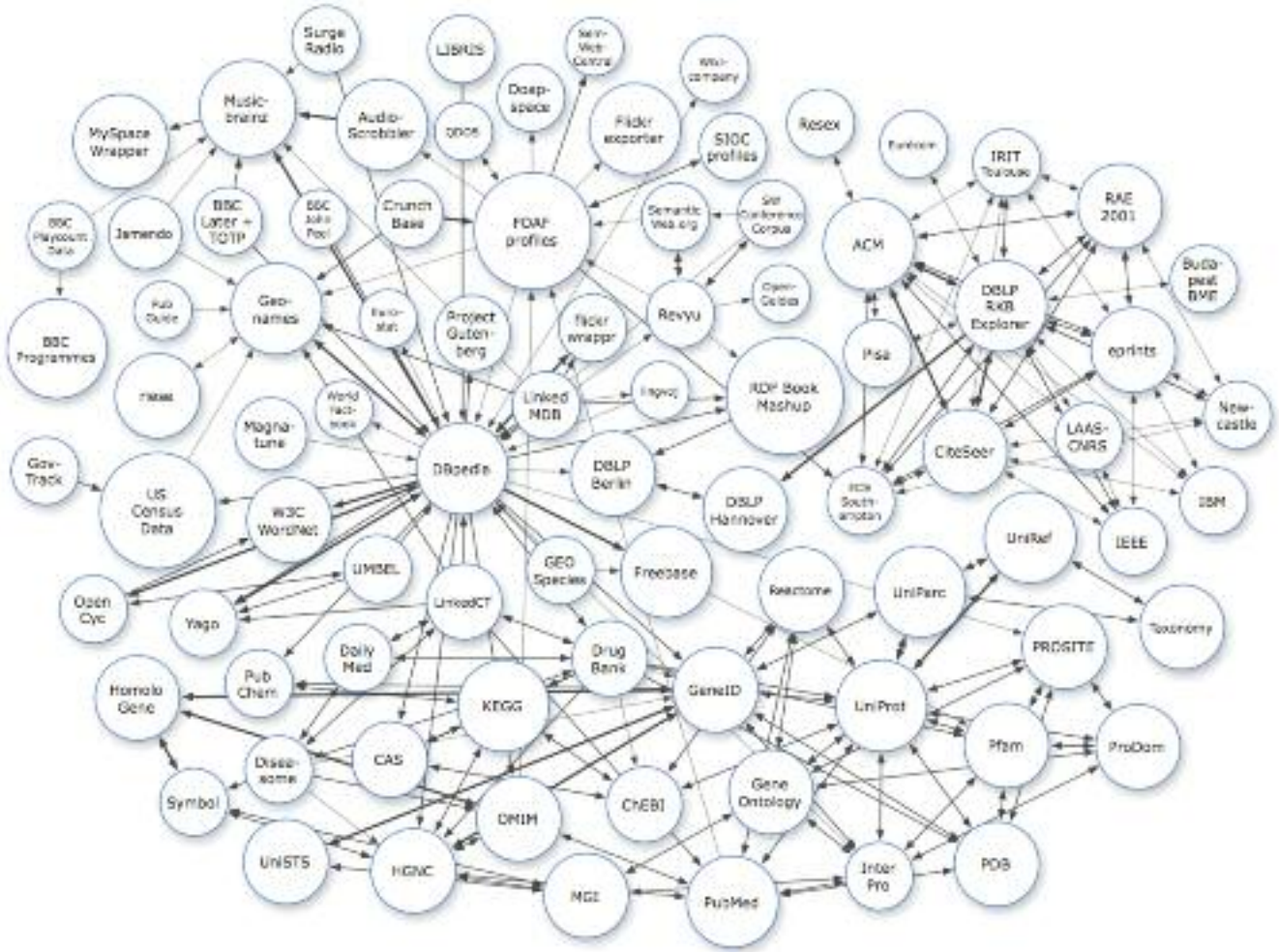
## Linked Data vinder frem

Der findes allerede mange forskellige såkaldte dataset, der hver for sig indsamler og katalogiserer, data indenfor forskellige vidensdomæner. F.eks. er Musicbrainz et initiativ, hvor frivillige indsamler metadata om musik, og Geonames er et dataset med geografiske data fra hele verden, og der findes mange lignende projekter. Via Linking Open Data-projektet prøver W3C at skabe en samlet platform for disse mange forskellige dataset i en slags databrønd. Også på nationalt niveau har man fået øjnene op for Linked Data, og såvel den amerikanske som den britiske stat har startet Linked Data-projekter med offentlige data, der stilles frit til rådighed.

Siden Linking Open Data-projektet blev søsat i maj 2007 har flere og flere dataset-initiativer tilsluttet sig, og ved den seneste optælling i november 2009 var der ca. 13.1 milliarder formaliserede metadataposter i RDF-format. Der er altså meget, der tyder på, at Linked Data meget vel kan blive den fremtidige infrastruktur for The Web of Data, hvorpå man via forskellige web-services, api'er eller programmer kan skabe semantiske applikationer og dermed være medvirkende til, at den semantiske web bliver en realitet i den nærmeste fremtid.

## Bibliotekernes rolle

Traditionelt set har det indtil internettets indtog været bibliotekernes rolle at udvælge og strukturere verdens informationer og stille dem til rådighed via deres katalogiseringssystemer. Dette er sket gennem en streng styret proces, hvor det er bibliotekerne selv, der har kontrolleret og defineret systemerne. Der har med andre ord været tale om en høj grad af topstyring.




Virkeligheden i dag er helt anderledes. Internettets flade og dynamiske struktur har vist sig at gøre det umuligt at topstyre informationsstrømmene, og derfor kan initiativer som Linked Data og Web of Data meget vel vise sig at være nøglen til den semantiske web. Dette bør bibliotekerne for alvor begynde at forholde sig til og begynde at understøtte.

Det kunne bl.a. gøres ved, at bibliotekerne stiller deres egne datadomæner i form af deres bibliografiske indekser til rådighed som Linked Data, så de kan indgå i Linking Open Data-projektet. Der er allerede eksempler på biblioteker, som er begyndt herpå, eksempelvis CERN's bibliotek, Open Library m.fl.

Der findes rigtig meget ekspertviden på bibliotekerne, og derfor er det oplagt, at bibliotekerne spiller en aktiv rolle i forædlingen af data f.eks. fra sociale communities som Wikipedia m.v.

Hidtil har det været dataloger og de enkelte vidensdomæner, der har udviklet de ontologier og taksonomier, der ligger til grund for de forskellige datasets som Linked Data består af. Dette arbejde kunne bibliotekerne også bidrage med i kraft af deres mangeårige tradition med at opbygge systemer og strukturere data. Og endelig bør der være et kæmpe potentiale for bibliotekerne, hvis de åbner deres kataloger og for alvor begynder at supplere og værdiberige katalogposter med Linked Data.

Der er med andre ord flere måder, hvor bibliotekerne kan bidrage i bestræbelserne på at gøre den semantiske web og Linked Data til en realitet. Samtidig sker udviklingen lige nu med stormskridt, så hvis bibliotekerne vil have indflydelse, er det på tide at indgå som en aktiv medspiller i Linked Data-communityet. Til glæde og gavn for alle. 

#### Litteraturliste:

Tim Berners-Lee: Linked Data - Design Issues, W3C, juli 2006 – <http://www.w3.org/DesignIssues/LinkedData.html>

Ted: Tim Berners-Lee on the next Web, februar 2009 – [http://www.ted.com/index.php/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/index.php/talks/tim_berniers_lee_on_the_next_web.html)

Alexander Korth: The Web of Data: Creating Machine-Accessible Information, Read Write Web, april 2009 – [http://www.readwriteweb.com/archives/web\\_of\\_data\\_machine\\_accessible\\_information.php](http://www.readwriteweb.com/archives/web_of_data_machine_accessible_information.php)

Richard MacManus: Understanding the New Web Era: Web 3.0, Linked Data, Semantic Web, Read Write Web, maj 2009 – [http://www.readwriteweb.com/archives/understanding\\_the\\_new\\_web\\_era\\_web\\_30\\_linked\\_data\\_s.php](http://www.readwriteweb.com/archives/understanding_the_new_web_era_web_30_linked_data_s.php)

Richard MacManus: ReadWriteWeb Interview With Tim Berners-Lee, Part 1: Linked Data, Read Write Web, juli 2009 – [http://www.readwriteweb.com/archives/interview\\_with\\_tim\\_berniers-lee\\_part\\_1.php](http://www.readwriteweb.com/archives/interview_with_tim_berniers-lee_part_1.php)

Richard MacManus: ReadWriteWeb Interview With Tim Berners-Lee, Part 2: Search Engines, User Interfaces for Data, Wolfram Alpha, And More..., Read Write Web, juli 2009 – [http://www.readwriteweb.com/archives/readwriteweb\\_interview\\_with\\_tim\\_berniers-lee\\_part\\_2.php](http://www.readwriteweb.com/archives/readwriteweb_interview_with_tim_berniers-lee_part_2.php)

Wikipedia: Linked Data, januar 2010 – [http://en.wikipedia.org/wiki/Linked\\_Data](http://en.wikipedia.org/wiki/Linked_Data)

Wikipedia: Semantic Web, januar 2010 – [http://en.wikipedia.org/wiki/Semantic\\_web](http://en.wikipedia.org/wiki/Semantic_web)

Linking Open Data – <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

The CERN Library publishes its book catalog as Open Data, januar 2010 – <http://gs-service-bookdata.web.cern.ch/gs-service-bookdata/announcement.html>

About MusicBrainz – [http://musicbrainz.org/doc/About\\_MusicBrainz](http://musicbrainz.org/doc/About_MusicBrainz)

About Geonames – <http://www.geonames.org/about.html>