



# SCHEMA.ORG

## – SØGEMASKINERNES METADATAFORMAT

---

*De store søgemaskiner har lanceret et værktøj til professionelle webudbydere: metadataformatet schema.org. Metadataformat lyder som en biblioteksting, men det er det ikke. I denne sammenhæng er biblioteker bare et lille hjørne.*

AF LEIF ANDRESEN, CHEFKONSULENT, DET KONGELIGE BIBLIOTEK  
LEIF@KB.DK

De store søgemaskiner Google, Bing og Yahoo lancerede i juni 2011 et format for metadata om et website. Formatet kaldes Schema.org og er i kraftig vækst. Allerede to år efter, at de store søgemaskiner med Google i spidsen lancerede dette metadataformat, anvender mere end 10 % af alle websites indekseret i Google Schema.org. Det er heller ikke underligt, at Schema.org, der er et værktøj som lover bedre opmærksomhed fra Google og de andre søgemaskiner, bliver brugt i vidt omfang. Schema.org er som en varedeklaration for websites, der gør det muligt for generelle søgemaskiner som Google, Bing m.fl. at præsentere delinformationer fra websites i en relevant sammenhæng. Eksempelvis ved at præsentere en ”musiker” profil, når der søges på Bruce Springsteen, sammen med kunstnerens største sange og næste koncerter.

### Hvad kan bibliotekerne bruge dette til?

Med denne metode tilbyder Google m.v. at crawle også vores - bibliotekernes - data og bruge dem i søgemaskinernes brugergrænseflader ud fra strukturerede metadata. Da bibliotekernes brugere i stort omfang kommer til bibliotekerne fra Google, så er det vigtigt med metoder, som kan gøre bibliotekerne mere synlige. OCLC lagde juni 2012 WorldCat’s metadata ud som schema.org indbygget i visning i worldcat.org. Dette er fulgt op ultimo februar 2014 med en slags værktøj, hvor schema.org præsenteres knyttet til et værkniveau – med henvisninger til de mange enkeltregistreringer.

De strukturerede data gør det muligt for søgemaskinerne at præsentere enkelt elementer fra et website med kontekst og på den måde gøre søgeresultaterne mere relevante for brugerne. Dette er en stærk mulighed som biblioteker bør udnytte! Ved at opmærke omtale af en koncert som en *MusicEvent* kan Google og andre typer søgemaskiner bruge denne information.

Men det betyder ikke, at schema.org kan bruges til kvalificeret udveksling mellem biblioteker. Hertil kræves der mere end schema.org, som et bredt dække format kan tilbyde. Det er som bekendt aldrig lykket at gøre MARC-format spiseligt for verdenen udenfor biblioteker. Det var sådan set heller ikke den oprindelige tanke med MARC – det var netop interoperabilitet og genbrug mellem biblioteker. Selvom det nye BIBFRAME er i XML og LinkedData, så er dette format også primært for samspil mellem biblioteker.

Men netop fordi alle domæner nødvendigvis må være opmærksomme på schema.org, så betyder det også, at der er et potentiale i at bruge schema.org ikke blot til at gøre sig nemt tilgængelige overfor søgemaskiner, men også som et format til samspil med andre sektorer/domæner – d.v.s. som et format for generel interoperabilitet.

Schema.org ser derfor ud til at blive måden hvorpå bibliotekerne kan gøre sig lækre og tilgængelige for webudviklere i andre brancher end lige vores egen!

Det kan nævnes, at Zotero anvender Schema.org sammen med COinS som en metode til at dette bibliografiske værktøj let kan hente strukturerede bibliografiske oplysninger direkte fra et website, uden at brugeren skal bruge særlige eksport/import funktioner. Dette udnyttes bl.a. af Den Danske Forskningsdatabase.

### Opbygningen af Schema.org

Schema.org er udviklet til et ret omfattende format med mere end 700 entiteter. Toppen dette univers er *Thing*, som bliver opdelt på: *Action*, *BroadcastService*, *Class*, *CreativeWork*, *Event*, *MedicalEntity*, *Organization*, *Person*, *Place*, *Product* og *Property*. Herunder ligger udfoldninger, f.eks. *Thing* > *Organization* > *LocalBusiness* > *Library*.

## Udvidelser for biblioteker

Der er en WC3C arbejdsgruppe (Schema Bib Extend Community Group) i gang med at udfolde schema.org således at bibliotekernes ressourcer kan beskrives bedre. Der foreslås nogle tilføjelser til *CreativeWork*, såsom *translationOf* og *basedOn* samt udvidelser af underliggende entiteter til *Article*. En af pointerne er, at vi som biblioteksfolk må acceptere, at det er andre, som har fastlagt sprogbroden. Det hedder f.eks. ikke "Title" men "Name".

## Bedre resultat i Google?

Det er naturligt at spørge om brugen af schema.org for ejeren af et website giver nogen kontant udbytte i form af en bedre placering i f.eks. Googles skærmpresentation og ranking. Det fremgår af Googles dansk-sprogede FAQ (februar 2014), at det gør det! Google understøtter aktuelt følgende datatyper: *Artikel*, *Begivenheder*, *Lokale virksomheder*, *Film*, *Produkter*, *Restauranter* og *Softwareapplikationer*. Da Google konstant arbejder på at udvide deres funktioner, indeholder schema.org mange nye typer, som Google ret sikkert vil bruge i fremtidige applikationer.

## Hvorfor har søgemaskiner taget dette initiativ?

Formålet er optimering af søgemaskinernes indeksering af data. Det giver mere målrettede metadata at arbejde med for søgemaskinerne. Men det betyder også, at søgemaskinernes funktion kan målrettes bedre. Når alle slags forretninger (spisesteder, boghandlere både fysiske og e-, web-site baserede services m.m.m.) "gør sig til" for Google m.v., så betyder det også, at Google kan overtage markedet fra mere specielle søgemaskiner. Der er ingen grund til at gå til de mange specielle vejvisere m.v. for aktuelle begivenheder, for butikker – herunder også for biblioteker – hvis svaret kan poppe op i en Google søgning.

Så der er ikke tale om, at søgemaskinerne er godgørende – det er en nøgtern forretningsstrategi: det handler om annoncer.

Men det skal ikke hindre, at vi i bibliotekerne skal indrette os efter at få det bedst mulige ud af det – ved at lægge vores data klar til søgemaskinerne i schema.org – både vores katalogposter og servicebeskrivelser af de fysiske og virtuelle biblioteker. De danske bibliotekers poster ligger allerede i worldcat.org som schema.org, men det er værd at undersøge, hvilke fordele der vil være ved en mere direkte præsentation.

## Referencer:

Schema.org <http://schema.org/>

Ofte stillede spørgsmål om schema.org (Google) <https://support.google.com/webmasters/answer/1211158?hl=da>

Linked Data for Libraries: Great Progress, but What Is the Benefit? / Richard Wallis <http://swib.org/swib13/programme.php> (abstract, slides og video)

Schema Bib Extend Community Group <http://www.w3.org/community/schemabibex/> på over 200.000 artikel poster registreret i universiteternes repositorier mellem 2000-2012. 🍷

