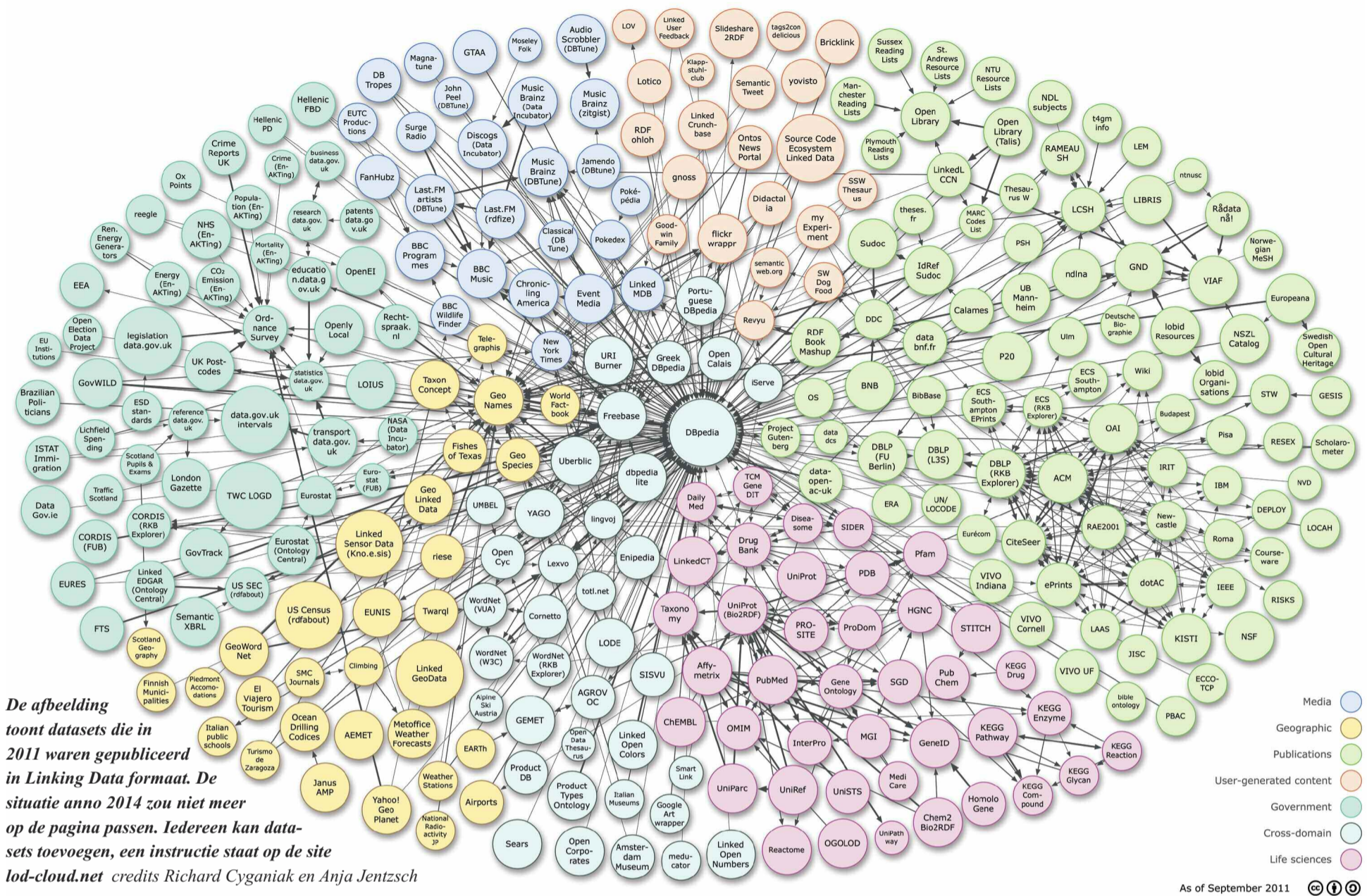


Linked Open Data nader bekeken



De afbeelding toont datasets die in 2011 waren gepubliceerd in Linking Data formaat. De situatie anno 2014 zou niet meer op de pagina passen. Iedereen kan datasets toevoegen, een instructie staat op de site lod-cloud.net credits Richard Cyganiak en Anja Jentzsch

Het Linked Open Data-principe is niet bij iedereen bekend. Een eerste kennismaking met voorbeelden uit de theorie en de praktijk. *René van Horik*

Een 'bank' kan verwijzen naar een financiële instelling of naar een meubelstuk. Wat de term betekent, wordt bepaald door de context. Als de term en de context volgens eenzelfde standaard zijn geformuleerd, kunnen deze informatiebronnen aan elkaar worden gekoppeld. En kunnen ze worden gebruikt voor bijvoorbeeld een nauwkeurigere zoekopdracht. Dit is het principe achter *Linked Open Data*. Door verbanden tussen informatie aan te brengen, ontstaan nieuwe (wetenschappelijke) inzichten. En dat biedt mogelijkheden voor archieven, bibliotheken en andere informatiebeheerders om de kwaliteit van informatiebronnen te verhogen.

Niet alleen maar rozengeur

Linked Open Data zorgen ook voor nieuwe uitdagingen. Door de vele koppelingen tussen en doorverwijzingen naar verschillende informatiebronnen is de kwaliteitsgarantie en de verantwoordelijkheid voor die kwaliteit niet altijd even duidelijk. Ook veranderen informatiebronnen van inhoud, waardoor versiebeheer een punt van aandacht wordt. En er ontstaat nieuwe kennis. De vraag dient zich aan, wie de actualiteit van deze kennis bewaakt en garanties geeft voor de correctheid van de kennis. De reputatie van de dataleverancier komt hierbij in het geding.

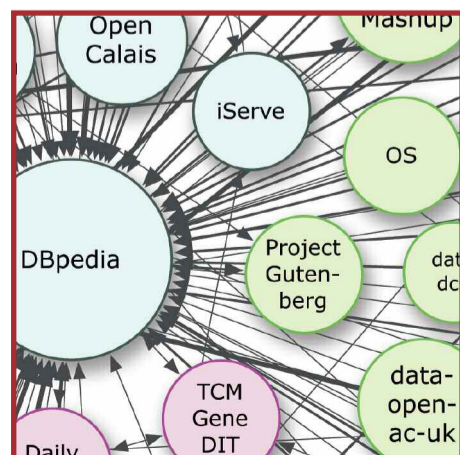
Om *Linked Open Data* in de praktijk toe te kunnen passen en de uitdagingen het hoofd te bieden, is het essentieel dat betrokkenen gebruik maken van dezelfde standaarden.

Het internationale orgaan voor webstandaarden, het World Wide Web Consortium (W3C), geeft leiding aan de ontwikkeling van deze standaarden. Op de website linkeddata.org staan informatiebronnen die als *Linked Open Data* beschikbaar en met elkaar verbonden zijn.

Langetermijntoegang

Ook op projectniveau wordt onderzoek gedaan naar de werkbaarheid van *Linked Open Data*. Een voorbeeld van zo'n project is *Preserving Linked Data (PRELIDA)*. Dit project richt zich op het verbeteren van de langetermijntoegang tot *Linked Open Data* en brengt onderzoekers bij elkaar die actief zijn op het gebied van digitale duurzaamheid en het semantisch web.

linkeddata.org
prelida.eu



Een praktijkvoorbeeld: 'hondsmoeilijk maar de moeite waard'

Linked Open Data in het juridische domein: hoe toegankelijk zijn rechterlijke uitspraken? We spreken Frank van Harmelen, hoogleraar kennisrepresentatie. *Rutger Nugteren*

"Onze vakgroep heeft enkele *Linked Open Data*-projecten gedaan op het gebied van toegankelijkheid van data binnen de rechtspraak. Het bleek hondsmoeilijk om binnen het domein van de rechtspraak te formaliseren en standaardiseren."

Regels van de taal

Dit blijkt ook uit het proefschrift van Marc van Opijnen. Hij stelt dat wetsteksten ongestructureerd zijn en in allerlei combinaties en stijlen voorkomen. Er zijn wel standaarden om te structureren, maar die worden niet overal toegepast. Dit komt mede doordat gerechtelijke uitspraken zijn opgesteld in de natuurlijke taal. De regels van deze taal

zijn afgeleid uit het gangbare gebruik, in plaats van opgesteld op formele wijze, zoals bij computertaal. En de schrijvers van wetsteksten nemen aan dat de lezers allerlei voor de hand liggende zaken zelf weten. Voor een computer is dat echter niet duidelijk. Wanneer je bijvoorbeeld in een formele taal moet definiëren wat een voertuig is, dan moet je vastleggen dat een voertuig gebruikt kan worden door een mens, voordat je een ongeluk met een mens in een voertuig kan beschrijven. In het medische domein zijn dit soort formele relaties vastgelegd in bestaande taxonomieën, terminologieën en classificaties. De mate van standaardisatie is daar veel groter, toch is het niet fundamenteel anders.

Nog hooguit 5 jaar

Volgens Van Harmelen zijn semantische webtechnieken dan ook toepasbaar binnen rechterlijke uitspraken, het zal alleen aanzienlijk meer moeite en tijd kosten. Dat is

ook het beeld dat Van Opijnen oproept.

Een ander obstakel bij rechterlijke uitspraken is de beschikbaarheid van brongegevens. In het proefschrift van Van Opijnen staat: "door de aanwezigheid van steeds meer data, en de beschikbaarheid ervan in voor hergebruik geschikte formaten, zijn de onderzoeksmogelijkheden in een paar jaar explosief toegenomen." "Echter," merkt Van Harmelen op: "we zijn er nog lang niet. Juist de beschikbaarheid van *Linked Open Data* laat nog zeer te wensen over. Veel context ontbreekt doordat uitgeverij hun artikelen nog niet 'open' op internet zetten. De verwachting is dat dit snel zal veranderen. Onze vakgroep ontwikkelt voor Elsevier prototypes die het mogelijk maken om van artikelen naar data te linken. Het zal nog hooguit vijf jaar duren voordat de uitgeverijwereld om is, en artikelen en data beschikbaar stellen en aan elkaar linken," aldus Van Harmelen.