



Forum Standaardisatie

Expertadvies RDFa

Datum 2 april 2014

Colofon

Projectnaam	Expertadvies RDFa
Versienummer	1.0
Locatie	
Organisatie	Forum Standaardisatie Postbus 96810 2509 JE Den Haag forumstandaardisatie@logius.nl
Auteurs	Dr. Ir. E.J.A. Folmer Ir. L.M. Punter

Inhoud

Colofon	2
Inhoud	3
Managementsamenvatting.....	4
1 Doelstelling expertadvies.....	6
1.1 <i>Achtergrond.....</i>	<i>6</i>
1.2 <i>Vraagstelling.....</i>	<i>6</i>
1.3 <i>Gevolgde proces.....</i>	<i>6</i>
1.4 <i>Leeswijzer</i>	<i>7</i>
2 De standaard RDFa	8
2.1 <i>Inleiding.....</i>	<i>8</i>
2.2 <i>RDF</i>	<i>8</i>
2.3 <i>RDFa – inhoudelijk.....</i>	<i>9</i>
2.4 <i>Het gebruik van RDFa</i>	<i>10</i>
2.5 <i>Concurrerende standaarden</i>	<i>12</i>
2.6 <i>Gerelateerde standaarden</i>	<i>13</i>
2.7 <i>Samengevat</i>	<i>17</i>
3 Rondgang langs een aantal betrokkenen	18
3.1 <i>Overzicht.....</i>	<i>18</i>
3.2 <i>Logius / Webrichtlijnen.....</i>	<i>18</i>
3.3 <i>CBS</i>	<i>19</i>
3.4 <i>Kennis- en exploitatiecentrum Officiële OverheidsPublicaties (KOOP)</i>	<i>20</i>
3.5 <i>BKWI.....</i>	<i>20</i>
3.6 <i>Conclusies op basis van de gesprekken</i>	<i>21</i>
4 Advies aan Forum en College	22
4.1 <i>Overwegingen.....</i>	<i>22</i>
4.2 <i>Opname op de lijsten van het Forum Standaardisatie.....</i>	<i>22</i>
4.3 <i>Vervolgstappen</i>	<i>23</i>
4.4 <i>Samengevat</i>	<i>24</i>
5 Referenties	25

Managementsamenvatting

In dit verkennende expertonderzoek is onderzocht welke strategie het Forum Standaardisatie zou kunnen volgen ten aanzien van de standaard RDFa.

Wat houdt de standaard RDFa in?

De standaard RDFa maakt het mogelijk om gestructureerde data, bijvoorbeeld metadata, toe te voegen aan een document (content). Vaak is dit document een webpagina, maar het zou ook een ODF-document kunnen zijn. Metadata is bijvoorbeeld een locatie, een kenmerk, een auteur of een titel.

Deze gestructureerde data kan vervolgens ingelezen worden door een computersysteem. Hiermee wordt het makkelijker om webpagina's te indexeren, content aan elkaar te relateren en content te hergebruiken. Door zoekmachines wordt RDFa gebruikt om bijvoorbeeld producten en prijzen te filteren uit webshops (b.v. voor vergelijkingswebsites) en om content te hergebruiken in social media websites (b.v. Facebook).

Welk proces is gevolgd?

De standaard RDFa is aangemeld door Steve Pemberton van CWI (Centrum voor Wiskunde en Informatica). Hij is ook betrokken bij de ontwikkeling van de standaard door beheerder W3C. Naar aanleiding van deze aanmelding is vervolgens dit verkennende expertonderzoek uitgevoerd.

In het expertonderzoek is onderzocht welke toepassingsgebieden mogelijk zijn, welke relatie en samenhang er is met andere standaarden en wat het draagvlak is voor de standaard binnen de overheid.

Welke samenhang is er met andere standaarden?

Binnen OWMS (de overheidsstandaard voor web-metadata) is het mogelijk om RDFa te gebruiken als syntax (techniek om gegevens vast te leggen) voor metadata. OWMS staat echter ook andere syntaxen toe. Met metadata is het bijvoorbeeld mogelijk om aan te geven wie de auteur is van een document en over welk onderwerp het document gaat.

Binnen de overheid en het onderwijs wordt gebruik gemaakt van specifieke vormen van metadata uitwisseling. Bij de overheid wordt (o.a.) voor een aantal publicaties van decentrale overheden gebruik gemaakt van een eigen XML-formaat en een webservice voor het verzamelen van de gegevens. Deze worden vervolgens via een landelijk portaal (voorziening) getoond. In het onderwijs wordt gebruik gemaakt van OAI-PMH (opgenomen op 'pas toe of leg uit') voor het verzamelen van metadata van educatieve content. Hoewel RDFa technisch nieuwer is, is niet op voorhand duidelijk of er in deze twee gevallen een positieve business case is voor een migratie naar deze standaard.

Voor bepaalde toepassingen van RDFa vormen de standaarden Microdata en Microformats een alternatief. Met name Microdata (onderdeel van HTML5) is concurrerend. Deze standaard biedt minder mogelijkheden, maar is iets gemakkelijker te implementeren.

Wat is het draagvlak voor de standaard?

In het kader van dit verkennende onderzoek is met enkele (potentieel) betrokkenen gesproken. Uit deze gesprekken komt naar voren dat RDFa positief benaderd wordt. Inhoudelijk vinden de meeste mensen het een standaard met een interessant toepassingsperspectief. Er is echter nog geen sprake van daadwerkelijk grootschalig gebruik of een duidelijk vastomlijnde business case en toepassingsgebied.

Wat is het advies aan het Forum Standaardisatie?

Nu niet opnemen op één van de lijsten

Geadviseerd wordt om op dit moment RDFa nog niet op te nemen op één van de lijsten:

- De standaard is nog niet gangbaar; opname op de lijst met gangbare standaarden is daardoor niet aan de orde.
- Er is nog geen vastomlijnde business case met bijbehorend interoperabiliteitsvraagstuk op grond waarvan opname op de lijst voor 'pas toe of leg uit' overwogen kan worden.

Aanvullend advies

- Gezien de interesse voor de standaard is het verstandig om over enige tijd nogmaals te kijken in hoeverre de standaard gebruikt wordt en/of er een meer vastomlijnde business case is ontstaan. Een goed trigger moment hiervoor kan de afronding van het project PILOD zijn waarbij de deelnemers als vervolgactie aangespoort kunnen worden om concreet aan de slag te gaan met de standaard.
- Er is wel een inhoudelijke toepassingsperspectief, dat gedeeld wordt door de betrokkenen die ten behoeve van dit verkennende onderzoek zijn geïnterviewd. De standaard is goed toepasbaar voor metadatering, zoals o.a. in de OWMS-standaard wordt beschreven. Geadviseerd wordt om twee specifieke toepassingsgebieden verder te verkennen en op haalbaarheid te toetsen:
 1. De publicatie van overheidsinformatie op het internet
Hierbij gaat het specifiek om publicaties als vergunningen en regelgeving die nu nog via proprietary mechanismen worden gepubliceerd en geïndexeerd.
 2. Het publiceren van linked open data
Hierbij gaat het om het bredere vraagstuk van linked open data, waar ook andere standaarden een rol kunnen spelen.

Voor deze toepassingsgebieden zou een sponsor gezocht moeten worden vanuit de overheid. Dit zouden respectievelijk het Ministerie van BZK (rechtstreeks of via KOOP; de beheerder van OWMS en de diverse publicatievoorzieningen) en het Platform Implementatie Linked Open Data (PiLOD) kunnen zijn.

- Daarnaast wordt geadviseerd om thematisch te kijken naar het generieke vraagstuk rondom standaarden voor het semantic web. Dit betreft dan niet alleen standaarden voor linked open data, maar b.v. ook standaarden voor rules, trust en security. Dit heeft betrekking op een breder interoperabiliteitsvraagstuk, waardoor het voor de hand ligt dat het Forum Standaardisatie hier zelf het initiatief voor neemt.

1 Doelstelling expertadvies

1.1 Achtergrond

In 2007 is door het kabinet besloten tot een actieplan Nederland Open in Verbinding [1]. Het doel van dit actieplan is om de informatievoorziening toegankelijker te maken, onafhankelijkheid van ICT-leveranciers te creëren en de weg vrij te maken voor innovatie.

Eén van de maatregelen van het actieplan is het gebruik van een lijst met standaarden, die vallen onder het principe 'pas toe of leg uit' (comply-or-explain) [2]. Het College Standaardisatie, dat in 2006 door het kabinet is ingesteld, spreekt zich uit over de standaarden die op de lijst zullen worden opgenomen, o.a. op basis van een expertbeoordeling van de standaard [3]. Het College Standaardisatie wordt geadviseerd door het Forum Standaardisatie. Bureau Forum Standaardisatie ondersteunt beide instellingen.

Onderwerp van dit expertadvies is RDFa. Deze standaard is aangemeld door Steve Pemberton van CWI (Centrum voor Wiskunde en Informatica) voor opname op de lijst met open standaarden voor 'pas toe of leg uit'.

1.2 Vraagstelling

Op basis van de intake is door het Forum Standaardisatie besloten om een verkennend expertadvies.

Concreet zijn de volgende vragen gesteld:

- Wat is het mogelijke toepassingsgebied van RDFa binnen de Nederlandse overheid?
- Wat is de relatie en samenhang met andere standaarden in dit toepassingsgebied?
- Wat is het draagvlak voor de standaard binnen de Nederlandse overheid? Wordt de standaard al gebruikt?
- Voor welke lijst is de standaard het meest van toepassing?

Aan de hand van dit verkennende expertadvies wil het Forum Standaardisatie komen tot nadere besluitvorming rondom RDFa.

1.3 Gevolgde proces

Dit verkennende expertadvies is opgesteld door:

- Dr. Ir. Erwin Folmer, expert open standaarden bij TNO en betrokken bij PiLOD, het Platform Implementatie Linked Open Data.
- Ir. Matthijs Punter, researcher en consultant connected business en open standaarden bij TNO.

Het advies is tot stand gekomen op basis van:

- Gesprekken met de indiener.
- Desk research en navraag/toetsing bij andere inhoudelijke experts.
- Gesprekken met een aantal betrokkenen vanuit de overheid.

1.4 Leeswijzer

Dit document is als volgt opgebouwd:

- Hoofdstuk 2 bevat een introductie van RDFa en beschrijft de relatie met andere standaarden.
- Hoofdstuk 3 bevat een weergave van een aantal gesprekken die met betrokkenen binnen de overheid zijn gevoerd.
- Hoofdstuk 4 bevat de conclusies en het advies aan het Forum Standaardisatie.

2 De standaard RDFa

2.1 Inleiding

De standaard RDFa maakt het mogelijk RDF te gebruiken in webpagina's en documenten.

In dit hoofdstuk is een korte beschrijving opgenomen van achtereenvolgens:

- RDF – de algemene standaard waar RDFa op gebaseerd is
- De standaard RDFa
- Het gebruik van RDFa
- Concurrerende standaarden
- Gerelateerde standaarden

2.2 RDF

De standaard 'RDF' (Resource Description Framework) is een generieke standaard die een hoeksteen vormt van het semantisch web. RDF is opgenomen op de lijst met gangbare standaarden. RDFa is gebaseerd op de RDF-standaard.

RDF beschrijft een metamodel om relaties vast te leggen tussen 'concepten': gedigitaliseerde beschrijvingen van de werkelijkheid.

In RDF wordt vastgelegd: een resource, een eigenschap en een waarde.

Bijvoorbeeld: "De trein heeft de kleur geel" kan in RDF worden vastgelegd als "De trein" (resource), "heeft kleur" (eigenschap), "geel" (waarde). In RDF wordt dit een 'statement' genoemd. Omdat het altijd over drie zaken gaat wordt een dergelijke statement ook wel een 'triple' genoemd.

Het is mogelijk om in een metamodel de mogelijke eigenschappen vast te leggen. Hierdoor kan de eigenschap van verschillende resources vergeleken worden en worden geïnterpreteerd. Zo zou bijvoorbeeld vastgelegd kunnen zijn dat voor een vervoermiddel de eigenschap "heeft kleur" beschreven kan of moet worden.

Als dan bijvoorbeeld elders is vastgelegd: "De auto heeft de kleur geel" dan kan op basis daarvan geconcludeerd worden dat "de trein" en "de auto" dezelfde kleur hebben, nl. "geel".

Vanwege de structuur is deze semantiek computer-interpreteerbaar geworden. In het voorbeeld kan een computer afleiden dat de trein en de auto dezelfde kleur hebben. Of zoeken naar alle vervoermiddelen met de kleur geel; dit kan niet wanneer deze informatie alleen in een zin of tekst is weergegeven. Hierdoor is RDF een belangrijke standaard in het semantisch web.

De standaard RDF beschrijft in essentie niet meer dan dit metamodel en geen syntax. Wel wordt in de standaard een XML-syntax meegeleverd (RDF/XML)¹.

2.3 RDFa – inhoudelijk

De standaard RDFa is een specifieke toepassing van RDF. RDFa is ook ontwikkeld door W3C en wordt gebruikt om RDF-informatie toe te voegen aan andere XML-documenten.

De meest gebruikte toepassing is het gebruik van RDFa binnen webpagina's (XHTML en HTML). Het is echter ook mogelijk om RDFa te gebruiken binnen ODF-documenten.

De eerste versie van RDFa dateert uit 2007. De laatste versie (1.1) is gepubliceerd als W3C-recommendation medio 2012.

Het doel van RDFa is om relatief ongestructureerde content gemakkelijker computer interpreteerbaar te maken door het toevoegen van gestructureerde data. Onderstaand voorbeeld² geeft dit weer.

Een bepaalde webpagina zou bijvoorbeeld onderstaande HTML-inhoud kunnen bevatten:

```
<html>
<head>
  ...
</head>
<body>
  ...
  <h2>The Trouble with Bob</h2>
  <p>Date: 2011-09-10</p>
  ...
</body>
</html>
```

Hierin is een stuk tekst opgenomen "The Trouble with Bob" en "Date: 2011-09-10". Het is echter niet bij voorbaat duidelijk wat dit voor tekst is (de HTML-tags zeggen enkel iets over de opmaak van die tekst).

Door middel van RDFa kan dit echter wel expliciet worden gemaakt, zoals aangegeven in onderstaand voorbeeld:

```
<html>
<head>
  ...
</head>
<body>
  ...
  <h2 property="http://purl.org/dc/terms/title">
    The Trouble with Bob
  </h2>
```

¹ Maar er zijn echter ook alternatieve syntaxvormen mogelijk, zoals Turtle, N3 (Notation 3) of JSON-LD, ieder met eigen voor- en nadelen.

² Afkomstig uit de RDFa primer [5]

```

<p>Date:
  <span property="http://purl.org/dc/terms/created">
    2011-09-10
  </span>
</p>
...
</body>
</html>

```

Vetgedrukt zijn er nu twee elementen toegevoegd die een *property* (eigenschap) omschrijven. Een computer kan nu interpreteren dat "The Trouble with Bob" de *titel* is van het document en dat "2011-09-10" de *creatiedatum* van het document is.

Naast het element "property" zijn er nog diverse andere attributen die RDFa toevoegt, zoals het datatype, relaties tussen elementen en zelfs de mogelijkheid om computer interpreteerbare content toe te voegen.

RDFa voegt op deze manier een computer interpreteerbare laag toe aan (relatief) ongestructureerde documenten.

2.4 Het gebruik van RDFa

2.4.1 Algemeen

Indien er RDFa-content is opgenomen in een (web-) document dan zijn er diverse gebruikersmogelijkheden³.

- Een browser waarmee het document bekeken wordt kan op basis van de opgenomen gestructureerde data aanvullende functionaliteit bieden. Bijvoorbeeld:
 - Indien een datum getoond wordt kan automatisch een link gemaakt worden naar de kalender van de gebruiker.
 - Indien er een locatie getoond wordt kan automatisch een link gemaakt worden naar een kaartenapplicatie om de locatie op de kaart te tonen.
 - Etc.
- Een crawler/zoekmachine kan de gegevens inlezen en gestructureerd weergeven. Bijvoorbeeld:
 - Algemene metadata, bijvoorbeeld op basis van Dublin Core [6], kan worden gebruikt voor indexering.
 - Specifieke data-elementen ('rich snippets [8]'), zoals:
 - Producten en prijzen.
 - Namen en adressen.
 Deze gegevens kunnen vervolgens gebruikt worden op indexerings- of vergelijkingssites.
 - Gegevens om een pagina op te nemen op social media websites (b.v. Facebook gebruikt RDFa ten behoeve van 'Vind ik leuk' in het OpenGraph-protocol [9]).
- Gestructureerde data zou als website getoond kunnen worden, maar – doordat aan die website met RDFa gegevens zijn toegevoegd – ook geautomatiseerde ingelezen kunnen worden in

³ De gebruikersmogelijkheden in de praktijk zijn afhankelijk van de content/documenten waar met RDFa gestructureerde gegevens aan zijn toegevoegd en de inhoud van die gestructureerde gegevens.

apps/toepassingen van derden. Dit is een laagdrempelige vorm van publiceren van linked open data. Bijvoorbeeld: een tabel met statistische informatie. Deze is dan niet alleen leesbaar voor mensen, maar ook door systemen die deze informatie verder kunnen inlezen/verwerken.

2.4.2 *Mogelijke toepassingsgebieden voor de overheid.*

Als we dit vertalen naar mogelijke functionele toepassingsgebieden voor de overheid, zoals gebruikt binnen het kader van de lijsten van het Forum Standaardisatie, dan zijn er een aantal zaken denkbaar:

1. Het metadateren van websites

Hierbij gaat het om het toevoegen van basale metadata (b.v. Dublin Core) aan een document. Dit verbetert de algemene indexeerbaarheid en vindbaarheid van documenten.

2. Het publiceren van min of meer gestructureerde content

Hierbij gaat het om documenten die in zichzelf een beperkte set gestructureerde data hebben, die verder gaat dan basale metadata. Denk bijvoorbeeld aan: vergunningen, meldingen openbare ruimte, bekendmakingen of regelgeving. Deze hebben vaak unieke identifiers, locaties, verwijzingen naar regelgeving, etc.

Door deze uit te drukken in RDFa wordt het mogelijk om apps te ontwikkelen die specifiek kunnen omgaan met deze domeinspecifieke gegevens. Bijvoorbeeld: het tonen van vergunningen op een kaart, het relateren van regelgeving, etc.

Deze toepassing is vergelijkbaar met het toevoegen van specifieke data-elementen ('rich snippets') voor bijvoorbeeld producten en prijzen die vervolgens in een vergelijkingswebsite opgenomen kunnen worden.

3. Het publiceren van grote sets (linked) open data

Hierbij ligt de focus op de data zelf. Bijvoorbeeld statistische of financiële data, die nu enkel via een computerbestand beschikbaar wordt gesteld. Het gebruik van RDFa zou als voordeel kunnen hebben dat deze data nu als webcontent beschikbaar wordt gesteld, maar tegelijkertijd nog wel computer-interpreteerbaar blijft.

Het mag duidelijk zijn dat dit drie toepassingsgebieden zijn die niet heel hard afgebakend kunnen worden, maar in elkaar overlopen. In de gesprekken met de diverse betrokkenen is getoetst wat het meest geëigende toepassingsgebied zou kunnen zijn.

Voorbeeldtoepassing: London Gazette

In het Verenigd Koninkrijk maakt de London Gazette (vergelijkbaar met de Nederlandse Staatscourant) gebruik van RDFa voor officiële publicaties.

The screenshot shows the homepage of the London Gazette website. On the left is a vertical navigation menu with links such as Home, Policy-Makers, Data Re-Use, Insolvency, Browse, Placing a Notice, Indexes, My Account, My Notices, About the Gazette, Help, and Registered Users. The main content area features a search bar, a 'First Time User' button, and several service-oriented sections: 'Browse Recent Notices' with a category selector, 'Browse Publications' with a publication selector, 'Placing a Notice' with buttons for 'Placing a Notice' and 'Calculate Cost', 'Services' with buttons for 'Business Services' and 'Services for Individuals', and 'Historians' with buttons for 'Search Builder' and 'Search the Archive'. A 'Latest News' section is visible at the bottom.

<http://www.london-gazette.co.uk>

2.5 Concurrerende standaarden

Er zijn een aantal standaarden die gezien kunnen worden als 'concurrerend': *Microformats* (beheerd door CommerceNet) en *Microdata* (beheerd door W3C).

Beide standaarden zijn een verlengstuk van HTML en richten zich daardoor vooral op websites, terwijl RDFa ook in andere XML-documenten gebruikt kan worden.

RDFa is ten opzichte van beide standaarden technisch-inhoudelijk het meest compleet en veelzijdig qua mogelijkheden. Na RDFa volgt Microdata. Microformats biedt de minste mogelijkheden van de drie standaarden⁴.

⁴ Zie bijvoorbeeld <http://manu.sporny.org/2011/uber-comparison-rdfa-md-uf/> voor een technisch-inhoudelijke vergelijking.

2.5.1 *Microformats*

De standaard Microformats maakt het mogelijk om in HTML bepaalde gestructureerde content op te nemen. Het gaat dan om bijvoorbeeld:

- hCalendar → voor agenda items (gebaseerd op vCalendar)
- hProduct → voor product items
- hCard → voor contactinformatie (gebaseerd op vCard)
- etc.

De standaard is ontwikkeld vanaf 2005 door CommerceNet.

2.5.2 *Microdata*

De standaard Microdata maakt onderdeel uit van de HTML5-standaard. Ten opzichte van RDFa biedt het (deels) vergelijkbare mogelijkheden, met name voor wat betreft het metadateren van websites.

Maar er zijn wel verschillen.

- Het uitgangspunt voor Microdata is HTML en het daaraan toevoegen van gestructureerde content.
- Het uitgangspunt van RDFa is de relatief 'rijkere' wereld van het semantisch web/RDF en het inbedden van deze wereld in HTML- (en andere XML-) content.

Hierdoor biedt RDFa meer mogelijkheden, maar is het ook technisch complexer. Microdata biedt minder mogelijkheden, maar is technisch minder complex. Dit leidt tot veel discussies tussen inhoudelijk experts over de wenselijkheid van de ene of de andere standaard.

In de praktijk ondersteunen veel zoekmachines beide standaarden. Enkele zoekmachines, waaronder Google, geven voor sommige toepassingen wel de voorkeur aan Microdata. Vanwege de sterkere koppeling met HTML5 is de standaard (iets) gemakkelijker te implementeren. Daarnaast is er een goede ondersteuning door de website <http://www.schema.org>. Deze bekende website biedt vocabulaires/metadataschema's voor een aantal veelgebruikte toepassingen. Voor deze schema's geldt echter ook weer dat deze via een transformatie ook uit te drukken zijn in RDFa.

Kortom: Microdata is in veel opzichten een concurrent van RDFa. In een aantal opzichten is RDFa technisch/inhoudelijk beter (generieker). Voor een aantal praktische toepassingen geven gebruikers en zoekmachines soms de voorkeur aan Microdata.

2.6 **Gerelateerde standaarden**

2.6.1 *Overheid Web Metadata Standaard (OWMS)*

De standaard OWMS versie 4.0 is opgenomen op de lijst voor 'pas toe of leg uit'. De standaard beschrijft op welke wijze de overheid webdocumenten moet metadateren.

De standaard maakt onderscheid tussen:

- eigenschappen die overheidsinformatie beschrijven
- lijsten van waarden die deze eigenschappen kunnen hebben
- syntax van waarden die deze eigenschappen kunnen hebben

De standaard is gebaseerd op de Dublin Core specificatie, maar vult deze nader in voor overheidstoepassingen.

Versie 4.0 van OWMS bevat de mogelijkheid om de metadata uit te drukken in XHTML+RDFa, naast andere syntaxmethoden.

2.6.2

Contentmodellen

Op basis van OWMS zijn voor specifieke toepassingen contentmodellen ontwikkeld. Deze contentmodellen (ook wel IPM – Informatie Publicatie Modellen – genoemd) beschrijven voor specifieke toepassingen welke metadata opgenomen moet worden.

Voorbeelden van contentmodellen zijn:

- Basis Wetten Bestand (opgenomen op de lijst voor 'pas toe of leg uit')
- Bekendmakingen
- Decentrale Regelgeving
- ...

Hoewel deze contentmodellen uitgedrukt kunnen worden in HTML i.c.m. RDFa wordt op dit moment vaak gekozen voor een eigen XML-formaat om de metadata uit te wisselen met een centrale overheidsvoorzieningen (zie ook de volgende paragraaf).

2.6.3

Methoden voor uitwisseling van metadata met RDFa

Op dit moment zijn er diverse methoden van uitwisseling van metadata met behulp van RDFa.

De standaard RDFa gaat er van uit dat metadata is opgenomen in een document zelf en dat een crawler/zoekmachine deze informatie inleest en indexeert. Er is dus geen sprake van een interactie tussen de aanbieder van het document en het systeem dat de metadata indexeert, anders dan het inlezen van alle documenten en bijbehorende metadata.

Er kan echter ook sprake zijn van een grotere hoeveelheid (meta-)data, van waaruit een *selectie* gemaakt zou moeten worden. In dat geval zou:

- gebruik gemaakt kunnen worden van een z.g.n. 'triple-store' op basis van RDF, waarin de (meta-) data wordt opgeslagen. Bijvoorbeeld over boeken en auteurs.
- Vervolgens kunnen gegevens van daaruit met een specifieke zoektaal (vaak: SPARQL) opgevraagd worden. Bijvoorbeeld: "toon alle boeken met als auteur Jan Cremer".

In oudere metadata-omgevingen wordt op een andere manier hier mee omgegaan. In dit verband zijn twee situaties relevant, vanwege de relatie met standaarden voor 'pas toe of leg uit': onderwijs en overheidsinformatie. We lichten de werking van beide op hoofdlijnen toe.

Onderwijs: OAI-PMH

Ten behoeve van de uitwisseling van metadata over lesmateriaal is op de lijst voor 'pas toe of leg uit' de standaard OAI-PMH opgenomen (Open Archives Initiative, Protocol for Metadata Harvesting). Deze standaard gaat uit van een 'repository' (bibliotheek) met bronmateriaal en een 'harvester' die deze bibliotheken bevraagt:

- De 'harvester' vraagt periodiek gegevens op bij aangesloten bibliotheken. Het systeem gebruikt hiervoor het OAI-PMH protocol

en de daarin beschreven methoden. Zo kan de harvester vragen om alle wijzigingen of alle records die aan bepaalde kenmerken voldoen. De bibliotheek antwoordt vervolgens met de metadata van de gevraagde set content.

- De 'harvester' slaat deze metadata vervolgens op.
- Een gebruiker kan vervolgens terecht bij de 'harvester' om te zoeken in de metadata van alle daarbij aangesloten bibliotheken.

Op deze manier wordt het protocol gebruikt in het basisonderwijs en voortgezet onderwijs voor het zoeken naar lesmateriaal in bibliotheken van verschillende uitgevers.

Overheidsinformatie: Overheids zoekdiensten en portalen

Enkele jaren geleden is er veel aandacht geweest voor een overheidsbrede zoekdienst. Dit heeft uiteindelijk geleid tot een aantal omgevingen waarin gegevens van verschillende overheidinstanties wordt verzameld.

Onderliggend werken deze omgevingen met twee soorten standaarden:

- Contentmodellen of IPMs (Informatie Publicatie Modellen) waarin de metadata van de te ontsluiten gegevens wordt gespecificeerd. Deze metadata wordt doorgaans als XML-bestand op een webserver geplaatst.
- Webservice-koppelingen met een centrale zoekdienst, waarmee periodiek deze XML-bestanden worden uitgewisseld.

De centrale zoekdienst heeft zo alle gegevens beschikbaar en combineert deze in een centrale index⁵. Hierin kunnen gebruikers dan zoeken.

Potentiële overlap met RDFa

Het mag duidelijk zijn dat er in zowel de overheids- als de onderwijssituatie (potentieel) overlap is met RDFa:

- Er is een andere syntax voor het vastleggen van de metadata.
- Er is een specifieke vorm van verzameling van deze metadata, die afwijkt van wat gangbaar is bij toepassing van RDF(a).

Deels kunnen deze verschillen verklaard worden vanuit verschillende behoeften en aard van de systemen. Deels ook door technische vernieuwing (RDFa is een jongere standaard).

Hoewel een overstap naar RDFa mogelijk technische en functionele verbeteringen biedt (bijvoorbeeld: een verbeterde doorzoekbaarheid van overheidswebsites), is niet op voorhand duidelijk of er ook een business case is voor een dergelijke transitie.

2.6.4

Webrichtlijnen

De Webrichtlijnen, die ook is opgenomen op de lijst voor 'pas toe of leg uit', bevatten regels die bedoeld zijn om de toegankelijkheid van websites te vergroten. Het correct toepassen van HTML en metadata maakt hier onderdeel van uit.

In dit verband zou er een relatie kunnen zijn met RDFa. Immers: het opnemen van gestructureerde content d.m.v. RDFa zou wellicht meer mogelijkheden kunnen bieden om een website te tonen in een aangepaste omgeving.

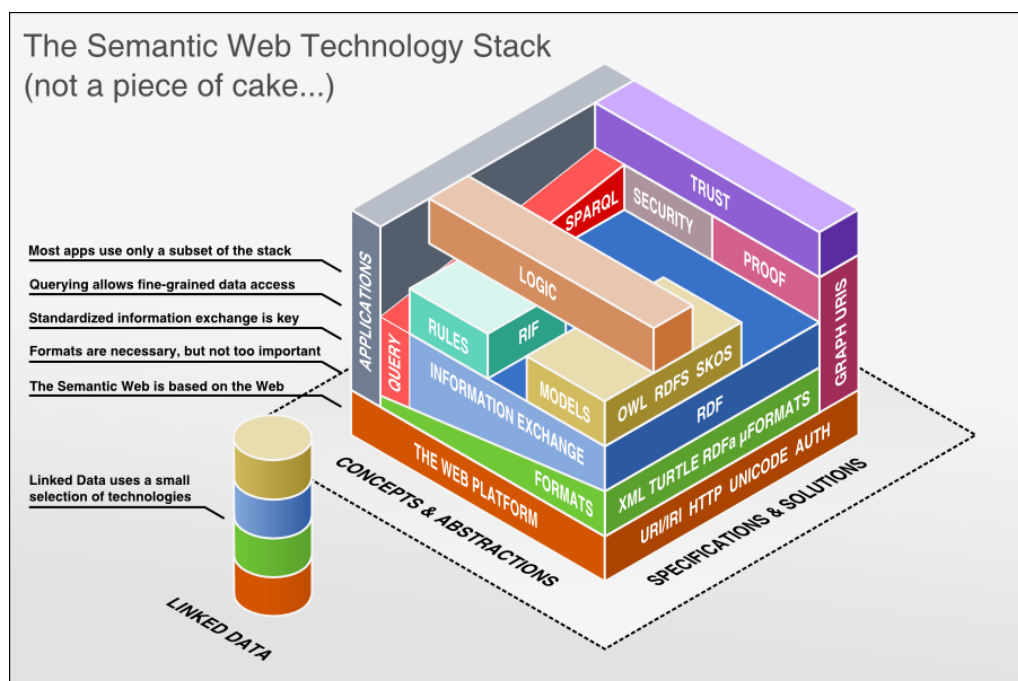
⁵ Zie bijvoorbeeld: <http://koop.overheid.nl/producten/lokale-bekendmakingen/documentatie-lokale-bekendmakingen> voor het 'lokale bekendmakingen'

In een gesprek met Raph de Rooij (betrokken bij het beheer van de Webrichtlijnen) is dit nader getoetst (zie het volgende hoofdstuk). Hierin bleek dat het opnemen van RDFa in de Webrichtlijnen naar zijn mening niet wenselijk is.

2.6.5 Standaarden voor Linked Open Data en Semantic Web

RDFa is één van de (vele) standaarden die gebruikt worden binnen het semantisch web en open linked data.

Onderstaande figuur geeft een overzicht van deze 'stack aan standaarden'.



Figuur: standaarden in het semantic web [7]

Voor het *semantisch web* is een brede set aan technologieën vereist, die betrekking hebben op:

- Het onderliggende web platform
- Formaten en syntaxen
- Informatie-uitwisseling
- Regels
- Modellen
- Applicaties
- Query-talen
- Logica
- Security
- Trust

Deze technologieën hangen samen met diverse standaarden, die nu nog niet opgenomen zijn op de lijst voor 'pas toe of leg uit' of de lijst met gangbare standaarden.

Open linked data wordt gerealiseerd met een subset van deze technologieën en standaarden. Vaak wordt daar de volgende set aan standaarden als basis gezien:

- URI (Uniform Resource Identifier) voor het identificeren van zaken
- HTTP URIs als methode om data beschikbaar te stellen
- RDF voor data
- SPARQL als query taal
- Al of niet aangevuld met andere standaarden.

Binnen het programma PiLOD (waar ook het Forum Standaardisatie aan deelneemt) wordt op een implementatie-gerichte wijze uitgetoond hoe dergelijke standaarden en technologieën ingezet kunnen worden.

2.7 **Samengevat**

- RDFa kan gebruikt worden voor:
 - Het metadateren van webdocumenten.
 - Het ontsluiten van min of meer gestructureerde content, zoals bijvoorbeeld vergunningsinformatie en andere officiële publicaties.
 - Het publiceren van grote sets (open) linked data in een vorm die zowel voor mensen (webpagina) als voor computers (datasets) interpreteerbaar is.
- De belangrijkste concurrerende standaard is Microdata, onderdeel van HTML5. Deze is eenvoudiger te implementeren, maar is functioneel minder rijk. Een andere concurrerende standaard is Microformats.
- OWMS ondersteunt reeds het uitdrukken van overheidsmetadata d.m.v. RDFa.
- Er zijn bij de overheid (via IPM-Contentmodellen en een aantal voorzieningen) en in de onderwijssector (via OAI-PMH) vormen van uitwisseling van metadata in gebruik waarbij andere technologische keuzes zijn gemaakt. RDFa is technologisch nieuwer en biedt mogelijke meerwaarde. Het is echter niet op voorhand duidelijk of een overstap naar RDFa opweegt tegen de kosten.
- RDFa is onderdeel van de set van standaarden t.b.v. het semantisch web en linked open data.

3 Rondgang langs een aantal betrokkenen

3.1 Overzicht

In het kader van dit verkennende expertonderzoek zijn er gesprekken gevoerd met een aantal potentiële betrokkenen binnen de overheid.

Hierbij is ondermeer gevraagd:

- Of men bekend is met de standaard
- Welke toepassingen men ziet voor RDFa
- Of de standaard in aanmerking zou moeten komen voor opname op de lijst voor 'pas toe of leg uit'

Dit hoofdstuk bevat een weergave van de gevoerde gesprekken en geeft de mening van de betrokkenen weer.

3.2 Logius / Webrichtlijnen

Gesproken is met Raph de Rooij, productmanager van de Webrichtlijnen bij Logius.

Hij is enthousiast over de standaard RDFa en de technische mogelijkheden die dit biedt. Intern heeft hij geëxperimenteerd met het gebruik van RDFa. RDFa biedt in zijn optiek goede mogelijkheden om context mee te geven aan informatie die op een website wordt gepubliceerd.

Naast RDFa heeft hij ook geëxperimenteerd met Microformats. Hoewel ook deze standaard context kan meegeven aan webcontext, is het naar zijn mening beperkter qua mogelijkheden.

De standaard RDFa is naar zijn mening vrij eenvoudig te implementeren in situaties waarbij een web-publicatieomgeving is aangesloten op een database. Vanuit deze database kan informatie over de betekenis van de content gemakkelijk worden toegevoegd.

Enkele jaren geleden is er veel energie gestopt in een overheidsbrede zoekdienst waarmee burgers en ondernemers konden zoeken naar overheidsinformatie. Dit is destijds onder verantwoordelijkheid van het ministerie van BZK gedaan. Dit heeft geleid tot de ontwikkeling van de zogenaamde Informatie Publicatie Modellen (IPM's). Deze IPM-standaarden gaan uit van een vaste relatie tussen content aanbieder en zoekdienst.

Eigenlijk zou naar zijn mening met de technologie van nu (waaronder RDFa) hier nog een keer naar gekeken moeten worden. Een overheidsbrede zoekdienst, maar ook portalen voor specifieke toepassingen, zouden veel eenvoudiger op te zetten zijn indien overheidsinformatie op websites structureel gecodeerd zou zijn d.m.v. RDFa. Ook zou dat meer mogelijkheden bieden voor derde partijen om diensten te ontwikkelen op basis van de gepubliceerde informatie.

Opname van RDFa in de Webrichtlijnen ligt volgens De Rooij niet voor de hand, aangezien de Webrichtlijnen doorgaans niet het gebruik van specifieke standaarden voorschrijven.

Opname op de lijst voor 'pas toe of leg uit' zou een goede optie zijn om het gebruik van RDFa aan te jagen, mits een duidelijk toepassingsgebied gedefinieerd kan worden.

Samengevat:

- RDFa is een goede kandidaat standaard voor het publiceren van overheidsinformatie op het internet. Het gaat dan met name om min of meer gestructureerde gegevens, die nu vaak gemetadateert worden via IPM-contentmodellen (vergunningen, bekendmakingen, etc.).
- RDFa zou in de toekomst mogelijk opgenomen kunnen worden op de lijst voor 'pas toe of leg uit', mits er een goed toepassingsgebied gedefinieerd kan worden.
- Opname in de Webrichtlijnen ligt niet voor de hand.

3.3

CBS

Gesproken is met Dick Woensdregt (enterprise architect binnen het CIO Office) en Dennis Raymond (projectmanager).

Het CBS is bij uitstek een partij die grote collecties gestructureerde (statistische) data beheert. Deze data wordt momenteel ontsloten via de StatLine webapplicatie. Via deze applicatie kunnen belanghebbenden en geïnteresseerden statistische informatie opvragen.

Momenteel wordt door het CBS gewerkt aan mogelijkheden om deze data ook geautomatiseerd op te vragen. Hiervoor wordt gekeken naar de OData standaard. OData is het Open Data Protocol, oorspronkelijk ontwikkeld door Microsoft maar thans in beheer bij OASIS. De standaard is ontwikkeld om via het web CRUD-interacties mogelijk te maken op database (toevoegen, opvragen, wijzigen en verwijderen van gegevens).

OData wordt echter primair ingezet voor machine-to-machine koppelingen. Bijvoorbeeld een applicatie van derden die informatie van het CBS wil opvragen en daarvoor het OData-protocol gebruikt.

CBS heeft geen ervaring met het gebruik van RDFa. Wel kent men de standaard RDF.

Het potentieel van RDFa wordt wel herkend. Het zou dan met name gaan om het meegeven van aanvullende metadata bij publicaties. Zo zou bijvoorbeeld een betere link gelegd kunnen worden tussen de rapporten van het CBS met statistische analyses, rapporten onderling en de daaraan gekoppelde data. Op dit moment worden dergelijke koppeling op de website nog niet gelegd. Bij de vervanging van het Content Management Systeem is dit wel een aspect waar men mogelijk naar wil kijken.

Het CBS wijst er op dat dit niet alleen een technisch vraagstuk is, maar ook een organisatorisch vraagstuk. Immers: deze koppelingen zullen wel (vaak: handmatig) gelegd moeten worden en moeten worden beheerd. Dat gebeurt op dit moment nog niet altijd.

Een verplichting voor RDFa via 'pas toe of leg uit' komt voor het CBS nu nog te vroeg. Er zou eerst meer aandacht moeten zijn aan kennisdeling rondom de standaard en invoeringsscenario's.

Samengevat:

- RDFa is een interessante standaard, maar CBS heeft er nog geen ervaring mee.
- Opname op 'pas toe of leg uit' komt nu nog te vroeg, eerst zou er meer aandacht moeten zijn voor kennisdeling en invoeringsscenario's.

3.4 Kennis- en exploitatiecentrum Officiële OverheidsPublicaties (KOOP)

Gesproken is met Hans Overbeek, adviseur metadata, open data en open standaarden.

KOOP is onderdeel van De Werkmaatschappij en is o.a. verantwoordelijk voor de OWMS (Overheid.nl Web Metadata) standaard en diverse portalen voor publicaties van de overheid (regelgeving, bekendmakingen, etc.). Deze portalen werken (grotendeels) met de IPM-standaarden: publicatiemodellen voor specifieke toepassingen.

RDFa wordt gekwalificeerd als een veelbelovende standaard. Er is een intern verkennend vooronderzoek gedaan, waarbij ook gekeken is naar standaarden als Microdata en Microformats.

Twee jaar geleden is RDFa ook opgenomen als mogelijke standaard voor het uitdrukken van metadata in de OWMS-standaard. Hierbij is ook gekeken naar de relatie met schema.org voor het publiceren van metadataschema's voor specifieke toepassingen.

De omgevingen voor het publiceren en indexeren van decentrale regelgeving en overheidspublicaties maken voor het verzamelen van deze content gebruik van specifieke standaarden (zie ook 2.6.2). Een overstap naar RDFa is op dit moment niet voorzien.

Samengevat:

- RDFa wordt gekwalificeerd als een veelbelovende standaard. Behalve intern vooronderzoek wordt er nog niet actief gebruik van gemaakt.
- In OWMS wordt de mogelijkheid geboden om d.m.v. RDFa metadata uit te drukken.

3.5 BKWI

Gesproken is met Willem Kossen, ICT-architect bij BKWI. Hij ziet in RDFa een belangrijk middel om meer metadata mee te geven bij webcontent (de standaard metadata op basis van Dublin Core is immers beperkt qua rijkheid). Door het meegeven van additionele metadata – rijkere semantische beschrijvingen – is het mogelijk voor app ontwikkelaars en zoekmachines om op een betere manier content in te lezen.

Dit is met name relevant in omgevingen waar er min of meer gestructureerde data wordt gepubliceerd. Bijvoorbeeld:

- Datasets op data.overheid.nl
- Diverse rapportage omgevingen

Tegelijkertijd zijn er ook veel toepassingen waar op dit moment nog geen metadata of structuur wordt meegegeven. Het handmatig opgeven van dergelijke metadata kost veel tijd en energie.

Samengevat

De toepassing van RDFa zou zich dus vooral moeten richten op:

- Toepassingen waar geautomatiseerd content wordt geïndexeerd of wordt verzameld.
- Toepassingen waar redelijkerwijs hiertoe extra metadata meegegeven kan worden.

3.6 Conclusies op basis van de gesprekken

- RDFa wordt over het algemeen positief benaderd.
- Het kennisniveau van de standaard is relatief beperkt.
- Er worden verschillende toepassingsgebieden genoemd.
- Uit de gesprekken komt niet een duidelijke business case naar voren met bijbehorend (eenduidig) interoperabiliteitsvraagstuk.

4 Advies aan Forum en College

4.1 Overwegingen

4.1.1 *Interessant toepassingsperspectief*

De standaard RDFa biedt een interessant toepassingsperspectief: het toevoegen van semantische informatie aan ongestructureerde webcontent. Hiermee wordt het gemakkelijker om webcontent te indexeren en te gebruiken in zoekmachines, apps en andere toepassingen.

4.1.2 *Nog weinig kennis en ervaring bij gebruikers binnen de overheid*

Veel mensen kijken op dit moment naar de mogelijkheden die de standaard biedt. Dit heeft zich evenwel bij de geïnterviewde personen nog niet vertaald naar officiële pilots of proefprojecten voor het gebruik van de standaard. De kennis over het gebruik van de standaard is ook nog beperkt.

4.1.3 *Nog geen duidelijke businesscase en bijbehorend interoperabiliteitsvraagstuk*

Ondanks het toepassingsperspectief, is er op dit moment geen duidelijke en eenduidige business case om het gebruik van RDFa te bevorderen. Hierdoor is het ook niet mogelijk om een vastomlijnd toepassingsgebied te definiëren voor opname van RDFa op de lijst voor 'pas toe of leg uit'. De mogelijke toepassing varieert van traditionele web-metadata tot toepassingen op het gebied van linked open data. Voor traditionele metadata voorziet OWMS al in het gebruik van RDFa.

Voor sommige toepassingen geldt dat RDFa concurrerend is met of aanvullend is aan andere standaarden. Zo zou RDFa in theorie bijvoorbeeld gebruikt kunnen worden in overheidsvoorzieningen waar op dit moment nog gebruikt wordt gemaakt van OWMS/IPM-contentmodellen in combinatie met een eigen XML formaat (voor het publiceren van overheidsinformatie - vergunningen, bekendmakingen, etc.).

4.2 Opname op de lijsten van het Forum Standaardisatie

Gegeven de genoemde overwegingen is opname op één van de lijsten van het Forum Standaardisatie op dit moment nog niet aan de orde:

- Het is nog geen gangbare standaard: er zijn nog weinig overheidsinstanties die gebruik maken van RDFa en de kennis over de toepassingsmogelijkheden is beperkt. Opname op de lijst met gangbare standaarden is hierdoor niet aan de orde.
- Opname op de lijst voor 'pas toe of leg uit' is nog prematuur: er is weliswaar een toepassingsperspectief, maar nog geen vastomlijnde business case op basis waarvan de standaard gestimuleerd zou moeten worden via 'pas toe of leg uit'. Voor sommige toepassingen zijn er op dit moment concurrerende standaarden opgenomen of anderzijds in gebruik.

Geadviseerd wordt om daarom nu niet over te gaan tot opname van de standaard op de lijst voor 'pas toe of leg uit' en 'gangbaar'.

4.3 Vervolgstappen

Onbekend maakt in dit geval echter nog niet onbemind.

De standaard wordt over het algemeen positief benaderd. Het zou goed zijn om na een bepaalde periode (bv. 2 jaar) opnieuw te kijken of er meer overheidsinstanties kennis hebben van en/of gebruik maken van RDFa.

In de tussentijd zou op een aantal manieren de business case voor de standaard verder geconcretiseerd kunnen worden. Deze business case zal in vrijwel alle gevallen niet alleen betrekking hebben op RDFa, maar ook op andere standaarden in samenhang.

1. Specifieke toepassingsgebieden verder verkennen en op haalbaarheid toetsen

Het huidige verkennende expertonderzoek geeft aanknopingspunten voor de volgende toepassingsgebieden:

- De publicatie van overheidsinformatie op het internet
Onderzocht zou kunnen worden hoe RDFa gebruikt kan worden in aanvulling op de bestaande IPM-contentmodellen voor het publiceren van overheidsinformatie. In dit onderzoek zou ook gekeken kunnen worden naar de toekomstige architectuur voor overheidszoekdiensten en -portalen. Daar zou ook een standaard als Microdata bij betrokken kunnen worden.
- Publiceren van linked open data
Onderzocht zou kunnen worden hoe d.m.v. RDFa op een meer laagdrempelige manier linked open data gepubliceerd kan worden. Hierbij zouden dan ook andere standaarden voor linked open data bij betrokken moeten worden.

Voor deze toepassingsgebieden zou een sponsor vanuit de overheid gezocht moeten worden. Voor de genoemde gebieden zouden die bijvoorbeeld respectievelijk het ministerie van BZK (rechtstreeks of via KOOP) of overheidspartijen actief in PiLOD (Platform Implementatie Linked Open Data, voor het publiceren van linked open data) kunnen zijn.

2. Thematisch kijken naar standaarden voor het semantic web

Een andere invalshoek zou kunnen zijn door te kijken naar het meer generieke vraagstuk rondom standaarden voor het semantic web, met een veel generieker toepassingsgebied. Hierbij zou dan niet alleen gekeken moeten worden naar RDFa, maar ook naar andere gerelateerde standaarden. Dit betreft dan niet alleen standaarden voor linked open data, maar b.v. ook standaarden voor rules, trust en security.

Een dergelijke thematische aanpak heeft betrekking op een breder interoperabiliteitsvraagstuk. Daarom ligt het in dit geval voor de hand dat het Forum Standaardisatie hier zelf het initiatief voor neemt.

4.4 Samengevat

Op dit moment is opname van RDFa op één van de lijsten niet een logische keuze. De standaard is nog niet gangbaar en er is een onvoldoende eenduidige business case en bijbehorend interoperabiliteitsprobleem om de standaard op te nemen op de lijst voor 'pas toe of leg uit'.

Wel is er een interessant toepassingsperspectief en wordt de standaard over het algemeen positief benaderd. Geadviseerd wordt daarom om over twee jaar opnieuw te kijken naar deze standaard.

In de tussentijd zou het Forum Standaardisatie de ontwikkelingen op dit gebied kunnen versnellen door:

- Twee specifieke toepassingsgebieden nader te beschouwen (publiceren van overheidsinformatie, linked open data). Hierbij zouden dan sponsors vanuit de overheid gezocht moeten worden.
- Zelf thematisch te kijken naar standaarden voor het semantisch web, waarbij RDFa één van de standaarden is.

5 Referenties

- [1] *Actieplan Nederland Open in Verbinding*, 's-Gravenhage: Ministerie van Economische Zaken, 2007.
- [2] "Pas toe of leg uit" is vastgelegd in de "Instructie rijksdienst bij aanschaf ICT-diensten of ICT- producten" van 8 november 2008, en daarnaast in convenanten en afspraken met decentrale overheden. <http://www.forumstandaardisatie.nl/open-standaarden/voor-overheden/pas-toe-of-leg-uit-regime/>
- [3] "Instellingsbesluit College en Forum Standaardisatie 2010", <https://zoek.officielebekendmakingen.nl/stcrt-2010-4499.html>
- [4] "Toetsingprocedure en criteria voor indieners en experts", <http://www.forumstandaardisatie.nl/open-standaarden/aanmelden-en-toetsing/toetsingscriteria/>
- [5] RDFa Primer
<http://www.w3.org/TR/rdfa-primer>
- [6] Dublin Core Metadata Initiatieve
<http://dublincore.org>
- [7] Bron:
<https://joinup.ec.europa.eu/community/ods/document/online-training-material>
- [8] Google over Rich Snippets:
<https://support.google.com/webmasters/answer/99170?hl=nl>
- [9] RDFa OpenGraph bij Facebook:
<https://developers.facebook.com/docs/web/tutorials/scrumptious/open-graph-object/>