

# Van metadata naar kennis

## *Een ontologie creëren met Topic Maps*

Soms zijn metadata te beperkt om een document gemakkelijk terug te vinden en zouden de metadata eigenlijk met meer kennis verrijkt moeten worden. Met de Topic Maps-standaard is het mogelijk om een ontologie ofwel een kenniskaart te creëren waarin metadata deel uitmaken van een netwerk van kennis. Hierdoor wordt het mogelijk om documenten via verschillende paden en zonder belemmeringen te vinden.

*Pieter Brandwijk en Gabriel Hopmans*

Metadata, zoals Dublin Core, zijn direct gekoppeld aan een bron. Een bron (hier opgevat als opgeslagen data) is dan gemetadateerd met gegevens als auteur, titel, uitgever, formaat et cetera. Op deze manier is een document gemakkelijk terug te vinden als men een of meer van deze gegevens kent. Moeilijker is het als men alleen weet dat de auteur werkt voor een bepaalde organisatie maar diens naam niet kent. Of wanneer het trefwoord waarop men zoekt niet direct in de metadata staat, maar hier wel aan gerelateerd is. In deze gevallen zou het mooi zijn als ook de metadata met meer kennis te verrijken zijn. Een auteur werkt bijvoorbeeld voor een bepaalde organisatie of is expert op een bepaald gebied, en een trefwoord is gerelateerd aan andere trefwoorden. Met de Topic Maps-standaard kan een ontologie, een kenniskaart, worden gecreëerd waarin metadata deel uitmaken van een netwerk van kennis. Het wordt dan mogelijk om via verschillende paden en zonder belemmeringen gezochte documenten te vinden. Verder voorziet de standaard in mechanismen om ontologieën uit verschillende domeinen samen te voegen, waardoor een steeds groter kennisdomein ontstaat. Ook biedt de standaard een effectieve scheiding tussen

naamgeving en identificatie, waardoor de willekeur van naamgeving niet langer leidend is voor de identificatie van objecten.

We geven hier een overzicht van in de praktijk voorkomende mechanismen om metadata te classificeren en te beheren. Daarna volgt een introductie in Topic Maps en geven we een voorbeeld van een ontologie gemodelleerd volgens Topic Maps.

### *Metadata, classificatie en beheer*

Bestaande classificatiemechanismen binnen organisaties worden vaak niet gebruikt of onvoldoende benut. Taxonomieën, thesauri en facetten zijn voorbeelden van dergelijke mechanismen die vaak wel bestaan, maar toch niet in staat zijn het zoekproces significant te verkorten. In deze paragraaf wordt duidelijk hoe deze mechanismen met elkaar te maken hebben en hoe Topic Maps in staat is hun krachten te bundelen.

### *Classificatiemechanismen*

#### **Taxonomieën**

Een eenvoudig classificatiemechanisme dat op iedere computer voorkomt is het bestands-

## Samenvatting

Met de Topic Maps-standaard kan een ontologie worden gemaakt waarin metadata deel uitmaken van een netwerk van kennis. Documenten kunnen dan via verschillende paden en zonder belemmeringen gevonden worden. De standaard voorziet in mechanismen om ontologieën uit verschillende domeinen samen te voegen, waardoor een steeds groter kennisdomein ontstaat, en biedt een effectieve scheiding tussen naamgeving en identificatie.

systeem. Dit is in feite te beschouwen als een taxonomie. Mappen worden hiërarchisch ingedeeld van algemeen naar specifiek, waardoor de gebruiker wordt geholpen bij het gezochte bestand uit te komen. De taxonomie heeft wel een aantal beperkingen. Zo kunnen objecten die onder verschillende categorieën horen eigenlijk maar onder één categorie geplaatst worden. In de praktijk betekent dat voor bestandssystemen dat bestanden gekopieerd worden naar een andere map en daar een eigen leven gaan leiden. Een andere oplossing is dat een bestand toch maar onder één categorie wordt geplaatst, met als gevolg dat het bestand onvindbaar is wanneer er via een andere weg wordt gezocht.

### Thesauri

Een thesaurus is eigenlijk een extensie van een taxonomie. De strikt verticale hiërarchie van de taxonomie wordt uitgebreid met de mogelijkheid om dwarsverbanden te leggen tussen termen. Thesauri worden vaak gebruikt voor de lijst met onderwerpen van een document en in online woordenboeken zoals [www.wordnik.com](http://www.wordnik.com).

De thesaurus heeft al grote potentie voor het ondersteunen van het zoekproces omdat hij meerdere zoekpaden mogelijk maakt. Toch wordt de thesaurus in de praktijk vaak niet zo gebruikt en is deze meer een applicatie die aan de zijlijn staat en waarin snel een geschikte term kan worden opgezocht.

### Facetten

Het idee achter facetclassificatie is dat een documentbron meervoudig wordt geclassificeerd door voor elk facet één term te kiezen, waardoor een document over verschillende assen wordt beschreven. Dit lijkt anders dan een thesaurus, maar is in feite een zeer gedisciplineerde manier om een thesaurus te maken en deze te gebruiken voor classificatie. Bekende voorbeelden zijn de facetgebaseerde zoeksystemen van bijvoorbeeld Marktplaats en Bol.

Deze voorbeelden laten ook zien dat facetclassificatie goede structurele mogelijkheden biedt om bronnen te vinden. De facetten leiden tot het steeds verder inperken van de zoekset, totdat de gezochte bron gevonden is. Van deze bron is natuurlijk nog te achterhalen wat de facetten zijn, maar het is niet mogelijk een bron zelf te relateren aan andere bronnen. Immers, de 'facet van'-relatie voldoet hiervoor niet. Hiervoor is een open ontologie nodig, waarin typen associaties vrij gedefinieerd kunnen worden.

### Beheer

In het voorgaande zijn een aantal classificatiemechanismen aan bod gekomen. De potentie van deze mechanismen in relatie tot het classificeren van metadata is aangegeven en er is uitgelegd hoe een taxonomie, thesaurus en facetclassificatie samenhangen. We beschrijven hier het beheer van verschillende bestaande systemen in een Topic Maps-ontologie. Als laatste komt aan bod hoe de ontologie is uit te breiden met een open taal voor het leggen van nieuwe associaties in het kennisdomein.

### Problemen uit de traditionele IT

In de traditionele IT doen zich diverse problemen voor op het gebied van databeheer:

- Veel bestaande systemen staan op zichzelf en een koppeling tussen systemen is vaak moeilijk te realiseren. De terminologie is bijvoorbeeld anders of de systemen hebben een ander datamodel waardoor de samenhang onduidelijk blijft.
- Het zoeken in verschillende systemen is omslachtig. Als het al mogelijk is, komt het vaak neer op een verzamelbak van de resultaten uit de verschillende systemen. Een sortering op relevantie is hierbij uitgesloten.
- De zoekmechanismen zelf zijn vaak alleen tekstgebaseerd, wat wil zeggen dat de kennis van gerelateerde termen uit bijvoorbeeld een thesaurus bij het zoeken ongebruikt blijft. Als die wel gebruikt wordt, is het vaak maar voor één systeem en niet voor meerdere systemen.



### Eén kennisdomein met Topic Maps

Topic Maps pakt deze problemen aan door zowel de classificatiemechanismen als de data in één geïntegreerd kennisdomein op te slaan. Voor de bestaande systemen geldt dan dat eenmalig een vertaalslag gemaakt moet worden tussen het legacymodel en Topic Maps. Zodra bestaande systemen zijn ontsloten in Topic Maps, zijn de systemen geïntegreerd navigeerbaar, op voorwaarde dat de oorspronkelijke systemen een zekere overlap hebben.

In een *topic map*<sup>1</sup> is het dus mogelijk om kennis van verbanden tussen classificaties van metadata te koppelen en deze geïntegreerde kennis direct te gebruiken bij het zoeken naar data.

### Het kennisdomein uitbreiden

Naast het geïntegreerd ontsluiten van bestaande systemen biedt Topic Maps ook de mogelijkheid om nieuwe associaties toe te voegen. Er zijn geen restricties aan wat voor typen associaties dit zijn, want Topic Maps voorziet in een open ontologie. Dit houdt in dat wanneer het voor de eindgebruiker nuttig is om terug te kunnen vinden voor welke organisatie een auteur van een document werkt, dit gegeven in de topic map kan worden vastgelegd. Voor het zoekproces betekent dit dat het niet langer alleen mogelijk is om documenten terug te vinden op basis van de auteur, maar ook op basis van organisatie. Dit kan omdat organisaties geassocieerd zijn aan auteurs, en auteurs op hun beurt weer geassocieerd zijn aan documenten.

## Introductie in Topic Maps

### Topics en associations

Een topic map is een netwerk of graaf of, meer nauwkeurig, een *hypergraph*. De knopen van een topic map representeren onderwerpen (*subjects*). Deze representaties worden *topics* genoemd. De verbindingen tussen topics heten *associations*. Een association drukt de relatie uit tussen een of meer topics. Een association is ongericht. Het onderscheid in welke functie de individuele topics in een association vervullen, wordt bepaald door de rol die zij in de association spelen.

Associations geven de Topic Maps-standaard zijn expressieve kracht. Bij de meeste andere datamodellen zijn relaties tussen objecten alleen in één richting te leggen. Of, als een ongerichte relatie wel uit te drukken is, dit kan vaak alleen maar tussen twee objecten. Dit is bijvoorbeeld het geval bij een relationele database. Een vreemde sleutel is een enkelvoudige relatie van één tabel naar een andere; vanuit de tweede tabel kan men niet terug naar de eerste. Een ongerichte relatie is te leggen door in een tussentabel sleutels van twee andere tabellen op te slaan; de relatie is dan ongericht, maar kan alleen tussen twee objecten worden uitgedrukt. Dankzij associations kent Topic Maps deze beperkingen niet.

Figuur 1 illustreert de concepten topic en association. Als voorbeeld worden termen uit de thesauri van de Rijksoverheid gebruikt. Een topic hierin is de term 'AOW'. Dit topic heeft een relatie met het topic 'SVB' (Sociale Verzekeringsbank); de SVB keert namelijk de AOW uit (in opdracht van het ministerie van SZW). Dit gegeven is een association tussen de twee topics, wat wordt weergegeven als een lijn tussen de twee topics.

1. We gebruiken de term 'Topic Maps' om te verwijzen naar de standaard Topic Maps en 'topic map' om te verwijzen naar een instantie van de standaard Topic Maps.



Figuur 1. Voorbeeld van association tussen twee topics

### Rollen

Figuur 1 bevat een simplificatie. Topics zijn namelijk niet direct onderdeel van een association, maar zijn de spelers van een rol. De association uit het voorgaande voorbeeld is een association tussen een type uitkering en een uitkeringsinstantie. Het topic 'AOW' speelt in deze association de rol van 'uitkering' en het topic 'SVB' speelt de rol van 'uitkeringsinstantie'. Met rollen erbij komt de association er dan uit te zien als in figuur 2.



Figuur 2. Voorbeeld van association tussen twee topics, inclusief rollen

## Occurrences

*Occurrences* zijn waarden die aan een topic gekoppeld kunnen worden. Een voorbeeld is de URL waar het document dat door het topic wordt gepresenteerd te vinden is, indien het document online staat. Of eventueel kan het oorspronkelijke document in een occurrence worden opgeslagen. Een occurrence is in feite een dataveld dat specifiek bij een bepaald topic hoort. Het gegeven is dan wel van belang, maar maakt niet zelf meer deel uit van de ontologie. Zo wordt de geboortedatum van een persoon doorgaans als occurrence opgeslagen.

## Types

'AOW' is een type pensioen en 'SVB' is een type bank. In een topic map wordt dit weergegeven met een voorgedefinieerde association, namelijk de *topic type* association. Een topic type is zelf ook weer een topic, en in een association van dit type speelt het ene topic de rol van topic type en het andere de rol van *instance*.

Ook associations en rollen hebben een type, respectievelijk het *association type* en het *role type*. Ook deze types zijn zelf weer topics. Zo is uit het voorgaande voorbeeld de association tussen 'AOW' en pensioen van het type 'uitgekeerd door'. De role types van de twee gespeelde rollen in de association zijn 'uitkering' en 'instantie'. De volledige weergave van de association ziet er nu uit als in figuur 3.

## Topic names en identifiers

Topic Maps maakt een cruciaal onderscheid tussen de naam van een topic en de identificatie van een topic. Een topic kan zowel meerdere namen hebben als meerdere identificaties, maar een naam identificeert een topic niet. Hierboven kwam het topic met de naam 'AOW' al aan de orde. Hieraan kan een tweede naam worden toegevoegd, bijvoorbeeld 'Algemene Ouderdomswet'. Beide namen gaan over hetzelfde topic, namelijk het topic dat

representeert wat men in de volksmond kent als de AOW.

De identificatie van het topic dat de AOW representeert gebeurt met een *subject identifier*. Dit is een URI die per topic map maar bij één topic mag voorkomen. Zoals gezegd mag een topic wel meerdere subject identifiers hebben. In het geval van het topic voor de AOW is een mogelijke subject identifier bijvoorbeeld 'http://www.rijksoverheid.nl/aow'.

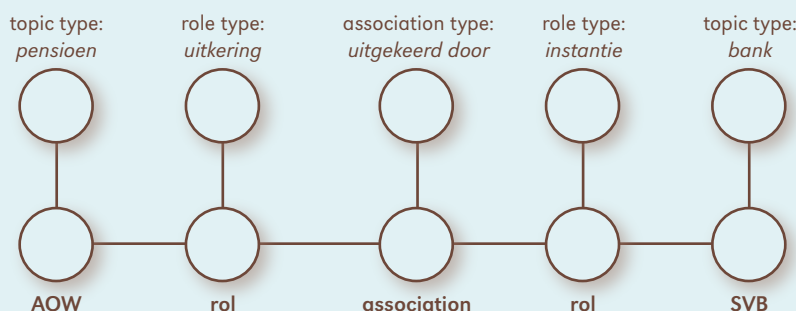
## Merging

Een belangrijk gevolg van het onderscheid tussen topic names en identifiers is dat in een topic map meerdere topics kunnen voorkomen met dezelfde naam of namen. Wanneer in de topic map twee topics voorkomen met dezelfde subject indicator, representeren deze topics hetzelfde subject en worden ze automatisch samengevoegd. Dit heet *topic merging*; hierbij worden alle namen, identifiers, associations en occurrences van de twee topics in een enkel topic opgeslagen.

Merging kan ook plaatsvinden tussen twee topic maps. Er worden dan in feite twee kennisdomeinen samengevoegd. De winst van Topic Maps is dat hiervoor geen enkele conversie nodig is, topic maps kunnen altijd worden samengevoegd. Dit is een krachtig mechanisme wanneer data uit verschillende systemen moeten worden samengevoegd, bijvoorbeeld tussen departementen of zelfs tussen verschillende organisaties.

## Scope

In het voorbeeld van de twee namen voor het topic over de AOW staan nu twee gelijkwaardige namen, namelijk 'AOW' en 'Algemene Ouderdomswet'. Er is nu geen duidelijkheid over welke van deze twee namen in welke situatie te prefereren is. Hiervoor kent Topic Maps het concept *scope*. Met *scope* wordt aangegeven dat 'AOW' een afkorting is en 'Algemene Ouderdomswet' de naam voluit is.



Figuur 3. Volledige weergave van de association

Hierdoor kan afhankelijk van de context bepaald worden welke naam te prefereren is.

**Reification**

De term *reification* komt van het Latijnse *res* (ding) en *facere* (maken) en betekent vrij vertaald ‘een ding maken van’. In Topic Maps betekent reification dat ‘dingen’ die geen topic zijn, bijvoorbeeld associations en topic names, een topic krijgen dat hen representeert. Dit topic is dan weer te gebruiken voor nieuwe associations. Dit is een krachtig mechanisme om zeer specifieke kennis op te slaan. Een reden om reification te gebruiken is bijvoorbeeld om iets over een association te zeggen, zoals in welke periode een wet geldig is.

**Voorbeeldontologie**

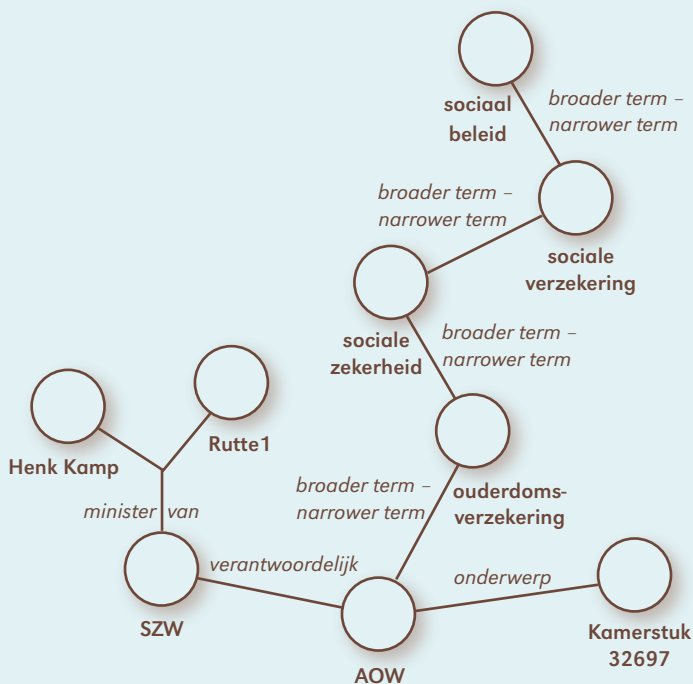
Hiervoor is het voorbeeld van thesaurustermen gebruikt. Dit voorbeeld wordt hier verder uitgewerkt om te laten zien hoe met behulp van een ontologie verbanden kunnen worden gelegd tussen verschillende systemen. Het uitgangspunt is weer de AOW.

De ontologie (zie figuur 4) laat zien dat de AOW valt onder de verantwoordelijkheid van het ministerie van SZW. In de kabinetsperiode Rutte 1 is Henk Kamp minister van SZW. Verder laat de ontologie de thesaurushiërarchie zien van AOW tot sociaal beleid en is er een wetsvoorstel te zien dat betrekking heeft op de AOW. Van het wetsvoorstel is te zien wie het ingediend heeft en op welke wet het betrekking heeft. ‘AOW’ is nu niet langer meer een losse term die aan een document gekoppeld is, maar is een zelfstandige entiteit in een kennisdomein geworden.

Iedere knoop stelt een topic voor en de uitdrukking eronder is een naam van het topic. De lijnen die topics met elkaar verbinden zijn associations. In het geval van AOW, SZW en Henk Kamp is dat een drievoudige relatie. Om de afbeelding overzichtelijk te houden zijn de rollen van de topics niet weergegeven en zijn ook niet alle associations benoemd.

**Topic Maps en het semantische web**

Zo'n zes jaar geleden werden Topic Maps en het semantische web als concurrerende standaarden gezien en werden er ook felle discussies gevoerd. Binnen toepassingen van het semantische web waarin de RDF- en OWL-standaarden worden toegepast, geldt de Open World Assumption (OWA), de veronderstelling dat ‘alles waarvan we niet weten dat het waar is, onwaar is’. Topic Maps wordt in vergelijking met het semantische



Figuur 4. Ontologie voor AOW

web echter vooral toegepast in een afgebakend domein en er wordt uitgegaan van een complete kennisbank; dit is dus eigenlijk de tegenhanger van OWA, de Closed World Assumption.

Er werd rond 2005 en 2006 ook veel energie gestoken in het verhelderen van verschillen en overeenkomsten tussen gebruiksdoelen en concepten om de RDF-standaard uit te drukken in de Topic Maps-standaard. Volgens een W3C-rapport (W3C 2006) is de omgekeerde route om Topic Maps in RDF uit te drukken niet een op een mogelijk en moet dus van geval tot geval worden onderzocht of sprake is van kennisverlies. Toch wordt RDF de de-facto-uitwisselingsstandaard voor Linked Data. Topic Maps is hier een aanvulling op omdat het beter identiteitsmanagement biedt, en ook het concept van merging. Het semantische web en Topic Maps gaan dus uitstekend samen, mits de onderscheiden gebruiksdoelen in beeld blijven. Het semantische web heeft een veel bredere community op het web en bij onderzoeksinstituten, maar het is gewoon voor een ander doeleinde geschikt.<sup>2</sup>

## Semantische kennistoepassingen in de praktijk

### TopicView

TopicView is een toepassing ontwikkeld bij de Politie Amsterdam Amstelland waarmee het werk van de analisten sneller en nauwkeuriger kan worden gedaan. TopicView integreert informatie uit meerdere losstaande primaire systemen tot één overzichtelijk kennissysteem waarin alle informatie in één greep beschikbaar is. Analisten werkzaam voor risicojeugdgroepen kunnen hiermee meer verbanden leggen en hun eigen beweringen en conclusies toevoegen (ministerie van Justitie, 2009). Het kennisgedeelte wordt opgebouwd door middel van taxonomieën van indicatoren; hierin komen ook trefwoorden voor die ook weer voorkomen in de ontsloten bronnen. TopicView wordt uitgebreider beschreven in Hopmans (2011).

### Noorse overheid

De Noorse overheid heeft een webportaal ontwikkeld waar gebruikers snel publicaties en relevante informatie rond de gang van een wetsvoorstel moeten kunnen vinden. De oplossing die hiervoor is gebruikt is de integratie van een zoekmachine met een Topic Maps-navigatie met groepering- en filteringmechanismen om de relevante zaken rond het wetsvoorstel te vinden (zie [www.stortinget.no/no/Saker-og-publikasjoner/Saker/Sak/?p=45513](http://www.stortinget.no/no/Saker-og-publikasjoner/Saker/Sak/?p=45513)).

### ICTU

Bij het ICTU wordt Topic Maps nu ingezet met Ontopia voor de publicatie van formele documenten door de Nederlandse nationale overheid. Er zijn heel veel documenten, in verschillende formaten, met zeer strikte procedures voor publicatie en modificatie. Het doel om Topic Maps in te zetten is om alle ondersteunende documenten op een aparte site te publiceren en specificaties toegankelijker te maken.

### Rechtspraak

Bij de rechtspraak zal Topic Maps worden ingezet voor het beheren, metadateren en publiceren van rechterlijke uitspraken. De keuze voor het gebruik van Topic Maps is vooral ingegeven door de flexibiliteit van Topic Maps om op eenvoudige wijze de eigen collectie documenten te kunnen combineren of te integreren met informatie van derden, zoals wetgeving en rechtswetenschappelijke literatuur.

### Conclusie

De Topic Maps-standaard voorziet in een model waarin verschillende in de praktijk voorkomende classificatiemodellen samenkomen en verrijkt worden. Metadatering van bronnen is hierin de eerste stap. De metadata maken vervolgens zelf weer deel uit van een groter kennisnetwerk, zoals de hiërarchie van een taxonomie of de termrelaties van een thesaurus. Topic Maps gaat echter nog verder door een open ontologie mogelijk te maken. De gebruiker van een classificatiemechanisme is niet langer afhankelijk van door het model voorgedefinieerde relatietypes, maar kan zelf definiëren welke typen associaties er kunnen bestaan tussen bronnen en metadata (topics).

Het Topic Maps-model is ook bij uitstek geschikt voor het semantische web en Linked Data. Onderwerpen (subjects) staan in het model centraal en ieder onderwerp is verbonden met de rest van het kennisdomein. Het koppelen van ontologieën tussen systemen zit al in de standaard ingebakken door de concepten van identificatie en merging. Door via Topic Maps verschillende datasets te ontsluiten creëer je automatisch een geïntegreerd kennisdomein. Een vertalingsmechanisme voor het bestaande datamodel naar Topic Maps en een API op de dataset vormen de kern van een dergelijke integratie.

Topic Maps heeft een levendige community waarin veel materiaal en software beschikbaar zijn. Voor vrijwel alle meest gebruikte talen op het internet zijn implementaties van de standaard

2. Er is pas ook weer een discussie opgestart met de titel 'OWL versus Topic Maps', zie [www.linkedin.com/groups/OWL-versus-Topic-Maps-](http://www.linkedin.com/groups/OWL-versus-Topic-Maps-)

## »»Het Topic Maps-model is bij uitstek geschikt voor het semantische web en Linked Data««

beschikbaar en is het mogelijk om direct met Topic Maps aan de slag te gaan.<sup>3</sup> Complexe en kennisintensieve processen dienen sneller en gemakkelijker te worden aangeboden en dit kan niet anders dan door de traditionele informatietechnologie te gaan verrijken met kennismodellen, contextspecifieke classificaties zoals taxonomieën, ontologieën, thesauri, verbeterde vormen van metadata, en kennisarchitectuur. Al zo'n twintig jaar is er een stroming vanuit de kunstmatige intelligentie die kennistechnologie heet en Topic Maps is een open kennisintegratiestandaard die veel krachtige kennisoplossingen biedt waarin de bovenstaande disciplines worden verenigd.

### Literatuur

- Hopmans, G. (2011). Van applicatielandschap naar informatielandschap. *InformatieProfessional*, nr. 1/2, pp. 24-28, <http://mssm.nl/wp-content/uploads/2011/05/hopmanTopics.pdf>.
- Ministerie van Justitie (2009). Samen effectief: de praktijk als inspiratie. Uitgave ter gelegenheid van de realisatie van het landelijk dekkend netwerk Veiligheidshuizen, [/www.veiligheidshuizen.nl/doc/publicaties/Samen\\_effectief.pdf](http://www.veiligheidshuizen.nl/doc/publicaties/Samen_effectief.pdf).
- W3C (2006). A Survey of RDF/Topic Maps Interoperability Proposals. Semantic Web Best Practices and Development Working Group, [www.w3.org/TR/rdfm-survey](http://www.w3.org/TR/rdfm-survey).

### Pieter Brandwijk

is software-engineer bij Morpheus Kennistechnologie BV.  
E-mail: [p.brandwijk@mssm.nl](mailto:p.brandwijk@mssm.nl).

### Drs. Gabriel Hopmans

is medeoprichter en CEO Morpheus Kennistechnologie BV.  
E-mail: [g.hopmans@mssm.nl](mailto:g.hopmans@mssm.nl).

3. Een tutorial voor de allereerste stappen met Topic Maps is beschikbaar op [www.mssm.nl/tutorials](http://www.mssm.nl/tutorials).