

Eindrapportage Onderzoek OPEN DATA EN HET GEBRUIK VAN STANDAARDEN

OP WEG NAAR VERBETERING AANBOD- EN DISTRIBUTIEPROCES VAN OPEN DATA

Versie: 1.03
Datum: 10 april 2012

Definitief

Inhoud

Managementsamenvatting	4
1. Aanbodproces	4
2. Registratieproces	5
3. Distributieproces	5
1. Ontwerpprincipes voor inrichting publicatieproces	5
2. PTLU-lijst en 'Ruwe-data-nu'-aanpak	6
3. Datavelden voor aanmelding dataset bij data.overheid.nl	6
1. Aanbodproces	6
2. Registratieproces	7
3. Distributieproces	7
1 Inleiding	9
1.1 Aanleiding onderzoek	9
1.2 Centrale onderzoeksvraag	10
1.3 Scope van het onderzoek: de vraag en aanbodketen	11
1.4 Inhoudelijke begrippen	12
1.4.1 Open data	12
1.4.2 Open standaarden	14
1.5 Inrichting van het onderzoeksproces	15
1.6 Onderzoekspopulatie	15
1.7 Leeswijzer	17
2 Beantwoording van de Centrale onderzoeksvraag	18
2.1 Centrale onderzoeksvraag	18
2.2 Belangrijkste bevindingen uit het onderzoek	18
2.2.1 Aanbodproces	18
2.2.2 Registratieproces	22
2.2.3 Distributieproces	23
2.3 Conclusies	24
2.3.1 Ontwerpprincipes voor inrichting publicatieproces	24
2.3.2 PTLU-lijst en 'Ruwe-data-nu'-aanpak	25
2.3.3 Datavelden voor aanmelding dataset bij data.overheid.nl	26
2.4 Aanbevelingen	27
2.4.1 Aanbodproces	27
2.4.2 Registratieproces	27
2.4.3 Distributieproces	28
3 Oordeel per aandachtsgebied	29
3.1 Inleiding	29
3.2 Verschillende perspectieven	30

3.3	A. Herbruikbaarheid (in welk formaat publiceert ik mijn data?)	30
3.3.1	Wat zijn de verschillende soorten van formaat waarin data wordt gepubliceerd?	30
3.3.2	Wat is het beste formaat waarin data kan worden gepubliceerd?	32
3.3.3	Zouden er ten behoeve van Open data nieuwe standaarden op de 'pas toe of leg uit'- lijst moeten worden opgenomen?	33
3.4	B. Vindbaarheid (hoe zorg ik dat mijn data gevonden wordt?)	35
3.4.1	Wat zijn de verschillende mogelijkheden om data te vinden?	35
3.4.2	Op welke wijze i.c. organisatorisch, technisch en procedureel wordt gewaarborgd dat data gevonden worden?	37
3.5	C. De wijze waarop de ontsluiting/infrastructuur wordt ingericht (via welke systematiek publiceert ik mijn data?)	38
3.5.1	Welke systematiek wordt gehanteerd om data te publiceren?	38
3.5.2	Met welke systematiek wordt data op een effectieve en efficiënte wijze gepubliceerd?	39
3.6	D. Relevante ontwikkelingen binnen de EU-beleidskaders	40
3.6.1	Wat zijn de relevante ontwikkelingen binnen de EU-beleidskaders?	41
3.6.2	Wat betekent dit voor de huidige praktijk?	41
3.6.3	Wat zijn de effecten hiervan op het publiceren, vinden en te hanteren systematiek op korte en langere termijn?	41
3.7	E. Lessons learned webrichtlijnen	42
3.7.1	Zijn de ontwikkeling op het gebied van webrichtlijnen vergelijkbaar met de ontwikkeling bij opendata en standaarden?	42
3.7.2	Welke lessen kunnen worden meegenomen vanuit webrichtlijnen?	42

Managementsamenvatting

Data.overheid.nl loopt in de praktijk tegen een aantal belemmeringen aan. Het Forum Standaardisatie (het Forum), ondersteund door het Bureau Forum Standaardisatie (BFS) wil ondersteuning bieden in de vorm van een onderzoek, resulterend in een advies. Hiervoor wil het Forum een onderzoek laten uitvoeren dat resulteert in een advies voor het wegnemen van belemmeringen, gerelateerd aan het gebruik van standaarden. Logica Business Consulting zal dit onderzoek in opdracht van het Forum uitvoeren.

Het uit te voeren onderzoek heeft als doel inzicht te geven in de ervaren technische belemmeringen gerelateerd aan de publicatie van data door publieke partijen en in de mogelijke oplossingen hoe deze belemmeringen kunnen worden weggenomen. Voor het uitvoeren van dit onderzoek is door het Forum de volgende centrale onderzoeksvraag geformuleerd:

Formuleer, op basis van de vragen en analyse van een aantal cases, voorstellen om het aanbod-, registratie- en distributieproces van Open data te verbeteren en stel op basis daarvan een handreiking op aan partijen die betrokken zijn bij publicatie, registratie en gebruik van Open data

Hierbij opmerkend dat dit onderzoek zich focust op aanbevelingen met betrekking tot te gebruiken standaarden, voor herbruikbaarheid, vindbaarheid en wijze waarop de ontsluiting/infrastructuur wordt ingericht vanuit een technische invalshoek vanuit de aanbodzijde. Juridische en beleidmatige aspecten behoren niet tot de scope van het onderzoek.

Om de centrale onderzoeksvraag te kunnen beantwoorden hebben de onderzoekers interviews uitgevoerd bij aanbieders en hergebruikers. De bevindingen hiervan zijn vastgelegd in een casusbeschrijving (totaal 8). Daarnaast is een internationale benchmark uitgevoerd en is via een GroupDecisionRoom-sessie gekomen tot een Syntheserapport. Op basis van deze documenten hebben de onderzoekers zich een feitelijk beeld gevormd van de stand van zaken en hebben de onderzoekers aanbevelingen geformuleerd.

De feitelijke situatie

De belangrijkste bevindingen vanuit de casusbeschrijvingen, de internationale benchmark en het Syntheserapport zijn:

1. Aanbodproces

- Er zijn geen technische drempels om data te publiceren
- Er zijn grote verschillen qua ontwikkelingsniveau tussen de verschillende aanbieders van Open data
- De discussie over Open data kent verschillende werelden: die van overheidsaanbieders en die van de meer academische en de beleidswereld. Overheidsaanbieders zien het groeipad naar het 3de niveau van TBL als haar verantwoordelijkheid, terwijl vooral de meer academische wereld streeft naar het 4de en 5de TBL-niveau. Op beleidsmatig niveau is slechts de uitspraak gedaan dat informatie bij voorkeur in open standaarden moet worden aangeboden.

- Het ontbreekt binnen het beleidsterrein aan een eenduidig vocabulaire, aan de verschillende begrippen worden meerdere betekenissen gegeven.
- Aanbieders van Open data zijn beperkt bekend met nationale initiatieven rondom Open data
- Er is behoefte aan de doorontwikkeling i.c. uitbreiding en meer samenhang van de 'pas toe of leg uit'-lijst
- Nederland stelt internationaal gezien vanuit de aanbieders hogere eisen aan Open data; zoals aan het formaat waarin de data wordt aangeboden en het belang van de beschrijving van de context waarin de data is verzameld.

2. Registratieproces

- Er is bij overheidsaanbieders behoefte aan een richtlijn met betrekking tot metadata
- De waarde van een Nationale catalogus met een verwijzing (verwijsindex) naar alle beschikbare Open datasets wordt onderkend.

3. Distributieproces

- Hosting van de data bij de bronhouder / eigenaar
- Cloud speelt geen onderscheidende rol ten aanzien van de distributie van Open data
- 3^{de} TBL-niveau wordt als het minimale niveau gezien om data te ontsluiten.

Conclusies

Vanuit de vaststelling dat er geen technische drempels zijn om data te publiceren, komt Logica Business Consulting op basis van voorgaande bevindingen tot de volgende conclusies:

1. Ontwerpprincipes voor inrichting publicatieproces

De best-practice voor de inrichting van het publicatieproces van Open data dient gebaseerd zijn op de volgende ontwerpprincipes:

- *Open data,..... tenzij.* Alle openbare data verzamelt door een overheidsorganisatie in het kader van de uitvoering van een publieke taak zou door overheden beschikbaar moeten worden gesteld als Open data met beperkte tenzij's.
- *'Ruwe-data-nu'-aanpak.* Overheidsorganisaties verzamelen data in het kader van de uitvoering van hun publieke taak. Data wordt gepubliceerd in het formaat waarin de data door de overheidsorganisatie is verzameld en/of bewerkt wordt voor de uitvoering van haar publieke taak
- *Prioritering in publicaties.* Overheidsaanbieders beschikken over meerdere datasets. Vanuit de 'ruwe-data-nu'-aanpak dienen al deze datasets te worden gepubliceerd. Indien vanuit een kosten c.q. capaciteitsperspectief tot prioritering moet worden gekomen is de behoefte bij de (her)gebruikers leidend
- *Metadatering.* Data is verzameld in het kader van een overheidstaak en dus binnen een context en met een betekenis. Derhalve moet de publicatie van

data altijd vergezeld gaan van een beschrijving van de context en betekenis i.c. metadatering

- *Hoogst beschikbare formaat.* Indien overheidsorganisatie een dataset in meerdere formaten beschikbaar hebben, dan dient bij publicatie van de dataset gekozen te worden voor het formaat dat zorgt voor het hoogst mogelijke TBL-niveau. Bij nieuwe datasets is het streven bij voorkeur 3de TBL-niveau.
- *Aanmelding dataset.* Na publicatie van de dataset dient de overheidsorganisatie deze te registreren bij het nationale open dataportaal, data.overheid.nl
- *(Her)gebruikersdialoog.* Na publicatie van de data dient de overheidsorganisatie in dialoog met de (her)gebruikers de verdere (door) ontwikkeling van de dataset op te pakken (met als navigatiepunt 3^{de} TBL-niveau).

Om het publicatieproces bij overheidsaanbieders te faciliteren dient te worden voorzien in een centrale kennis en expertisefunctie voor Open data. Een plek waar zowel aanbieders als (her)gebruikers terecht kunnen voor uitwisseling van kennis, informatie en ervaringen op gebied van Open data. Deze functie is daarnaast belast met de taak van verzamelen en beschrijven van de best-practices en met de (proactieve) communicatie rondom de Open data initiatieven van de (Rijks)overheid.

2. PTLU-lijst en 'Ruwe-data-nu'-aanpak

Logica Business Consulting is van mening dat gezien het huidige ontwikkelniveau van de verschillende aanbieders en op basis van de bevindingen uit het onderzoek de 'ruwe-data-nu'-aanpak het uitgangspunt zou moeten zijn van het Open data initiatief in Nederland. De PTLU-lijst biedt bij het aanbieden van datasets een goed aanknopingspunt voor de keuze van een open standaard. De PLTU-lijst dient alleen te worden uitgebreid en toegankelijker te worden gemaakt voor gebruik in het kader van Open data. Suggesties hiervoor zijn opgenomen in paragraaf 2.3.2.

3. Datavelden voor aanmelding dataset bij data.overheid.nl

Data.overheid.nl fungeert als centraal portaal voor de vindbaarheid van overheidsdatasets. Het nut en de noodzaak van zo'n centrale voorziening, in termen van een centrale catalogus (met metadata) voor de vindbaarheid en verwijzing naar de bronhouder, voor (her)gebruikers, wordt vanuit het onderzoek bevestigd. In paragraaf 2.3.3 worden suggesties gedaan voor metadata. De uitdaging voor de verder ontwikkeling van data.overheid.nl ligt in de toegankelijkheid en het vermijden van dubbelwerk i.c. metadata te exporteren en mappen / linken met individuele overheidsportalen, zodat dubbelingen van invoer (en onderhoud) wordt voorkomen.

Aanbevelingen

Op basis van bovenstaande conclusies doet Logica Business Consulting de volgende aanbevelingen

1. Aanbodproces

1. Inrichten van een centrale kennis en expertisefunctie voor Open data voor de uitwisseling van kennis, informatie en ervaringen op gebied van Open data,

belast met het verzamelen en beschrijven van de best-practices en actief met de communicatie rondom de Open data initiatieven van de (Rijks)overheid.

2. Uitbreiding en toegankelijker maken van de huidige PTLU-lijst,
 3. Opstellen van een variant op het TBL-model in het kader van de Nederlandse context i.c. toevoeging van niveau 3½. Niveau 3½ kent als extra toevoeging dat de gepubliceerde data voorzien moet zijn van een unieke 'identifier', wat van groot belang is voor de duurzaamheid, vindbaarheid en het gebruik van de data door (her)gebruikers.
2. Registratieproces
4. Ontwikkelen van een handleiding waarin het registratieproces op data.overheid.nl wordt beschreven. Uitgangspunt hierbij moet zijn dat de aanlevering van de dataset altijd vergezeld gaat van de beschrijving van de velden in een open formaat.
 5. Onderzoek of een combinatie mogelijk is van OWMS (PTLU-lijst) en DCAT waarbij tevens wordt aangesloten bij de ADMS ontwikkeling om te komen tot een minimale set van metadata.
 6. Op basis van het onderzoek vaststellen en publiceren van een lijst met de minimaal vereiste datavelden (metadata).
 7. Indien een uitbreiding op de lijst met vereiste metadata gewenst of nodig is, bijvoorbeeld in het geval van domeinspecifieke metadata, zorgdragen voor centrale communicatie van deze uitbreiding.
 8. Centrale communicatie over het registratieproces beleggen bij de centrale kennis en expertise functie (zie ook aanbeveling 1).
 9. Herontwerp van het aanmeldformulier van data.overheid.nl op basis van de minimale vereisten met betrekking tot metadata en toevoeging van het metadataveld contextbeschrijving.
 10. Inrichten van een terugmeldfaciliteit voor (her)gebruikers en afnemers waar zij inconsistenties in de metadatering kunnen melden om op deze wijze bij te dragen aan borging van kwaliteit van de data. Aanbeveling is dit centraal te faciliteren bijvoorbeeld via data.overheid.nl.
 11. Benut de vereiste metadata om vergelijkbare datasets beter vindbaar te maken door deze naar elkaar te laten verwijzen.
3. Distributieproces
12. Rondom het centrale portaal komen tot institutionalisering en invulling van de kennis- en expertise functie voor Open data. Als voorbeeld/best-practice hiervoor kan dienen de organisatie rondom het nationale Geo-register. Concreet betekent dit het komen tot een rijksbreed overheidsprogramma bestaande uit verschillende werkgroepen waarin (in dialoog met afnemers en (her)gebruikers) gezamenlijk wordt toegewerkt naar het oplossen van technische, organisatorische en juridische/beleidsmatige uitdagingen.
 13. Faciliteren van een contactoptie richting de originele bronhouder, bijvoorbeeld door contactgegevens vereist te stellen bij het aanmelden van een dataset op data.overheid.nl (zie ook subparagraaf 2.3.3).

14. Verkenning naar de mogelijkheden van shared services Open data om mogelijke problemen met opschaling op te lossen door uitwisseling ('sharen') van (server)capaciteit tussen overheidsorganisaties.

1 Inleiding

1.1 Aanleiding onderzoek

Om haar publieke taak goed uit te kunnen voeren produceert en verzamelt de overheid veel informatie. Een belangrijk deel van deze informatie is openbaar op grond van de Wet Openbaarheid Bestuur (Wob). Deze openbare overheidsinformatie kan voor meerdere doelen en door meerdere doelgroepen worden gebruikt. Dat noemen we hergebruik. Bijvoorbeeld door de overheid zelf om publieke taken efficiënter en effectiever uit te voeren. Daarnaast is deze informatie ook een belangrijke grondstof voor bedrijven om nieuwe producten en diensten, apps, te maken op basis van openbare overheidsinformatie. Daarnaast kan openbare overheidsinformatie bijdragen aan transparantie en het versterken van burgerschap doordat burgers over meer informatie beschikken.

Deze overheidsdata kan –meestal na verrijking door marktpartijen- commerciële waarde krijgen en daarmee bijdragen aan economische groei. Meer en meer landen bieden overheidsdata vrij toegankelijk aan. Ook in Nederland werken diverse publieke en private partijen aan de toegang van data. De Nederlandse overheid bevordert de beschikbaarheid van Open data; het actief en herbruikbaar aanbieden van openbare overheidsinformatie bijvoorbeeld via het open dataportaal data.overheid.nl. Zoals aangekondigd in de Digitale Agenda.nl en in de brief 'Hergebruik en Open data'¹ zijn de ministeries van Economische Zaken, Landbouw en Innovatie (EL&I) en Binnenlandse Zaken en Koninkrijksrelaties (BZK) van start gegaan met een programma Open data. Vanuit dit programma wordt zowel de aanbod, als de gebruikerskant van Open data de komende jaren gestimuleerd.

Data.overheid.nl loopt in de praktijk tegen een aantal belemmeringen aan. Het Forum Standaardisatie (het Forum), ondersteund door het Bureau Forum Standaardisatie (BFS) wil ondersteuning bieden in de vorm van een onderzoek, resulterend in een advies. Hiervoor wil het Forum een onderzoek laten uitvoeren dat resulteert in een advies voor het wegnemen van belemmeringen, gerelateerd aan het gebruik van standaarden. Logica Business Consulting zal dit onderzoek in opdracht van het Forum uitvoeren.

Het uit te voeren onderzoek heeft als doel inzicht te geven in de ervaren technische belemmeringen gerelateerd aan de publicatie van data door publieke partijen en in de mogelijke oplossingen hoe deze belemmeringen kunnen worden weggenomen. Het betreft hier de volgende belemmeringen:

1. Publieke partijen (potentiële aanbieders) weten niet in welk formaat hun data het best gepubliceerd kan worden
2. Potentiële gebruikers kunnen gepubliceerde bestanden met Open data niet vinden
3. Potentiële gebruikers hebben geen idee wat de betekenis van de aangeboden data is en daarmee in welke context hergebruik toegevoegde waarde krijgt

¹ TK 32 802 nr. 2, 31 mei 2011

4. Publieke partijen hebben aangegeven in een inventarisatieronde dat ze aarzelen (terughoudend zijn) om data aan te bieden, omdat hun ICT configuratie niet geschikt is om aan derden grootschalig data ter beschikking te stellen.

Randvoorwaarde voor een succesvol onderzoek is dat de geformuleerde aanbevelingen dusdanig concreet en toepasbaar zijn zodat deze aanbevelingen als input dienen voor het opstellen van een 'checklist' voor het aanleveren en registreren van open datasets aan data.overheid.nl.

1.2 Centrale onderzoeksvraag

Voor het uitvoeren van dit onderzoek is door het Forum de volgende centrale onderzoeksvraag geformuleerd:

Formuleer, op basis van de vragen en analyse van een aantal cases, voorstellen om het aanbod-, registratie- en distributieproces van Open data te verbeteren en stel op basis daarvan een handreiking op aan partijen die betrokken zijn bij publicatie, registratie en gebruik van Open data

Hierbij opmerkend dat dit onderzoek zich focust op aanbevelingen met betrekking tot te gebruiken standaarden, voor herbruikbaarheid, vindbaarheid en wijze waarop de ontsluiting/infrastructuur wordt ingericht vanuit een technische invalshoek vanuit de aanbodzijde. Juridische en beleidmatige aspecten behoren niet tot de scope van het onderzoek.

De handreiking heeft als doel organisaties te ondersteunen met het verder professionaliseren van het aanbod en de distributie van Open data. De handreiking maakt deel uit van het groter geheel van documenten die door het Forum in de komende maanden wordt opgesteld. De inhoud van de op te leveren handreiking is verdere concretisering van de best-practices op basis van het onderzoek of te wel op basis van de casestudy (casusbeschrijvingen en internationale benchmark).

Om de centrale onderzoeksvraag te beantwoorden dienen de onderzoekers van Logica Business Consulting zich eerst een oordeel te vormen - op basis van een casestudy i.c. casusbeschrijvingen en een internationale benchmark - over de best-practices op de volgende aandachtsgebieden c.q. aspecten:

- A. Vindbaarheid (hoe zorg ik dat mijn data gevonden wordt?)
- B. Herbruikbaarheid (in welk formaat publiceer ik mijn data?)
- C. De wijze waarop de ontsluiting/infrastructuur wordt ingericht (via welke systematiek publiceer ik mijn data?).
- D. EU-beleidskaders
- E. Lessons learned webstandaarden
- F. Internationale ontwikkelingen en best-practices (internationale benchmark).

Aandachtsgebieden E en F dienen als een toetsingskader voor de aandachtsgebieden A, B, C en D. Hiermee kan duiding en betekenis worden gegeven aan de bevindingen op de aandachtsgebieden A, B, C en D.

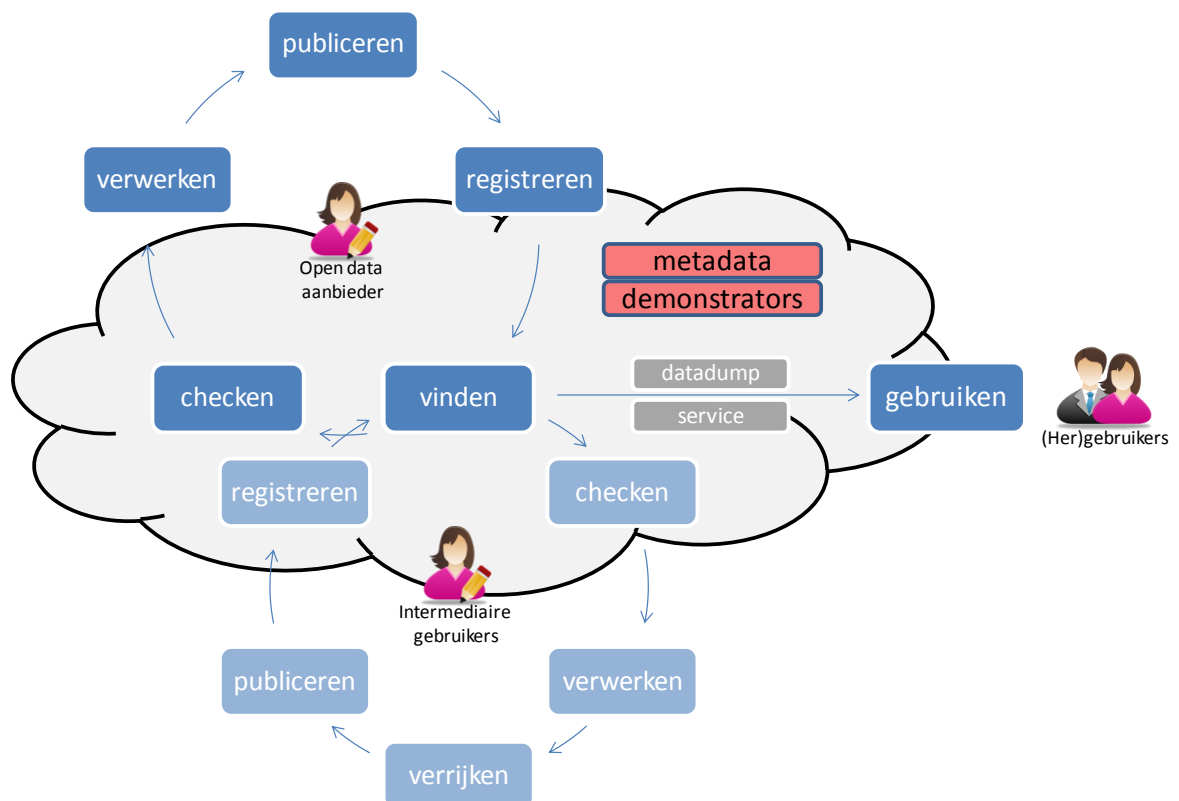
De resultaten van het onderzoek worden aangeboden aan de opdrachtgever van data.overheid.nl, het ministerie van BZK. De resultaten van aandachtsgebied C worden gecommuniceerd met BZK en DGOBR.

1.3 Scope van het onderzoek: de vraag en aanbodketen

De scope van het onderzoek is de waardeketen van Open data. Schematisch is dit vastgelegd in de onderstaande architectuurplaat. In deze architectuurplaat worden de drie belangrijke elementen die datasets definiëren in samenhang met elkaar verbonden:

- Vindbaarheid (hoe zorg ik dat mijn data gevonden wordt?)
- Herbruikbaarheid (in welk formaat publiceer ik mijn data?)
- De wijze waarop de ontsluiting/infrastructuur wordt ingericht (via welke systematiek publiceer ik mijn data?).

In de architectuurplaat wordt de vraag- en aanbodketen van Open data geschetst.



De publieke partijen (aanbieders van Open data) hebben een belangrijke rol als het gaat om de publicatie van data maar in onze visie ook in het aanjagen van het gebruik van Open data. Deze aanjaagfunctie bestaat uit het in Europees verband werken aan metadata waarmee Open data gevonden kan worden. Maar in onze visie heeft de overheid ook een taak als het gaat om het tonen van de mogelijkheden van het gebruik van Open data door zogenaamde demonstrators beschikbaar te stellen. Met name het concept van aanroepbare webservices is voor niet-ICTers moeilijk voor te stellen. Een demonstrator is een visualisatie die het mogelijk maakt om het webservices concept (dat als zodanig geen eigen gebruikers interface of 'voorkant' heeft) en gebruiksmogelijkheden op een eenvoudige manier uit te leggen en te tonen. Vanuit het gebruikersperspectief gaat het om het kunnen vinden van de Open data en het gebruiken van de Open data. Eventueel ook het verrijken van de data (door intermediaire gebruikers) en het publiceren van de (verrijkte) data.

De systematiek waarop de data wordt gepubliceerd bestaat in bovenstaande architectuurplaat uit de volgende processtappen:

- Dataset kan nadat deze gevonden is, er een juridische check heeft plaatsgevonden en de data op een geaggregeerd niveau is verwerkt worden gepubliceerd
- Aanbieder (data Publisher) voegt metadata toe en registreert de dataset (bijvoorbeeld bij data.overheid.nl): het 'wat' en 'hoe' is onderwerp van dit onderzoek
- Data wordt gepubliceerd als een datadump in geval van niet real-time data of als een aan te roepen service, op internet via cloud of eigen hosting: 'systematiek' is onderwerp van dit onderzoek
- De service is bereikbaar via REST of SOAP: 'formaat' is onderwerp van dit onderzoek. De service maakt rechtstreeks gebruik van brondata of van de gepubliceerde datadump
- De datadump is bereikbaar via REST of SOAP: 'formaat' is onderwerp van dit onderzoek
- De datadump wordt opgeslagen als bijvoorbeeld CSV, XML, KML etc.: 'standaarden' zijn onderwerp van dit onderzoek
- Gebruiker kan in data.overheid.nl databronnen zoeken. Via deze catalogus kan de gebruiker locatie en metadata van de bronnen vinden
- Gebruiker kan service en datadump benaderen op gevonden locatie, via de gedefinieerde protocollen.

1.4 Inhoudelijke begrippen

1.4.1 Open data

Dit onderzoek over Open data en standaarden gaat uit van bestaande wet- en regelgeving. De kernbegrippen voor dit onderzoek zijn gedefinieerd op basis van de beleidsbrief van minister Donner².

² TK 32 802 nr. 2, 31 mei 2011

Open overheidsdata

Overheidsinformatie is een breed begrip. Het gaat om bij de overheid berustende informatie die betrekking heeft op een bestuurlijke aangelegenheid, dat wil zeggen: gegenereerd of verzameld is in het kader van de uitvoering van een overheidstaak. In Nederland bepaalt de Wet openbaarheid van bestuur (Wob) welke informatie openbaar is.

Ook open overheidsdata (=Open data) valt onder het brede begrip overheidsinformatie. Open data zijn bronnen van onbewerkte overheidsinformatie die:

- Openbaar zij op grond van de Wob
- Bekostigd zijn uit publieke middelen, beschikbaar gesteld voor de uitvoering van die taak
- Bij voorkeur voldoen aan 'open standaarden'
- Bij voorkeur computerleesbaar zijn, zodat zoekmachines informatie in documenten kunnen vinden.

Open data wordt gezien als een grondstof voor innovatie vanwege het beoogde gratis hergebruik. Er kunnen diensten worden ontwikkeld die anders te duur of technisch niet realiseerbaar zijn; het biedt inzichten voor bestuur, planning en wetenschap; tools voor alternatieve besluitvorming, en nieuwe handelingsperspectieven voor burgers en organisaties.

En passant levert openheid daarnaast mogelijk een bijdrage aan een meer transparante overheid, omdat burgers desgewenst meer inzicht krijgen in de gegevens waarop beleid is gebaseerd.

Interoperabiliteit en standaarden worden gezien als een belangrijke randvoorwaarde om met Open data aan de slag te kunnen.

(On)gestructureerde data en metadata

Gestructureerde data omvat adresboeken, configuratieparameters, financiële transacties, technische tekeningen, etc. De data is met behulp van een richtlijn (zoals bijvoorbeeld XML) gestructureerd. Voorbeelden van ongestructureerde data zijn documentaire informatie (wet- en regelgeving en officiële publicaties), tekstuele documenten, spreadsheets, presentaties en webpagina's. Dit soort informatie wordt traditioneel buiten een relationele database opgeslagen. De reden hiervoor is dat de opbouw en structuur van dit soort informatie niet in een relationeel datamodel zijn vast te leggen.

De metadata betreft de beschrijving van de gegevens en niet de gegevens zelf. Metadata is belangrijk om het overzicht te bewaren van de beschikbare gegevens en om gemakkelijker naar de echte gegevens te kunnen zoeken. Het laat de gebruikers toe te weten welke gegevens bestaan, bij wie ze beschikbaar zijn, hoe deze gebruikt kunnen worden, na te gaan of de dataset dekkend is, etc. Metadata van ongestructureerde data is zelf weer gestructureerde data.

‘Vrij bruikbaar’

De omstandigheid dat data en datasets doorgaans ‘vrij bruikbaar’ zijn betekent niet dat de gegevens en gegevensverzamelingen per definitie ook rechtenvrij zijn. Veel auteurs-, nabuur- en databankrechtelijk beschermde, openbare overheidsinformatie kan zonder dat daartoe een verzoek dient te worden gedaan, al worden hergebruikt. Dit geldt in de eerste plaats voor de zogenoemde basisinformatie van de democratische rechtsstaat (wetten, besluiten en verordeningen etc.), maar ook alle overige door of vanwege de openbare macht openbaar gemaakte informatie mag vrijelijk worden hergebruikt, tenzij de openbare macht ten behoeve van zichzelf de intellectuele eigendomsrechten heeft voorbehouden³. Wanneer sprake is van een voorbehoud moet een verzoek tot hergebruik worden gedaan. Het overheidsorgaan mag dan deze rechten, ook bij elektronische vertrekking, niet voorbehouden.

1.4.2 Open standaarden

Standaarden zijn afspraken over de vorm van de uitwisseling van gegevens. Standaarden kunnen betrekking hebben op drie niveaus:

- Organisatorisch: afspraken over het overheidsbreed hanteren van eenduidige principes en uitgangspunten voor betere interoperabiliteit, afspraken over regelgeving, bedrijfsprocessen en uitvraagmomenten
- Semantisch: afspraken over de eenduidige betekenis van gegevens in ketens en werkprocessen
- Technisch: afspraken over transport en logistiek van de uitwisseling
- Een andere term voor standaard is ‘norm’ (NEN) of ‘recommendation’ (W3C). Standaarden kunnen gesloten of open zijn.

Kenmerkend voor open standaarden is dat er geen barrières zijn aan het gebruik van de standaarden door ICT-gebruikers en –aanbieders. Open standaarden staan tegenover gesloten standaarden die wel (potentiële) barrières kennen. Een standaard is volledig ‘open’ als:

- De standaard is goedgekeurd en zal worden gehandhaafd door een non-profit organisatie, en de lopende ontwikkeling gebeurt op basis van een open besluitvormingsprocedure die toegankelijk is voor alle belanghebbende partijen (consensus of meerderheidsbeschikking enz.)
- De standaard is gepubliceerd en over het specificatiedocument van de standaard kan vrijelijk worden beschikt of het is te verkrijgen tegen een nominale bijdrage. Het moet voor een ieder mogelijk zijn om het te kopiëren, beschikbaar te stellen en te gebruiken om niet of tegen een nominale prijs
- Het intellectuele eigendom – met betrekking tot mogelijk aanwezige patenten – van (delen) van de standaard is onherroepelijk ter beschikking gesteld op een ‘royalty-free’ basis
- Er zijn geen beperkingen omtrent het hergebruik van de standaard.

³ *kamerstukken II 2004/2005, 30188, nr 3, p.8*

Naast openheid geldt nog een aantal aanvullende criteria voor de opname van een standaard op de lijsten.

1.5 Inrichting van het onderzoeksproces

Het onderzoek moet leiden tot voorstellen om het aanbod- en distributieproces open data te verbeteren en tot een handreiking. De centrale onderzoeksvraag en de oordeelsvragen zijn beantwoord op basis van een casestudy (kwalitatief onderzoek). De casestudy behelst casusbeschrijvingen, een internationale benchmark en geïdentificeerde best-practices. Op basis van een analyse hiervan zijn door de onderzoekers aanbevelingen geformuleerd. De uitkomsten en aanbevelingen zijn in de vorm van stellingen door de onderzoekers via een G(roup)D(ecision)R(oom)-sessie voorgelegd aan de medewerkers van de organisaties waarvoor een casusbeschrijving is gemaakt. Deze medewerkers hebben hierop i.c. de best-practices en de aanbevelingen van de onderzoekers, kunnen reageren. De resultaten van deze GDR-sessie zijn uitgewerkt in een zogenaamd Syntheserapport.

Op basis van de uitkomsten uit het Syntheserapport hebben de onderzoekers de centrale onderzoeksvraag beantwoord en wordt een handreiking opgesteld.

Het onderzoeksproces is door de onderzoekers gefaseerd ingericht. De volgende onderzoeksfasen zijn onderscheiden:

- Fase 0: Opstellen Onderzoekskader
- Fase 1: Casestudy
- Fase 2: Synthese
- Fase 3: Voorstellen
- Fase 4: Handreiking.

De casusbeschrijvingen zijn 'ter wederhoor' voorgelegd aan de geïnterviewden. Het 'wederhoor' had betrekking op de juistheid, volledigheid en adequaatheid van de casusbeschrijving.

Het onderzoeksproces is in detail beschreven in het onderzoekskader. Het onderzoekskader is opgenomen in de onderzoekbijlage bij dit onderzoeksrapport.

1.6 Onderzoekspopulatie

De casestudy vormt de basis van het onderzoek. De casestudy bestaat uit twee onderdelen namelijk het opstellen van casusbeschrijvingen en het uitvoeren van een internationale benchmark. Op basis hiervan zijn (nationale en internationale) best-practices gedefinieerd en aanbevelingen geformuleerd. Via de beschrijving van in totaal acht casus hebben de onderzoekers zich een beeld gevormd van de feitelijke situatie en de best-practices in Nederland met betrekking tot het aanbod- en distributieproces Open data. In onderstaande tabel zijn het aantal casusbeschrijvingen vanuit de aanbodkant opgenomen.

Nr.	Organisatie	Datatypering	Omschrijving belang i.v.m. opdracht
1.	VU/ KOOP	Wetgeving Officiële Overheids Publicaties, wetten.overheid.nl	Voorbeeld casus voor gebruik van linked data (5*) Casus voor opzet/ toepassing/toetsing onderzoekresultaten
2.	CBS	Statistieken	Casus voor opzet/toepassing/ toetsing op te zetten raamwerk
3.	I&M	http://bag.vrom.nl/de_b ag_gebruiken/gebruik_d oor_private_partijen	Casus voor toetsing op te leveren raamwerk aan huidige werkwijze basisregistratie
4.	NDW	Data over het wegennet	Casus voor toepassing/ toetsing onderzoekresultaten
5.	Provincie Zuid Holland/ IPO	Geodata	Casus voor opzet/ toepassing en toetsing onderzoekresultaten.
6.	Gemeente Amersfoort/ Gemeente Rotterdam	Data betreffende bibliotheken, sport & recreatie, GGD en mobiliteit	Toetsing van de herbruikbaarheid van Open data in de voorgestelde vorm
7.	Planbureau voor de leefomgeving	Dataset leveranciers	Casus voor toepassing/ toetsing onderzoeksresultaten

In het onderzoek is het gebruikersperspectief betrokken door middel van drie organisaties. Het gebruikersperspectief is vastgelegd in één casusbeschrijving. In onderstaande tabel zijn de organisaties opgenomen die gezamenlijk de vraagkant casusbeschrijving vormen.

Nr.	Organisatie	Datatypering	Omschrijving belang i.v.m. opdracht
8.	Antwoord voor Bedrijven	Site waar afnemers Open data kunnen vinden	Afnemer van Open data
	Reed Elsevier	Afnemer van datasets, onderwijs, geneeskunde, etc.	Afnemer van Open data
	Open Sahara's	Representanten van potentiële gebruiker(s)	Afnemer van Open data

De acht casusbeschrijvingen zijn in de onderzoeksbijlage bij dit eindrapport opgenomen.

Naast de casusbeschrijvingen is ook een internationale benchmark uitgevoerd. Naast de Verenigde Staten en het Verenigd Koninkrijk als voorlopers, zijn ook Canada, in verband met een recent gestarte Open data pilot (data.gc.ca), en Nieuw Zeeland betrokken in de internationale benchmark.

De benchmark is uitgevoerd op basis van deskresearch, telefonische interviews (één per land) en de beoordeling van een concreet project per land. Voor wat betreft de projecten is gefocust op projecten die gerelateerd zijn aan een Open data portaal of een Overheidsregister.

Nr.	Organisatie	Datatypering	Omschrijving belang i.v.m. opdracht
1.	Open Data USA	data.gov	Kennis over Open data leverancier
2.	Open Data Canada	data.gc.ca	Kennis over Open data leverancier
3.	Open Data UK	data.gov.uk	Kennis over Open data leverancier
4.	Open Data Nieuw Zeeland	Leverancier	Kennis over Open data leverancier

De resultaten uit de internationale benchmark (inclusief het Syntheserapport) zijn ter consultatie voorgelegd aan het Open data initiatief in België. Relevante opmerkingen zoals gemaakt zijn verwerkt in het eindrapportage.

1.7 Leeswijzer

De achtergrond en aanleiding tot het voorliggende onderzoek is beschreven in hoofdstuk 1. In hoofdstuk 2 en 3 worden de resultaten vanuit het onderzoek gepresenteerd, te weten:

- Hoofdstuk 2: Beantwoording van de Centrale onderzoeksvraag (bevindingen, conclusies en aanbevelingen)
- Hoofdstuk 3: Oordeel per aandachtsgebied (vanuit het perspectief van de aanbieders, (her)gebruikers en internationaal).

In hoofdstuk 2 wordt in paragraaf 2.2. een samenvatting gegeven van de belangrijkste bevindingen uit het onderzoek. Dit is gedaan op basis van een uitgebreide beschrijving in hoofdstuk 3. De rode draad vanuit de casusbeschrijvingen zijn per aandachtsgebied vanuit de verschillende perspectieven – aanbieders, (her)gebruikers en internationaal - weergegeven in de verschillende paragrafen van hoofdstuk 3.

Alle onderzoeksproducten zijn opgenomen als bijlagen bij dit rapport. In de betreffende onderzoeksbijlage zijn de volgende documenten opgenomen:

- Internationale benchmark
- Syntheserapport
- Casusbeschrijvingen
- Additionele vragen onderzoekskader
- Onderzoekskader.

2 Beantwoording van de Centrale onderzoeksvraag

2.1 Centrale onderzoeksvraag

Voor het uitvoeren van dit onderzoek is door het Forum de volgende centrale onderzoeksvraag geformuleerd:

Formuleer, op basis van de vragen en analyse van een aantal cases, voorstellen om het aanbod-, registratie- en distributieproces van Open data te verbeteren en stel op basis daarvan een handreiking op aan partijen die betrokken zijn bij publicatie, registratie en gebruik van Open data

Hierbij opmerkend dat dit onderzoek zich focust op aanbevelingen met betrekking tot te gebruiken standaarden, voor herbruikbaarheid, vindbaarheid en wijze waarop de ontsluiting/infrastructuur wordt ingericht vanuit een technische invalshoek vanuit de aanbodzijde. Juridische en beleidmatige aspecten behoren niet tot de scope van het onderzoek.

In onderstaande paragraaf 2.2 wordt eerst ingegaan op de belangrijkste conclusies vanuit het onderzoek waarna in paragraaf 2.3 op basis van deze conclusies een aantal aanbevelingen worden gedaan.

2.2 Belangrijkste bevindingen uit het onderzoek

In onderstaande paragrafen 2.2.1 tot en met 2.2.3 wordt per proces – respectievelijk aanbod, registratie en distributie - ingegaan op de belangrijkste bevindingen vanuit het onderzoek c.q. de beelden uit de verschillende aandachtsgebieden c.q. aspecten met betrekking tot de casusbeschrijvingen, de internationale benchmark en het Syntheserapport.

2.2.1 Aanbodproces

De belangrijkste bevindingen vanuit het onderzoek ten aanzien van het aanbodproces:

- Er zijn geen technische drempels om data te publiceren
- Er zijn grote verschillen qua ontwikkelingsniveau tussen de verschillende aanbieders van Open data
- De discussie over Open data kent verschillende werelden: die van overheidsaanbieders en die van de meer academische en de beleidswereld. Overheidsaanbieders zien het groeipad naar het 3de niveau van TBL als haar verantwoordelijkheid, terwijl vooral de meer academische wereld streeft naar het 4de en 5de TBL-niveau. Op beleidsmatig niveau is slechts de uitspraak gedaan dat informatie bij voorkeur in open standaarden moet worden aangeboden.
- Het ontbreekt binnen het beleidsterrein aan een eenduidig vocabulaire, aan de verschillende begrippen worden meerdere betekenissen gegeven.
- Aanbieders van Open data zijn beperkt bekend met nationale initiatieven rondom Open data
- Er is behoefte aan de doorontwikkeling i.c. uitbreiding en meer samenhang van de 'pas toe of leg uit'-lijst

- Nederland stelt internationaal gezien vanuit de aanbieders hogere eisen aan Open data; zoals aan het formaat waarin de data wordt aangeboden en het belang van de beschrijving van de context waarin de data is verzameld.

Geen technische drempels

Belangrijkste bevinding uit het onderzoek is dat er geen technische drempels zijn om data te publiceren. Zowel in de casusbeschrijvingen alsmede tijdens de Synthesefase is duidelijk aangegeven dat er voor overheidsorganisaties geen technische drempels zijn voor het publiceren van Open data. De overheidsorganisaties zoals meegenomen in het onderzoek zijn bekend met het vraagstuk, de technische uitdagingen die dit vraagstuk met zich mee brengt en de wijze waarop tot een technische realisatie kan worden gekomen. Belangrijkste redenen die aangegeven worden voor het ontbreken van technische drempels bij de publicatie van Open data zijn:

- Open data stelt geen hoge eisen aan technische infrastructuur, systemen en applicaties
- Overheidsorganisaties beschikken over een adequate technische infrastructuur waarmee tot publicatie kan worden gekomen
- Indien er nog geen adequate technische infrastructuur is dan geven de organisaties aan op de hoogte te zijn welke technische instrumenten hiervoor benodigd zijn en dat deze op de markt te verkrijgen zijn.

Als enig mogelijk technische probleem wordt schaalbaarheid genoemd. Een dataportaal is iets anders dan een website, aldus de verschillende overheidsorganisaties betrokken bij het onderzoek. Met betrekking tot de schaalbaarheid wordt aangegeven dat de meeste organisaties het technisch beheer van de infrastructuur hebben uitbesteed zodat mogelijke opschaling op een relatieve eenvoudige wijze is te realiseren.

Daarnaast is opgemerkt dat in het proces van de publicatie van de Open data reeds al in het begin van het publicatieproces (bij de 'creatie') van Open data rekening moet worden gehouden met de technische (ver)eis(t)en om te kunnen publiceren en tevens met de mogelijke metadata voor registratie. Dit is geen technisch probleem maar een organisatorisch vraagstuk van het publicatieproces.

Verschillen tussen de aanbieders in ontwikkelingsniveau

De in het onderzoek meegenomen overheidsorganisaties kennen verschillende ontwikkelingsniveaus. Er zijn een aantal organisaties die ver gevorderd zijn en bijvoorbeeld al het merendeel van de data op Tim Berners-Lee (TBL)-niveau 3 (zie bijlage IV voor uitleg over dit vijf sterren model) van publiceren (bijvoorbeeld het Planbureau voor de Leefomgeving, ministerie van Infrastructuur & Milieu/BAG en het CBS). Daarnaast zijn er organisaties die nog aan het begin staan van het publiceren van Open data (zoals bijvoorbeeld de gemeente Amersfoort). Afhankelijk van het ontwikkelingsniveau van de betreffende organisatie wordt nagedacht over andere zaken ten aanzien van de publicatie van Open data. Onafhankelijk van het ontwikkelingsniveau waarin overheidsorganisaties zich bevinden geldt dat ze nu

onafhankelijk van elkaar instrumenten, beleid, processen etc. ontwikkelen om data te publiceren.

Verschillende werelden: die van aanbieders en die van meer academische en de beleidswereld

Er is een verschil in kennis van en meningen over Open data bij de daadwerkelijke overheidsaanbieders en de meer academische/beleidswereld.

Vanuit de overheidsaanbieders (en dan specifiek de gemeenten) wordt aangegeven dat 'men op weg' is met Open data. Men ziet dit als proces waarvan het merendeel van de overheidsorganisaties nog in de opstart fase zit. Daarnaast zijn de overheidsaanbieders het er over eens dat overheidsorganisaties de verantwoordelijkheid hebben om data op 3^{de} TBL-niveau aan te bieden. Het brengen van data naar niveau 4 en 5 wordt beschouwd als de verantwoordelijkheid van de markt. Verder wordt aangegeven dat data in een context moet worden geplaatst voordat het ter beschikking wordt gesteld aan eventuele (her)gebruikers en afnemers.

De discussie ten aanzien van Open data in de meer academische wereld is al verder gevorderd. Men onderschrijft het streven naar TBL-niveau 3, maar breidt de discussie uit richting het brengen van datasets in de cloud en linked open data (TBL-niveau 5). Op beleidsmatig niveau is slechts de uitspraak gedaan dat informatie bij voorkeur in open standaarden moet worden aangeboden. Op basis van de vergelijking met het proces van de webrichtlijnen moet geconstateerd worden dat het proces om te komen tot Open data langdurig van aard zal zijn. Waarbij de overheidsaanbieders in hun nieuwe rol moeten groeien zowel qua interne processen alsmede met de verschillende datasets.

Verschillende betekenissen

Op basis van de verschillende gesprekken tijdens het onderzoek komt duidelijk naar voren dat binnen het beleidsterrein geen eenduidige vocabulaire wordt gebruikt en niet altijd dezelfde betekenis wordt gegeven aan kernbegrippen.

De introductie van TBL in gesprekken werd als verhelderd ervaren. Het gaf een duidelijk kader waarin men elkaar kon vinden en elkaar kon begrijpen. Het model werd door iedereen omarmd.

Beperkt zicht op nationale initiatieven

De overheidsorganisaties die zijn bevestigd in het onderzoek hebben niet altijd zicht op de initiatieven op het vlak van Open data die vanuit de Rijksoverheid worden gestart. Niet altijd was men bekend met bijvoorbeeld: de rol en functie van het Bureau Forum Standaardisatie, de 'pas toe of leg uit' (PTLU)-lijst, de status van data.overheid.nl, activiteiten van de ICTU, de webrichtlijnen, wat wordt gekenmerkt als Open data, etc.

Doorontwikkelen 'pas toe of leg uit' (PTLU)-lijst

De PTLU-lijst kan voldoen als richtlijn voor formaten van te publiceren datasets. Sommige respondenten dragen enkele nieuwe standaarden voor de lijst aan (bijvoorbeeld JSON, XML, REST, CSV en KML zijn genoemd). Ook melden sommige

overheidsorganisaties dat verbetering van de lijst mogelijk is door de lijst per domein of beleidsterrein op te stellen. Vanuit het onderzoek is aangegeven dat het aanbeveling verdient om te streven naar meer samenhang in de PTLU-lijst.

Nederland stelt hogere eisen

De 'Raw-data-now'-aanpak die blijkens de internationale benchmark in het buitenland wordt toegepast, lijkt niet op veel draagvlak te kunnen rekenen bij Nederlandse overheidsorganisaties. Aanbieders van data zijn geen voorstander van het publiceren van ruwe data (zoals microdata) in de vorm waarin dit is verzameld maar stellen hogere eisen aan kwaliteit i.c. het formaat waarin de data wordt aangeboden en het belang van de beschrijving van de context waarin de data is verzameld vanuit de publieke taak. Afnemers en (her)gebruikers stellen minder eisen aan formaat, maar hechten meer belang aan de kwaliteit van data i.c. contextbeschrijving. Wel wordt vanuit Nederland het 'the best is the enemy of the good'-principe onderkend en dat die extra inspanningen (vanwege hogere eisen aan de kwaliteit) niet moet leiden tot extra kosten.

Daarnaast onderschrijft Nederland het belang van een nationale catalogus met een verwijzing (verwijsindex) naar alle beschikbare Open datasets, uniforme metadatering en de dialoog tussen aanbieder en (her)gebruiker/afnemer.

Toenemend belang Europese beleidskaders

Belangrijkste ontwikkelingen op het EU-niveau zijn de totstandkoming van de standaarden van INSPIRE en de ambities van SEIS (Shared Environmental Information Systems (<http://ec.europa.eu/environment/seis/>) waaronder SENSE (http://www.eionet.europa.eu/events/NRC_IS_2009/SOER_MAKES_SENSE.pdf). Op het gebied van metadata is het Asset Description Metadata Schema (ADMS) initiatief gestart (<https://joinup.ec.europa.eu/asset/adms/description>) als eerste stap richting een federatie van Europese repositories.

In toenemende mate stellen de ontwikkelingen op het gebied van de EU-beleidskaders eisen op nationaal niveau. Dit betekent ondermeer dat de kwaliteit van de datasets en het beschrijven hiervan in de context waarin de dataset tot stand is gekomen steeds belangrijk wordt voor zowel overheidsorganen als publieke organisaties. Daarnaast betekent dit ook de verantwoordelijkheid voor de data goed geregeld moet worden.

Lessen vanuit de webrichtlijnen

In het kader van Open data zijn lessen te trekken vanuit de parallelle ontwikkelingen rondom de webrichtlijnen. Deze ontwikkelingen rondom de webrichtlijnen zijn vergelijkbaar in termen van proces, maar niet inhoudelijk. De discussie rondom de webrichtlijnen had betrekking op doel versus middelen. In het geval van Open data is dit (nog) niet het geval, het doel is hier nog niet duidelijk. Er is een noodzaak onderkend, maar nog geen eenduidig doel gedefinieerd. Dit eenduidige doel zal er op termijn pas komen als het beleidsterrein verder wordt geïnstitutionaliseerd met een duidelijke politieke lobby. Belangrijkste lessen die qua proces vanuit de webrichtlijnen kunnen worden getrokken zijn:

- Zorg dat je in het proces van de creatie van de Open data als overheidsorganisatie al weet wat de vereisten zijn die verderop in de keten worden gesteld en baseer daarop je Open data beleid
- Borg de procesvereisten in het proces van Open data creatie. In het proces van creatie moeten die eisen op een eenvoudige wijze in het proces worden geborgd.
- Zorg dat er voor Open data een eigenaar komt en een lobby die druk kan uitoefenen. Met andere woorden organiseer het beleidsterrein.
- Heb een lange adem in het implementeren van de standaarden ("een jaar of 8").
- Sluit aan bij de reeds bestaande best-practices (niet alles van de grond af aan ontwikkelen).

2.2.2 Registratieproces

De belangrijkste bevindingen vanuit het onderzoek ten aanzien van het registratieproces zijn:

- Behoefte aan een richtlijn met betrekking tot metadata
- De waarde van een Nationale catalogus met een verwijzing (verwijsindex) naar alle beschikbare Open datasets wordt onderkend.

Behoefte aan richtlijn metadata

Er is behoefte aan een richtlijn c.q. leidraad met betrekking tot een minimale set metadata voor Open datasets. De genoemde metadata velden voor deze minimale set zijn: Onderwerp, Type (formaat), Licentie (rechten), Eigenaar (herkomst & betrouwbaarheid), Versie, Actualiteit (publicatiedatum & update frequentie) en Contact.

Het belang van metadata wordt door zowel overheidsorganisaties als afnemers en (her)gebruikers onderkend. Overheidsorganisaties zien metadataring als een lastig onderwerp, met name vanwege de te verwachten werklust die mogelijk gepaard gaat bij het voldoen aan eisen met betrekking tot metadata.

De standaarden OWMS en DCAT zijn genoemd als basis voor de richtlijn / leidraad voor de metadata standaard voor Open data. Verder is het Europese ADMS initiatief een aan te bevelen startpunt hiervoor.

Nationale catalogus met een verwijzing (verwijsindex) naar alle beschikbare Open datasets heeft waarde

De deelnemers aan het onderzoek waren het erover eens dat een centrale, voor iedereen toegankelijke plaats, functionerend als nationale catalogus met een verwijzing (verwijsindex) naar alle beschikbare Open datasets een duidelijke toegevoegde waarde heeft. In de huidige opzet voldoet de portal data.overheid.nl nog niet aan de eisen die gesteld worden aan een dergelijke centrale verwijsindex. Vanuit het onderzoek wordt aangedrongen op doorontwikkeling van data.overheid.nl en worden de volgende kanttekeningen geplaatst: de portal is lastig toegankelijk, vertoont nog te weinig structuur en verschaft onvoldoende duidelijkheid over het aanbod. Overheidsorganisatie vragen om een duidelijke beschrijving (handleiding) hoe data in de catalogus te plaatsen.

Vooralsnog zou de situatie zodanig moeten zijn dat de originele dataset fysiek op locatie staat van de originele bronhouder.

2.2.3 Distributieproces

De belangrijkste bevindingen vanuit het onderzoek ten aanzien van het distributieproces zijn:

- Hosting van de data bij de bronhouder / eigenaar
- Cloud speelt geen onderscheidende rol ten aanzien van de distributie van Open data
- 3^{de} TBL-niveau wordt als het minimale niveau gezien om data te ontsluiten.

Hosting bij data-eigenaar

In het kader van distributie is vanuit het onderzoek aangegeven dat er geen behoefte is aan een 'centraal portaal' waar overheidsdata fysiek te benaderen is. De toegevoegde waarde wordt niet gezien van één locatie waar alle data fysiek samenkomt / gehost wordt. Overheidsorganisaties hebben de voorkeur om de data op locatie van de data-eigenaar te hosten. Ook de (her)gebruikers hebben de voorkeur om de data bij de bronhouders af te nemen.

Zodra (her)gebruikers en afnemers data van verschillende aanbieders willen combineren, is het van belang dat dit probleemloos mogelijk is. Hiervoor is het van belang dat de overheid data op eenduidige wijze distribueert.

Cloud speelt geen rol

Alle overheidsorganisaties geven aan na te denken over het concept van cloud, maar zien hierbij geen directe relatie met het onderwerp Open data. Momenteel wordt de inzet van cloud in het kader van Open data dan ook niet door overheidsaanbieders overwogen.

3^{de} TBL-niveau wordt als het minimale niveau gezien om data te ontsluiten

Overheidsorganisaties die al langere tijd data publiceren (Planbureau voor de Leefomgeving, I&M en het CBS) hebben vastomlijnde werkwijzen en instrumenten om tot ontsluiting te komen. Organisaties die kortgeleden gestart zijn (bijvoorbeeld de gemeente Amersfoort) zien het belang maar missen deze procesvorming vooralsnog. Zij kiezen ervoor dit gaandeweg vorm te geven in een project, veelal startend met een concrete vraag vanuit de markt naar Open data of bestuurlijke ambitie om 'iets' te doen aan Open data en toe te groeien naar 3^{de} TBL-niveau.

Alle overheidsaanbieders gaven aan dat de 'pure' techniek in termen van distributie / infrastructuur niet als een grote uitdaging wordt ervaren. In geval van enorme datasets zoals bijvoorbeeld satellietinformatie is aangegeven dat dit kan leiden tot problemen voor afnemers en (her)gebruikers, omdat deze sets te groot zijn om direct te downloaden, in dat geval moet gezocht worden naar een technische workaround.

Alle partijen zijn het er over eens dat het minimale niveau om data te ontsluiten TBL-niveau 3 dient te zijn. Wel wordt het voorstel omarmt om een 'extra niveau' toe te voegen tussen niveau 3 en 4, op dit niveau zouden datasets unieke 'identifiers', de unieke verwijzende sleutels, moeten krijgen. Unieke 'identifiers' zijn van belang voor de borging van de duurzaamheid, vindbaarheid en het gebruik van de data door

(her)gebruikers. Het 4^{de} en 5^{de} TBL-niveau wordt door de overheidsaanbieders beschouwd als de verantwoordelijkheid van de markt.

2.3 Conclusies

Vanuit de vaststelling dat er geen technische drempels zijn om data te publiceren, komt Logica Business Consulting op basis van voorgaande bevindingen tot de volgende conclusies:

2.3.1 Ontwerpprincipes voor inrichting publicatieproces

Op basis van de bevindingen uit de casusbeschrijvingen, de internationale benchmark en het Syntheserapport concludeert Logica Business Consulting dat de best-practice voor de inrichting van het publicatieproces van Open data gebaseerd moet zijn op de volgende ontwerpprincipes:

- *Open data,..... tenzij.* Alle openbare data verzamelt door een overheidsorganisatie in het kader van de uitvoering van een publieke taak zou door overheden beschikbaar moeten worden gesteld als Open data met beperkte tenzij's.
- *'Ruwe-data-nu'-aanpak.* Overheidsorganisaties verzamelen data in het kader van de uitvoering van hun publieke taak. Data wordt gepubliceerd in het formaat waarin de data door de overheidsorganisatie is verzameld en/of bewerkt wordt voor de uitvoering van haar publieke taak
- *Prioritering in publicaties.* Overheidsaanbieders beschikken over meerdere datasets. Vanuit de 'ruwe-data-nu'-aanpak dienen al deze datasets te worden gepubliceerd. Indien vanuit een kosten c.q. capaciteitsperspectief tot prioritering moet worden gekomen is de behoefte bij de (her)gebruikers leidend
- *Metadatering.* Data is verzameld in het kader van een overheidstaak en dus binnen een context en met een betekenis. Derhalve moet de publicatie van data altijd vergezeld gaan van een beschrijving van de context en betekenis i.c. metadatering
- *Hoogst beschikbare formaat.* Indien overheidsorganisatie een dataset in meerdere formaten beschikbaar hebben, dan dient bij publicatie van de dataset gekozen te worden voor het formaat dat zorgt voor het hoogst mogelijke TBL-niveau. Bij nieuwe datasets is het streven bij voorkeur 3^{de} TBL-niveau.
- *Aanmelding dataset.* Na publicatie van de dataset dient de overheidsorganisatie deze te registeren bij het nationale open dataportaal, data.overheid.nl
- *(Her)gebruikersdialoog.* Na publicatie van de data dient de overheidsorganisatie in dialoog met de (her)gebruikers de verdere (door) ontwikkeling van de dataset op te pakken (met als navigatiepunt 3^{de} TBL-niveau).

De handreiking die organisaties moet ondersteunen met het verder professionaliseren van het aanbod en de distributie van Open data dient opgesteld te worden conform deze ontwerpprincipes. In bijlage I is een voorstel opgenomen voor de inhoudsopgave van deze handreiking.

Om het publicatieproces bij overheidsaanbieders te faciliteren dient te worden voorzien in een centrale kennis en expertisefunctie voor Open data. Een plek waar zowel aanbieders als (her)gebruikers terecht kunnen voor uitwisseling van kennis, informatie en ervaringen op gebied van Open data. Deze functie is daarnaast belast met de taak van verzamelen en beschrijven van de best-practices en met de (proactieve) communicatie rondom de Open data initiatieven van de (Rijks)overheid.

2.3.2 PTLU-lijst en 'Ruwe-data-nu'-aanpak

Logica Business Consulting is van mening dat gezien het huidige ontwikkelniveau van de verschillende aanbieders en op basis van de bevindingen uit het onderzoek de 'ruwe-data-nu'-aanpak het uitgangspunt zou moeten zijn van het Open data initiatief in Nederland. De PTLU-lijst biedt bij het aanbieden van datasets een goed aanknopingspunt voor de keuze van een open standaard. Overheidsorganisaties verzamelen data in het kader van de uitvoering van hun publieke taak en publiceren de data in het formaat waarin deze voor de uitoefening van haar publieke taak is verzameld en/of bewerkt. In het vervolg hierop moet de PTLU-lijst gezien worden als leidraad voor het brengen van data naar het 3^{de} of 4^{de} TBL-niveau. Toepassing van de standaarden op de PTLU-lijst, met uitzondering van standaarden van toepassing op processen (zoals NEN-ISO/IEC27001 voor wat betreft informatiebeveiliging in een netwerkorganisatie) en de documentformaat standaard PDF 1.7 en A-1 (niet computerleesbaar), leiden tot publicatie op 3^{de} en 4^{de} TBL-niveau.

De PTLU-lijst is een uitstekend uitgangspunt voor de formaatkeuze, maar de publicatie van Open data wordt niet bepaald door de formaatkeuze alleen. Qua formaatkeuze kan een overheidsstandaard al het 4^{de} TBL-niveau hebben, maar als geen algemeen gebruiksrecht is verleend, blijft de data op TBL-niveau 0.

In de casusbeschrijvingen, de internationale benchmark en het Syntheserapport zijn suggesties gedaan voor uitbreiding van PTLU-lijst namelijk:

- JSON
- DCAT (W3C): specifieke metadata voor data catalogi (OWMS-compliant)
- Data Cube: voor het publiceren van statistische/tabulaire gegevens
- IPM Open Data (nog te ontwikkelen, beschrijft o.a. hoe je datasets publiceert, hoe ze op DON terecht komen en met welke metadata)
- JuriConnect standaarden (NL): voor identificatie van juridische informatie zoals wetten, officiële publicaties, jurisprudentie, commentaren.
- NEN-3610: om geografische metadata te specificeren
- INSPIRE richtlijn (EU): beschrijving van ruimtelijke en milieu-informatie
- RDF/XML (W3C): semantische standaard voor relaties tussen data
- RDF Data cube vocabulary (Statistische data)
- SKOS (W3C): voor waardelijsten en hun relaties
- OWL (W3C): voor waardelijsten en hun relaties

- MIME: voor het herkennen van het formaat van een gedownload bestand, of voor 'content negotiation' (afhankelijk van de client wordt andere content gepresenteerd)
- HTML5: voor de lay-out van webpagina's
- RDFa/Microtagging: voor het semantisch annoteren van webpagina's, inline metadata
- CoolURIs: content negotiation voor linked data (Browser krijgt HTML pagina, RDF client krijgt RDF metadata), <http://www.w3.org/TR/cooluris/>
- IRI - Internationalised Resource Identifiers
- Language Tags: voor het aangeven van de gebruikte taal (cf. xml:lang, http://en.wikipedia.org/wiki/IETF_language_tag)
- VOID - Vocabulary of Interlinked Datasets
- PROV-O: vocabulair voor het beschrijven van herkomstinformatie ('provenance')
- Datex II
- SOAP (W3C): Simple Object Access Protocol
- REST
- CKAN, HTML 5, standaarden voor (gestructureerde) databases en audio- en videoformaten.
- SDMX (DSD)
- S-BVR, Semantics of Business Vocabulary and Rules
- OAI-PMH, of andere harvesting standaard: om metadata te harvesten
- NEN-3610: om geografische metadata te specificeren.

2.3.3 Datavelden voor aanmelding dataset bij data.overheid.nl

Data.overheid.nl fungeert als centraal portaal voor de vindbaarheid van overheidsdatasets. Het nut en de noodzaak van zo'n centrale voorziening, in termen van een centrale catalogus (met metadata) voor de vindbaarheid en verwijzing naar de bronhouder, voor (her)gebruikers, wordt vanuit het onderzoek bevestigd.

De minimale set van datavelden (metadata) voor de beschrijving van een dataset is:

- Onderwerp
- Type (formaat)
- Licentie (rechten)
- Eigenaar (herkomst & betrouwbaarheid)
- Versie
- Actualiteit (publicatiedatum & update frequentie)
- Contact.

De datavelden (metadata) voor het aanmelden van datasets bij data.overheid.nl zijn adequaat echter met de opmerking dat een dataveld moet worden toegevoegd waarin een contextbeschrijving wordt gegeven. Datasets worden in het kader van een overheidstaak verzameld en derhalve met kennisgeving hiervan (context) te worden

gedistribueerd en te worden voorzien van een bijsluiter waarin de juridische aspecten zijn opgenomen.

De uitdaging voor de verder ontwikkeling van data.overheid.nl ligt in de toegankelijkheid en het vermijden van dubbelwerk i.c. metadata te exporteren en mappen / linken met individuele overheidsportalen, zodat dubbelingen van invoer (en onderhoud) wordt voorkomen.

2.4 Aanbevelingen

In onderstaande subparagrafen wordt per proces – respectievelijk aanbod (het publiceren van ruwe data door een overheidsorganisatie) , registratie (toevoegen van metadata aan de dataset) en distributie (afnemen van een dataset door een (her)gebruiker) – een aantal concrete aanbevelingen gedaan.

2.4.1 Aanbodproces

Op basis van de bevindingen vanuit het onderzoek ten aanzien van het aanbodproces komt Logica Business Consulting tot de volgende aanbevelingen:

15. Inrichten van een centrale kennis en expertisefunctie voor Open data voor de uitwisseling van kennis, informatie en ervaringen op gebied van Open data, belast met het verzamelen en beschrijven van de best-practices en actief met de communicatie rondom de Open data initiatieven van de (Rijks)overheid.
16. Uitbreiding en toegankelijker maken van de huidige PTLU-lijst,
17. Opstellen van een variant op het TBL-model in het kader van de Nederlandse context i.c. toevoeging van niveau 3½. Niveau 3½ kent als extra toevoeging dat de gepubliceerde data voorzien moet zijn van een unieke 'identifier', wat van groot belang is voor de duurzaamheid, vindbaarheid en het gebruik van de data door (her)gebruikers.

2.4.2 Registratieproces

Op basis van de bevindingen vanuit het onderzoek ten aanzien van het registratieproces komt Logica Business Consulting tot de volgende aanbevelingen:

18. Ontwikkelen van een handleiding waarin het registratieproces op data.overheid.nl wordt beschreven. Uitgangspunt hierbij moet zijn dat de aanlevering van de dataset altijd vergezeld gaat van de beschrijving van de velden in een open formaat.
19. Onderzoek of een combinatie mogelijk is van OWMS (PTLU-lijst) en DCAT waarbij tevens wordt aangesloten bij de ADMS ontwikkeling om te komen tot een minimale set van metadata.
20. Op basis van het onderzoek vaststellen en publiceren van een lijst met de minimaal vereiste datavelden (metadata).
21. Indien een uitbreiding op de lijst met vereiste metadata gewenst of nodig is, bijvoorbeeld in het geval van domeinspecifieke metadata, zorgdragen voor centrale communicatie van deze uitbreiding.
22. Centrale communicatie over het registratieproces beleggen bij de centrale kennis en expertise functie (zie ook aanbeveling 1).

23. Herontwerp van het aanmeldformulier van data.overheid.nl op basis van de minimale vereisten met betrekking tot metadata en toevoeging van het metadataveld contextbeschrijving.
24. Inrichten van een terugmeldfaciliteit voor (her)gebruikers en afnemers waar zij inconsistenties in de metadatering kunnen melden om op deze wijze bij te dragen aan borging van kwaliteit van de data. Aanbeveling is dit centraal te faciliteren bijvoorbeeld via data.overheid.nl.
25. Benut de vereiste metadata om vergelijkbare datasets beter vindbaar te maken door deze naar elkaar te laten verwijzen.

Buiten de scope van de onderzoeksvraag worden de volgende aanbevelingen geformuleerd:

26. Opstellen van een overheidsbrede geldende juridische bijsluiter waarnaar de bronhouders verwijzen (te vinden op het centrale portaal).
27. Verkenning van mogelijke instrumenten ('tooling') om toevoegen van metadatering te vereenvoudigen (zie bijvoorbeeld http://www.atlasinfo.nl/documenten/invoeren_teksten_.pdf).

2.4.3 Distributieproces

Op basis van de bevindingen vanuit het onderzoek ten aanzien van het distributieproces komt Logica Business Consulting tot de volgende aanbevelingen:

28. Rondom het centrale portaal komen tot institutionalisering en invulling van de kennis- en expertise functie voor Open data. Als voorbeeld/best-practice hiervoor kan dienen de organisatie rondom het nationale Geo-register. Concreet betekent dit het komen tot een rijksbreed overheidsprogramma bestaande uit verschillende werkgroepen waarin (in dialoog met afnemers en (her)gebruikers) gezamenlijk wordt toegewerkt naar het oplossen van technische, organisatorische en juridische/beleidsmatige uitdagingen.
29. Faciliteren van een contactoptie richting de originele bronhouder, bijvoorbeeld door contactgegevens vereist te stellen bij het aanmelden van een dataset op data.overheid.nl (zie ook subparagraaf 2.3.3).
30. Verkenning naar de mogelijkheden van shared services Open data om mogelijke problemen met opschaling op te lossen door uitwisseling ('sharen') van (server)capaciteit tussen overheidsorganisaties.

3 Oordeel per aandachtsgebied

3.1 Inleiding

Voor het uitvoeren van het onderzoek is gekozen voor een aanpak en werkwijze waarbij de onderzoekers voor de beantwoording van de centrale onderzoeksvraag zich eerst een oordeel vormden over de feitelijke stand van zaken op een aantal aandachtsgebieden. Dit door per casus aan de hand van interviews en documentenanalyse de factfindingsvragen te beantwoorden. De oordeelsvorming per aandachtsgebied vindt plaats op basis van een aantal oordeelsvragen. In dit hoofdstuk worden per aandachtsgebied de oordeelsvragen beantwoord vanuit een aanbiederperspectief, gebruikersperspectief en een international perspectief. Dit betekent dat het aandachtsgebied van de internationale ontwikkelingen is verwerkt in de andere aandachtsgebieden. De uitwerking van de oordeelsvragen voor de internationale benchmark zijn opgenomen in de onderzoeksbijlage i.c. internationale benchmark.

In onderstaande tabel zijn de aandachtsgebieden opgenomen met bijbehorende oordeelsvragen. Aandachtsgebieden E en F dienen als een toetsingskader voor de aandachtsgebieden A, B, C en D. Hiermee kan duiding en betekenis worden gegeven aan de bevindingen op de aandachtsgebieden A, B, C en D.

Nr.	Aandachtsgebied	Nr.	Oordeelsvragen
A.	Herbruikbaarheid (In welk formaat publiceer ik mijn data?)	A1.	Wat zijn de verschillende soorten van formaat waarin data wordt gepubliceerd?
		A2.	Wat is het beste format waarin data kan worden gepubliceerd?
		A3.	Zouden er ten behoeve van Open data nieuwe standaarden op de 'pas toe of leg uit'-lijst moeten worden opgenomen?
B.	Vindbaarheid (Hoe zorg ik dat mijn data gevonden wordt?)	B1.	Wat zijn de verschillende mogelijkheden om data te vinden?
		B2.	Op welke wijze i.c. organisatorisch, technisch en procedureel wordt gewaarborgd dat data gevonden worden?
C.	De wijze waarop de ontsluiting / infrastructuur wordt ingericht (via welke Systematiek publiceer ik mijn data?)	C1.	Welke systematiek wordt gehanteerd om data te publiceren?
		C2.	Met welke systematiek wordt data op een effectieve en efficiënte wijze gepubliceerd?
D.	EU-beleidskaders	D1.	Wat zijn de relevante ontwikkelingen binnen de EU-beleidskaders?
		D2.	Wat betekent dit voor de huidige praktijk?
		D3.	Wat zijn de effecten hiervan op het publiceren, vinden en te hanteren systematiek op korte en langere termijn?

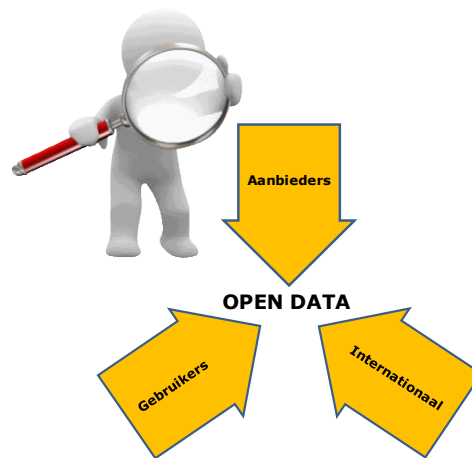
SYNTHESE

E.	Lessons learned webrichtlijnen	E1.	Zijn de ontwikkeling op het gebied van webstandaarden vergelijkbaar met de ontwikkeling bij opendata en standaarden?
		E2.	Welke lessen kunnen worden meegenomen vanuit webrichtlijnen?

F.	Internationale ontwikkelingen en best-practices	F1.	Wat zijn de relevante ontwikkelingen in het buitenland?
		F2.	Zijn er verschillen in de landen die worden meegenomen in het internationale benchmark
		F3.	Zijn er verschillen waar te nemen in de ontwikkeling bij gelijksoortige landen?
		F4.	Wat zijn lessons learned vanuit de benchmark?

3.2 Verschillende perspectieven

Zoals in paragraaf 1.6 is aangegeven worden in het onderzoek 3 verschillende perspectieven gehanteerd. Het perspectief van de aanbieders (7 casusbeschrijvingen te weten: VU/KOOP, CBS, I&M, NDW, Provincie Zuid Holland/IPO, Gemeente Amersfoort/ Gemeente Rotterdam en het Planbureau voor de Leefomgeving).



Het gebruikersperspectief ingevuld door één casusbeschrijving voor drie organisaties (Antwoord voor Bedrijven, Reed Elsevier en Open Sahara). En het internationale perspectief vanuit de internationale benchmark (Canada, UK, VS en New Zealand).

3.3 A. Herbruikbaarheid (in welk formaat publiceer ik mijn data?)

Een oordeel over het aandachtsgebied herbruikbaarheid wordt gevormd aan de hand van de beantwoording vanuit de verschillende perspectieven van de oordeelsvragen (subparagraaf 3.3.1 tot en met 3.3.3):

- Wat zijn de verschillende soorten van formaat waarin data wordt gepubliceerd?
- Wat is het beste formaat waarin data kan worden gepubliceerd?
- Zouden er ten behoeve van Open data nieuwe standaarden op de 'pas toe of leg uit'-lijst moeten worden opgenomen?

3.3.1 Wat zijn de verschillende soorten van formaat waarin data wordt gepubliceerd?



Aanbiederperspectief

Uit de casusbeschrijvingen komt naar voren dat:

- Overheidsorganisaties verschillende soorten van data onderkennen zoals microdata, databasegegevens, registers en indicatoren. Alle overheidsaanbieders waren van mening dat microdata niet als Open data dient te worden aangemerkt. Dit omdat dit gegevens zijn die niet in een context zijn geplaatst. Deze microgegevens worden verzameld op basis van een specifieke overheidstaak en dus in een context. Pas nadat de micro-data in een context is geplaatst (in de vorm van database, registers, indicatoren), wordt gesproken over Open data.
- Datasets kunnen worden getypeerd in termen van het overheidsproces waarin deze verzameld zijn i.c. administratieve, wetgevings- en beleidsprocessen in per proces verschillende formats of standaarden worden gehanteerd. Bijvoorbeeld: datasets verzameld in de administratieve processen dan is CSV gangbaar, bij beleidsprocessen is dat Microsoft Excel.
- Overheidsaanbieders data publiceren in open en gesloten standaarden (bijvoorbeeld CBS)
- Over de verschillende domeinen c.q. overheidsprocessen heen een grote overeenstemming is over de te hanteren formaten en standaarden. Deze komen terug in de PTLU-lijst en zijn aan te merken als algemene formaten en standaarden.
- Per domein, overheidstaak (administratief, wetgeving, geo, onderzoek etc.) in aanvulling op deze verschillende algemene formaten en standaarden gebruik wordt gemaakt van additionele c.q. domein specifieke formaten en standaarden. Redenen hiervoor zijn: (zie additionele vragen)

De volgende formaten zijn gedurende het onderzoek geïdentificeerd: JSON, DCAT, RDF Data Cube, IPM Open Data, JuriConnect standaarden (NL), NEN-3610, INSPIRE richtlijn, RDF-XML, SKOS, OWL, MIME, HTML5, RDFa/Microtagging, CoolURI's, IRI, Language Tags, VOID, PROV-O, Datex II, SOAP, REST, CKAN, SDMX, OAI-PMH, NEN-3610 en S-BVR.

De meest voorkomende standaarden (dus niet domein specifieke formaten) zijn: PDF, Excel en CSV. Voor veel aanbieders is er nog geen grote ambitie om verder te gaan dan deze drie formaten. Dit vanwege enerzijds het verwachte meerwerk om bestanden om te zetten naar andere formaten en anderzijds dat vanuit de (her)gebruikers en afnemers hier vooralsnog niet een substantiële vraag naar is.



Gebruikersperspectief

Uit de casusbeschrijving komt niet echt een beeld naar voren van de verschillende soorten formaat waarin data kan worden aangeboden. Belangrijk vanuit het perspectief van de (her)gebruikers is dat data kan worden gebruikt en toegang tot de data eenvoudig is – zonder het doorlopen van ingewikkelde aanvraagprocedures of zelfs het doorlopen van Wob-procedures.

Gebruikers stellen zich op het standpunt dat overheidsorganisaties data moeten aanleveren in gangbare formaten, zodat mogelijke conversiekosten voor gebruik in de interne systemen worden beperkt. Afgezien van domein specifieke formaten (zoals bijvoorbeeld in het Geo-

domein) hebben de (her)gebruikers en afnemers voorkeuren voor een open formaat i.c. CSV, JSON en XML.

Onder de gebruikers is geen consensus waar de verantwoordelijkheid ligt om databestanden aan te leveren in gewenste open formaten. Een deel van de hergebruikers is van mening dat overheid de databestanden moet aanleveren in CSV, JSON en XML. Het andere deel van de (her)gebruikers is van mening dat het de (her)gebruikers moeten zijn die de dataset om zetten in een (open) formaat. Hiermee wordt de drempel om te publiceren verkleind aldus deze groep van (her)gebruikers. Daarnaast is een gedeelte van de (her)gebruikers geïnteresseerd om de taak van het omzetten van dataformaten voor de overheid (al dan niet betaald) uit te voeren. Gedachte hierachter is dat private partijen dit tegen betaling kunnen doen voor de (her)gebruiker of de overheid als opdrachtgever (externe uitbesteding).



Internationaal perspectief

Uit de internationale benchmark komt naar voren dat in het buitenland niet gekoerst wordt op een standaard set van formaten. Het principe van "Raw data now"-benadering heeft de overhand. Dat wil zeggen publiceren in welk formaat dan ook (zonder sturing van bovenaf). Dan ontstaat een dynamiek waarin de overheidsorganisatie met hergebruikers en afnemers in dialoog gaat over de gewenste formaten.

3.3.2 Wat is het beste formaat waarin data kan worden gepubliceerd?



Aanbiederperspectief

Het aanbieden van data in een machine leesbare structuur, in een non-proprietary formaat (3^{de} TBL-niveau) heeft de voorkeur bij aanbieders. Hierbij opmerkend dat dit een groeipad moet zijn. De standaarden opgenomen in de PTLU-lijst zijn de juiste standaarden hierbij opmerkend dat er per domein op beleidsterreinen een andere formaat kan worden gebruikt.



Gebruikersperspectief

Er is niet echt sprake van één beste formaat waarin data kan worden gepubliceerd vanuit het perspectief van de gebruikers. Er dient hierbij onderscheid gemaakt te worden naar administratieve processen, beleidsprocessen en wetgevingsprocessen. Het gewenste formaat is afhankelijk van het doel waarvoor de data wordt verzameld, de context en gebruik.

Het aanbieden van data in een machine leesbare structuur, in een non-proprietary formaat (3^{de} TBL-niveau) heeft de voorkeur, publicatie op hogere niveaus door de overheid kan leiden tot hogere kosten en (nog meer) drempels. Verwachting is dat de markt en (her)gebruikers sneller en beter de gepubliceerde data verrijken tot 4^{de} en 5^{de} TBL-niveau.



Internationaal perspectief

Het aanbieden van data in een machine leesbare structuur, in een non-proprietary formaat (3^{de} TBL-niveau) heeft de voorkeur. Echter dit moet niet een bottleneck worden ("Best as the enemy of the good"-syndroom). Het gaat er om data te publiceren en in welk formaat maakt op dit moment niet uit.

3.3.3 Zouden er ten behoeve van Open data nieuwe standaarden op de 'pas toe of leg uit'-lijst moeten worden opgenomen?



Aanbiederperspectief

De deelnemers zijn het er over eens dat de overheid i.c. een centrale actor standaarden moet benoemen. Hergebruikers en data afnemers combineren verschillende datasets van verschillende overheidsorganisaties. Vanuit deze optiek moet gekomen worden tot uniformering in standaarden.

De betreffende standaarden dienen als minimale set te worden beschouwd. Conform deze minimale set dienen alle overheidsorganisaties te publiceren. In overleg met (her)gebruikers en afnemers moeten overheidsorganisaties kunnen besluiten om in specifieke domeinen additionele standaarden te gebruiken c.q. toe te voegen aan de deze minimale set.

Hierbij opmerkend dat zo'n lijst vooral niet te groot moet worden omdat door een overvloed aan standaarden de lijst zijn functie verliest: waarneer alles een standaard is een lijst overbodig en ontstaan dezelfde problemen als wanneer er geen lijst is.

Het merendeel van de geïnterviewden was niet bekend met het doel, inhoud en betekenis van de PTLU-lijst. Belangrijkste redenen hiervoor waren:

- Nog nooit eerder op geattendeerd
- Onduidelijke lijst ("Wat wordt er bedoeld?, wat staat er nu eigenlijk?, etc.")

Na uitleg over de lijst waren de geïnterviewden van mening:

- Kijkend met een niet-domein specifieke bril dat:
 - op de PTLU-lijst de juiste (open) standaarden staan
 - de lijst een dynamische karakter moet hebben
 - een aantal veel gebruikte standaarden er aan toegevoegd moeten worden (o.a. XML, JSON)
- Kijkend met een domein specifieke bril dat:
 - aan de PTLU-lijst andere standaarden moeten worden toegevoegd (zoals NDW-standaarden)
 - de PTLU-lijst te algemeen is en beter tot domein (of overheidstaak) specifieke lijsten moet worden gekomen.
 - aan de PLTU-lijst gangbare standaarden moeten toegevoegd (bijvoorbeeld Google maps).

Voor de verdere door ontwikkeling van de PTLU-lijst worden de volgende suggesties gedaan:

- Het dient een dynamische lijst te zijn waarop nieuwe standaarden kunnen worden toegevoegd (procedure hiervoor is onbekend)
- Dient in lijn te zijn met Europese richtlijnen (is op dit moment niet het geval)
- Dient gebruikt te kunnen worden als een leidraad voor overheidsorganisaties (is nu een niet begrijpbare lijst)
- De lijst data-aanbieders helpt met de keuze uit de vele standaarden die er zijn
- Er op toe wordt gezien dat deze lijst wordt gebruikt door data-aanbieders
- Aansluit bij de wensen van de (her)gebruikers en afnemers voor wat betreft gangbare toepassingen (willen Excel).



Gebruikersperspectief

Vanuit het perspectief van de gebruikers bestaat er consensus over dat de PTLU-lijst lijkt te voldoen. De lijst wordt als nuttig gezien als op de lijst formaten zijn terug te vinden, die gebruikt kunnen worden voor het publiceren van een dataset op het derde niveau van TBL. Hierbij opmerkend dat de standaarden niet open standaarden hoeven te zijn maar dat ook gangbare c.q. veel gebruikte standaarden worden opgenomen op deze lijst.

Vanuit het de (her)gebruikers en de afnemers werd het risico geïdentificeerd dat door nu teveel de nadruk te leggen op de (open) standaarden, dit zorgt voor een belemmering voor het publiceren van data. De roep om standaarden vanuit 'de markt' zal toenemen naarmate er meer data gepubliceerd wordt. De PTLU-lijst ondersteunt een laatste stap in een vaak lang bestuurlijk en organisatorisch proces om te komen tot de publicatie van Open data. Aan een versnelling van dit proces of aan een breder aanbod van Open data – waaraan vanuit het perspectief van de (her)gebruikers vooral behoefte is – draagt de lijst in beperkte mate bij.



Internationaal perspectief

In de meeste internationale casusbeschrijvingen wordt:

- Data gepubliceerd in de vorm zoals ze is verzameld of bewerkt door een overheidsorganisatie zodat de data owners i.c. de beleidsmedewerkers of onderzoekers verantwoordelijk voor de dataverzameling, front-end niet veel additionele processen moeten worden ingeregeld om te komen tot publicatie van data.
- De back-end processen verbeteren om herpubliceren en het updaten van de data eenvoudiger te maken. De up-front cost worden geminimaliseerd om weerstaand om tot publicatie over te gaan tegen te gaan. Daar de overheidsorganisaties investeringen om te publiceren als een niet gefinancierd mandaat beschouwen. Pas wanneer de overheidsorganisatie de voordelen ziet om open data te publiceren zijn de betreffende organisaties ook intrinsiek gemotiveerd om de front-end processen op een juiste wijze in te richten.

3.4 B. Vindbaarheid (hoe zorg ik dat mijn data gevonden wordt?)

Een oordeel over het aandachtsgebied Vindbaarheid wordt gevormd aan de hand van de beantwoording vanuit de verschillende perspectieven van de oordeelsvragen (subparagraaf 3.4.1 tot en met 3.4.2):

- Wat zijn de verschillende mogelijkheden om data te vinden?
- Op welke wijze i.c. organisatorisch, technisch en procedureel wordt gewaarborgd dat data gevonden kan worden?

3.4.1 Wat zijn de verschillende mogelijkheden om data te vinden?



Aanbiederperspectief

In de meeste gevallen is de (domein) afnemers bekend waar welke datasets gevonden kunnen worden. Nieuwe potentiële afnemers vinden de datasets door middel van Google (of andere zoekmachines). Tenslotte maken (her)gebruikers en afnemers gebruik van bilaterale afspraken tussen aanbieder en (her)gebruiker (bijvoorbeeld Reed Elsevier en de Hbo-raad).

De aanbieders zien een toegevoegde waarde van een landelijke catalogus. Dit omdat dan op een centrale plek de (her)gebruikers en afnemers een ingang hebben naar alle mogelijke datasets. Technisch gezien zijn er allerlei oplossingen om dit portaal vorm te geven. Belangrijkste aspecten om te komen tot deze centrale portaal zijn: een goede zoekfunctie/register, actualiteit, metadata en verwijzing naar de bron.

Een deel van de deelnemers aan het onderzoek (zoals bijvoorbeeld gemeente Rotterdam) geven aan ook het belang te zien van een organisatie specifieke catalogus met daarbij wel de ambitie om de verschillende catalogi aan elkaar te koppelen om de zoekresultaten voor (her)gebruikers zo volledig mogelijk te maken. Hiermee wordt het principe van "no wrong door" gerealiseerd.

Met betrekking tot metadatering spelen de volgende issues:

- Hoe classificeren we informatie?
- Welke trefwoordenlijst gebruiken we?
- Hoe sluiten we aan op INSPIRE?
- Hoe houden we de werklast hiervoor beperkt?

Als je gezamenlijk overeenstemming hebt op deze punten dan is metadatering niet echt moeilijk, aldus de meeste geïnterviewden. Relevante metadata standaarden zijn:

- OWS
- Dublin Core
- ISO 19115 and ISO 19128
- OAI-PMH: om metadata te harvesten
- NEN-3610: om geografische metadata te specificeren
- DCAT (Data Catalogus).

Vanuit de aanbieders wordt aangegeven dat het van belang is dat er een richtlijn komt met betrekking tot metadata voor de vindbaarheid, bruikbaarheid en betrouwbaarheid van data. Voor de daadwerkelijke vormgeving van zo'n richtlijn zijn de volgende suggesties gedaan, de richtlijn dient:

- Niet restrictief te zijn c.q. de status van de richtlijn is een handleiding c.q. advies
- Kent een minimale set aan metadata die door individuele data-aanbieder aangevuld kan worden of per beleidsterrein of domein kan worden bepaald
- Opgesteld te worden door een centrale/coördinerende actor op het beleidsterrein van Open data
- Aan te sluiten bij EU ontwikkelingen op dit gebied
- In overleg met de (her)gebruikers te worden vastgesteld.

Als mogelijke optie om de kwaliteit van de data te garanderen wordt gesuggereerd om een terugmeldvoorziening in het leven te roepen.



Gebruikersperspectief

Ervaren (her)gebruikers en afnemers binnen een domein kennen de verschillende overheidsorganisaties en weten de weg te vinden naar de bronhouders. Andere opties zijn via Google (of andere zoekmachines) of het opstarten van een Wob procedure bij een ministerie. Intreders op het domein vinden de overheidsorganisaties via "mond-op- mond"-reclame.

Een landelijke Open data portaal wordt als van toegevoegde waarde gezien omdat er dan een centrale plek is waar (her)gebruikers en afnemers een ingang kunnen vinden naar de verschillende datasets. Hierbij opmerkend dat (her)gebruikers en afnemers de betreffende datasets willen afnemen bij de bronhouders (vanwege de mogelijkheid van direct contact met de bronhouder).



Internationaal perspectief

Een breed benoemd voordeel van een open data-portaal is dat gerelateerde gegevens uit verschillende organisaties door gebruikers op een plek kunnen worden gevonden of verzameld ongeacht de bronhouder. Dit maakt het voor gebruikers gemakkelijker om gegevens met betrekking tot een specifieke onderwerp te vinden zonder te hoeven weten welke instanties deze produceren. Aan de andere kant is het voor gebruikers die weten wat ze willen eenvoudiger om naar de betreffende overheidsinstantie te gaan in plaats van deze data te gaan zoeken op een Open data portaal. Om deze tegenstrijdigheid op te lossen is het van belang dat het Open data portaal de datasets op verschillende wijze presenteert i.c. multiple listing of te wel door middel van taxonomies of ontologieën waarmee data kan worden gevonden. Hierbij opmerkend dat multiple listing een (tijds)investering vergt. De wens om dit te doen mag geen excuus worden om het publiceren van data uit te stellen.

3.4.2 Op welke wijze i.c. organisatorisch, technisch en procedureel wordt gewaarborgd dat data gevonden worden?



Aanbiederperspectief

Vinden van Open data kent technisch gezien geen uitdagingen. Het merendeel van de overheidsorganisaties gaf aan dat de betreffende techniek reeds in de organisatie voorhanden is. Eventuele aanvullingen zijn eenvoudig via de markt te krijgen. Kennis met betrekking tot de technische solutions is binnen de organisaties voorhanden.

De problemen betreffen:

- Het formuleren van beleidstandpunten c.q. opstellen van beleid in termen van:
 - gestructureerde/geaggregeerde data of de ruwe data?
 - wel of niet real-time data?
 - wel of niet creëren van een publiek data-archief.
- Daarnaast de organisatie rondom de dataset. Gemaakte beleidskeuzen hebben gevolgen voor de interne werkwijze binnen een organisatie. Wellicht is dit de grootste uitdaging om te komen tot Open data (zie casusbeschrijvingen Zuid-Holland/IPO, Plan bureau van de Leefomgeving, I&M en CBS). Wil een organisatie op een efficiënte en effectieve wijze komen tot het aanbieden van Open data dan dienen de processen rondom een dataset op een dusdanige wijze te worden georganiseerd dat bij de datacreatie (verzameling van microdata) al rekening wordt gehouden met de uiteindelijke publicatie van de betreffende data. Er vanuit gaande dat het gestructureerde data betreft. In het interne proces tot publicatie zijn dan verschillende schakels te onderkennen namelijk: onderzoekers die verantwoordelijk zijn voor de gegevensverzameling en het beheer van de dataset en/of indicatoren, en de text editors, webtechnics en Graphics editors die met behulp van een content management systeem zorg dragen voor de daadwerkelijke publicatie. Belangrijk om hierbij op te merken deze samenwerking met tooling wordt ondersteund. Tools zoals bijvoorbeeld een Wordmarco's waarmee de onderzoekers tijdens hun dagelijkse werkzaamheden al werken in een publicatieformaat en de metadata worden aangemaakt. Op basis van deze output van de onderzoeker kan de teksteditors dan direct aan slag met de publicatiewerkzaamheden. In het productieproces wordt dan geen tijd verloren met conversieslagen. Voorbeelden van zo'n succesvolle werkwijze is beschreven in de casusbeschrijving van het Plan bureau van de Leefomgeving. Ook CBS en I&M hebben de werkprocessen op een dusdanige wijze geregeld dat reeds al aan het begin van proces bij de dataverzameling rekening wordt gehouden met de (ver)eisen van publicatie.



Gebruikersperspectief

Vanuit gebruikersperspectief wordt een technische belemmering benoemd namelijk de enorme datasets zoals satellietinformatie. Dergelijke grote datasets maken het voor sommige gebruikers onmogelijk om direct te downloaden. Hiervoor moet een technische workaroud worden gecreëerd. Echter, ook dit wordt niet als onoverkomelijke drempel ervaren.

Data moeten beschikbaar zijn voor derden. De wijze waarop dit door de overheidsaanbieders dient te worden geregeld is niet de verantwoordelijkheid van de (her)gebruikers en afnemers. Voorkeur bij (her)gebruikers en afnemers is om de data te houden bij de bronhouders. Dit omdat daar de kennis is over de dataset. De wijze waarop het CBS en het Planbureau van de leefomgeving hun datasets ter beschikking

stellen zijn goede voorbeelden hoe de datasets gevonden kunnen worden. Door het gebruik van data zal de vindbaarheid toenemen. Hoe meer data beschikbaar komt en gebruikt wordt, hoe beter gewaarborgd wordt dat data vindbaar is en blijft. Hiervoor is het belangrijk dat de overheid data publiceert.

Daarnaast wordt data.overheid.nl gezien als een prima startpunt voor creatieve ondernemers en voor partijen die context aan data willen toevoegen. Een centraal punt waar inzicht wordt gegeven in de verschillende mogelijke datasets is wenselijk. Overheidsaanbieders zouden dan ook gestimuleerd moeten worden om hun datasets op te nemen in deze nationale catalogus.

De vindbaarheid van data, of beter het gebruik van data, kan gestimuleerd worden door vanuit de aanbieder enige metadata toe te voegen, zoals datum publicatie, wijzigingen ten opzichte van vorige publicaties, contactgegevens en eigenaar.



Internationaal perspectief

Een breed benoemd voordeel van een open dataportaal is dat gerelateerde gegevens uit verschillende organisaties door gebruikers op een plek kunnen worden gevonden of verzameld ongeacht de bronhouder.

Metadata speelt een belangrijke rol bij de vindbaarheid van data. In het algemeen zouden de semantiek, context en beperking van de data beschreven moeten worden om de kans van verkeerd gebruik en misinterpretatie te voorkomen.

3.5 C. De wijze waarop de ontsluiting/infrastructuur wordt ingericht (via welke systematiek publiceer ik mijn data?)

Een oordeel over het aandachtsgebied wijze waarop ontsluiting/infrastructuur wordt ingericht wordt gevormd aan de hand van de beantwoording vanuit de verschillende perspectieven van de oordeelsvragen (subparagraaf 3.5.1 tot en met 3.5.2):

- Welke systematiek wordt gehanteerd om data te publiceren?
- Met welke systematiek wordt data op een effectieve en efficiënte wijze gepubliceerd?

3.5.1 Welke systematiek wordt gehanteerd om data te publiceren?



Aanbiederperspectief

Alle overheidsaanbieders waren van mening dat de 'pure' techniek in termen van ontsluiting/infrastructuur niet als een uitdaging wordt ervaren. Het belangrijkste is dat je als data-aanbieder afspraken maakt over:

- Webservices (welke worden aangeboden)
- Informatiemodellen
- Schaalbaarheid in termen van beschikbaarheid en performance van dataservers in relatie tot de afnamebehoefte.

Bij overheidsaanbieders die starten met Open data wordt een werkwijze gehanteerd waarbij op basis van een concrete vraag vanuit de markt een project wordt gestart.

De werkwijze en de protocollering ontwikkeld in het project wordt via het olievlek principe verder uitgerold in de organisatie.

In verschillende casusbeschrijvingen (TBL-niveau 3) wordt als een kritieke succes factor opgemerkt een geïnstitutionaliseerd omgeving rondom de datasets.

Voorbeelden hiervan zijn Provinciaal register, BAG, NDW, Compendium voor de leefomgeving. Institutionaliseren rondom een dataset heeft betrekking op:

- Overleggremia waarin aanbieder, (her)gebruikers en afnemers zijn vertegenwoordigd en gezamenlijk (iedereen vanuit zijn eigen verantwoordelijkheid) vorm geven aan de verdere ontwikkelingen van de dataset
- Werkgroepen voor het oplossen of verder ontwikkelen van technische standaarden, oplossen distributie problemen, etc.
- Een terugmeldvoorziening met betrekking tot de kwaliteit van de datasets.



Gebruikersperspectief

Vanuit het perspectief van de gebruikers is er niet direct een visie op de te hanteren systematiek. Er wordt erkent dat de Cloud een rol kan spelen, net als een nationale voorziening om data op / mee te publiceren. Benadrukt is dat het publiceren van data belangrijker is dan de systematiek die daarbij gebruikt wordt.



Internationaal perspectief

Bestaande datasets worden gepubliceerd in de vorm waarin ze worden verzameld en/of bewerkt. Vanuit de internationale benchmark komt naar voren dat de back-end processen bij de overheidsorganisaties worden verbeterd om zodoende het herpubliceren en het updaten van de data eenvoudiger te maken. Tot deze aanpak wordt gekomen om de weerstand in een overheidsorganisatie tegen te gaan. Pas wanneer de overheidsorganisatie de voordelen ziet om Open data te publiceren worden de front-end processen in een overheidsorganisatie aangepast.

Uitgangspunt van de 'raw data now' aanpak is dat door (her)gebruikers en afnemers feedback wordt gestimuleerd en op basis hiervan overheidsorganisaties komen tot de juiste focus voor de verbetering van de kwaliteit van Open data. Hierbij opmerkend dat de tegenstanders van Open data van mening zijn dat duidelijkheid over privacy, kwaliteit en semantische issues randvoorwaardelijk zijn voor deze aanpak. Hiermee wordt verkeerd gebruik of misbruik van data namelijk voorkomen.

3.5.2 Met welke systematiek wordt data op een effectieve en efficiënte wijze gepubliceerd?



Aanbiederperspectief

Open data is een principiële keuze. Het proces binnen een overheidsorganisatie voor de productie van Open data moet zodanig ingericht worden dat het niet uit maakt of er vraag of geen vraag naar is, c.q. publicatie moet los staan van waarde in een specifieke context.

Vanwege tijd en kosten is het niet mogelijk om op korte termijn alle sets (openbaar) te publiceren (en zeker niet op het niveau van drie sterren). Je moet dan als organisatie gaan prioriteren. Je zou dan kunnen kiezen voor een aanpak waarbij een overheidsorganisatie bekend maakt wat een organisatie in huis heeft. Op basis van de vraag uit de markt kan je dan datasets selecteren en beschikbaar stellen, met het uiteindelijke doel om alle sets op termijn gepubliceerd te krijgen.

Cloud als een instrument om data op een effectieve en efficiënte wijze te publiceren werd door de overheidsorganisatie niet als zodanig onderkend. Voor hen geldt het principe dat het eigendom (en daarmee de bewerking van de originele data) ligt bij de bronhouder. Derden wordt dan ook niet, via een cloud-toepassing, toegang gegeven tot de originele dataset.

Met betrekking tot schaalbaarheid gaven de overheidsorganisaties aan dat dit een mogelijk probleem zou kunnen zijn maar dat dit is ondervangen door middel van de uitbesteding van het technische beheer. Eventuele opschaling van capaciteit is dan via de externe leveranciers mogelijk. Daarnaast is opgemerkt dat schaalbaarheid niet echt een issue is als data wordt afgenomen door brokers en app-bouwers, die data repliceren en daarmee eindgebruikervraag afdempen.



Gebruikersperspectief

Vanuit het perspectief van de gebruikers is er niet direct een visie op de te hanteren systematiek. Erkent wordt dat de Cloud een rol kan spelen, net als een nationale voorziening om data op / mee te publiceren. Benadrukt is dat het publiceren van data belangrijker is dan de systematiek die daarbij gebruikt wordt.



Internationaal perspectief

Indien overheidsorganisaties te weinig servercapaciteit hebben, dan wordt de suggestie gedaan om de schaalbaarheid te regelen via het centraal portaal (datasets worden op de server van dit portaal gehost). Tevens is het een optie dat overheidsorganisaties zelf onderling servercapaciteit uitwisselen.

3.6 D. Relevante ontwikkelingen binnen de EU-beleidskaders

Een oordeel over het aandachtsgebied ontwikkelingen binnen de EU-beleidskaders wordt gevormd aan de hand van de beantwoording vanuit de verschillende perspectieven van de oordeelsvragen (subparagraaf 3.6.1 tot en met 3.6.3):

- Wat zijn de relevante ontwikkelingen binnen de EU-beleidskaders?
- Wat betekent dit voor de huidige praktijk?
- Wat zijn de effecten hiervan op het publiceren, vinden en te hanteren systematiek op korte en langere termijn?

3.6.1 Wat zijn de relevante ontwikkelingen binnen de EU-beleidskaders?



Aanbiederperspectief

Het merendeel van de overheidsorganisaties volgt de EU-ontwikkelingen op afstand. Een actueel en up-to-date inzicht met betrekking tot de beleidsontwikkelingen is niet aanwezig bij de organisatie meegenomen in het onderzoek. De Digitale Agenda wordt gevolgd en is INSPIRE een belangrijke richtlijn.



Gebruikersperspectief

Binnen de EU wordt gewerkt aan een eigen Open data portal wat de vindbaarheid van data (en tussen landen) mogelijk vergroot.

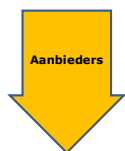
De aandacht van de EU voor Open data leidt tot meer awareness bij Nederlandse overheidsorganisaties. Dat is positief.



Internationaal perspectief

In de internationale benchmark is dit aandachtsgebied niet meegenomen.

3.6.2 Wat betekent dit voor de huidige praktijk?



Aanbiederperspectief

De EU dringt aan om alle overheidsdata te definiëren als Open data. Dit betekent dat gemeenten, semi-overheidsorganisaties, etc. met Open data aan de slag moeten. Dit is in Nederland nog niet het geval. Dit betekent dat nog veel overheidsorganisaties aan "de bak moeten" en dat er snel inzicht moeten komen in mogelijke best-practices van overheidsorganisaties.



Gebruikersperspectief

Mogelijk komt versneld meer data beschikbaar. Het kan werken als een trigger om toegang te krijgen tot datasets.



Internationaal perspectief

In de internationale benchmark is dit aandachtsgebied niet meegenomen.

3.6.3 Wat zijn de effecten hiervan op het publiceren, vinden en te hanteren systematiek op korte en langere termijn?



Aanbiederperspectief

De beleidsuitgangspunten zoals geformuleerd door Eurocommissaris mevrouw Kroes en de PSI Directive vormen de navigatiepunten voor de ontwikkeling van het Open data debat.

Belangrijke elementen hiervoor zijn:

- Alles open, tenzij... . Dit betekent dat het volume van beschikbare datasets naar verwachting de komende jaren exponentieel zal toenemen

- Alle overheidssectoren en semi-overheidsorganisaties vallen onder het beleid. Dit betekent dat gemeenten, provincies, ziekenhuizen, waterschappen en gesubsidieerde instellingen c.q. stichtingen datasets ter beschikking zullen / moeten gaan stellen
- Europese uniformering en koppeling van datasets. Het EU landen beleid zal verder worden geharmoniseerd met als doel dat gelijksoortige datasets uit verschillende landen door (her)gebruikers en afnemers met elkaar kunnen worden gekoppeld.



Gebruikersperspectief

Verwachting is dat er meer awareness komt voor Open data.



Internationaal perspectief

In de internationale benchmark is dit aandachtsgebied niet meegenomen.

3.7 E. Lessons learned webrichtlijnen

Een oordeel over het aandachtsgebied 'lessons learned' webrichtlijnen wordt gevormd aan de hand van de beantwoording vanuit de verschillende perspectieven van de oordeelsvragen (subparagraaf 3.7.1 tot en met 3.7.2):

- Zijn de ontwikkelingen op het gebied van webrichtlijnen vergelijkbaar met de ontwikkelingen bij Open data en standaarden?
- Welke lessen kunnen worden meegenomen vanuit de webrichtlijnen?

3.7.1 Zijn de ontwikkeling op het gebied van webrichtlijnen vergelijkbaar met de ontwikkeling bij opendata en standaarden?



Aanbiederperspectief

Het is vergelijkbaar in termen van proces. Inhoudelijk echter niet. De discussie van webrichtlijnen was een doel-middelen discussie. Dit is niet het geval bij Open data. Het doel is hier nog niet duidelijk. Er is een noodzaak maar nog geen duidelijk doel. Dit doel zal er pas komen als het beleidsterrein wordt geïnstitutionaliseerd met een duidelijke politieke lobby.



Gebruikersperspectief

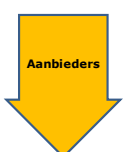
Gebruikers zijn niet bekend met deze discussie.



Internationaal perspectief

In de internationale benchmark is dit aandachtsgebied niet meegenomen.

3.7.2 Welke lessen kunnen worden meegenomen vanuit webrichtlijnen?



Aanbiederperspectief

- Er moeten een hoger doel komen voor Open data, dat is er nu nog niet, behalve dat het moet.

- Zorg dat je in het proces van de creatie van de Open data al weet wat de vereisten zijn die verderop in de keten worden gesteld en baseer daarop je Open data beleid.
- Borg de procesvereisten in het proces van creatie. In het proces van creatie moeten die eisen op een eenvoudige wijze in het proces worden geborgd.
- Zorg dat er voor Open data een eigenaar komt en een lobby die druk kan uitoefen. Met andere woorden organiseer het beleidsterrein.
- Heb een lange adem in het implementeren van de standaarden (een jaar of 8).
- Sluit aan bij de best-practices (niet alles van de grond af aan ontwikkelen).

**Gebruikersperspectief**

Bij de totstandkoming van de webrichtlijnen is veel aandacht besteed aan awareness en het bepalen van open formaten. Suggestie wordt gedaan om op beide vlakken – awareness & open formaten – te leren van de opgedane ervaringen.

**Internationaal perspectief**

In de internationale benchmark is dit aandachtsgebied niet meegenomen.

Bijlage I: Voorstel inhoudsopgave handleiding

Inhoudsopgave

introductie

Doelgroep

1.0 Open data

- 1.1 Waarom Open Data?
- 1.2 Wat is Open Data
- 1.3 Wat is Open?
- 1.4 Welke gegevens heb je het over?
- 1.5 Administratieve, wetgeving en beleidsprocessen
- 1.6 Open data domeinen
- 1.7 Voorbeelden van datasets

2.0 Publicatieproces

- 2.1 Ontwikkel beleid
- 2.2 Projectmatige aanpak
- 2.3 Inventariseer
- 2.4 Stel Open data catalogus samen
- 2.5 Kies dataset (s)
- 2.6 Open Licentie
- 2.7 Kies formaat en standaarden
- 2.8 Hoe gebruik ik een bepaald formaat en standaard?
- 2.9 Maak data beschikbaar (Technische openheid)
- 2.10 Best-practices publicatieproces

3.0 Registratieproces

- 3.1 Metadata
- 3.2 Metadata richtlijnen
- 3.2 Laat het weten!
- 3.3 Procedure centrale portal
- 3.4 Best-practices registratie

4.0 Distributieproces

- 4.1 Inrichting en vormgeving portaal
- 4.2 Gebruikersdialoog
- 4.3 Terugmeldvoorziening
- 4.4 Gebruikerscommunicatie
- 4.5 Best-practices distributie

5.0 Continue leren & verbeteren

- 5.1 Van een projectorganisatie naar beklijving in de organisatie
- 5.2 Verbeteringscyclus
- 5.3 Kennis en Expertise centrum Open data

Woordenlijst

Bijlage

- Bestandsformaten
- Een overzicht van bestandsformaten
- Open bestandsformaten
- Wat Legal (IP) rechten zijn er in Data (bases)

Bijlage II: Stuurgroepleden, begeleidingsgroepleden en onderzoeksteam

In onderstaande tabel zijn de leden van de stuurgroep opgenomen.

Stuurgroep Lid	Organisatie
Wim van Nunspeet (voorzitter/opdrachtgever)	CBS
Richard Blad	Ministerie van Economische zaken
Guus Bronkhorst (agendalid)	Ministerie van Binnenlandse Zaken
Maarten Hillenaar (agendalid)	Ministerie van Binnenlandse Zaken
Hans Overbeek	Koop
Marijke Salters	BFS / Logius
Imke Vrijling	Ministerie van Binnenlandse Zaken
Arianne de Man	IPO
Marcel Reuvers	GeoNovum

In onderstaande tabel zijn de leden van de begeleidingsgroep opgenomen.

Begeleidingsgroep Lid	Organisatie
Marijke Salters (voorzitter)	BFS / Logius
Bart Knubben	BFS / Logius
Hans Overbeek	KOOP / De Werkmaatschappij
Paul Suijkerbuijk	ICTU
Dorien van Gastel	ECP/EPN

In het onderstaande tabel zijn de namen opgenomen van het onderzoeksteam van Logica Business Consulting.

Onderzoeksteam Lid	Functie
Jaap Korteweg	Onderzoeksleider / Senior adviseur
Annemarie van Campen	Senior adviseur
Martin Borman	Senior adviseur
Angela Barendrecht	Junior adviseur
Mark Hartman	Lead adviseur
Jeff Rothenberg	Internationaal expert/ Senior adviseur

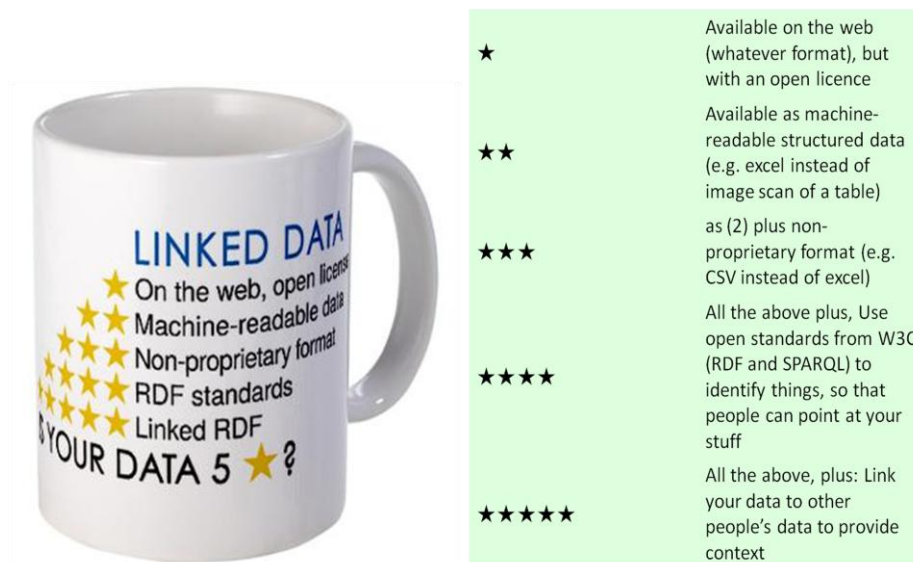
Bijlage III: Lijst geïnterviewde personen

Nr.	Naam	Organisatie
1	de heer F. Verschoor	Antwoord voor Bedrijven
2	de heer H. Coomans	ICTU/ I&M
3	mevrouw M. Bevelander	IPO
4	de heer J.F. van Arragon	Provincie Zuid-Holland
5	de heer H. Overbeek	KOOP
6	de heer D. Woensdregt	CBS
7	de heer G. Hoogeboom	NDW
8	de heer E. Felici	NDW
9	de heer P. Francissen	Gemeente Amersfoort
10	de heer G. van Duijn	Gemeente Amersfoort
11	de heer J. Dolmans	Plan bureau voor de Leefomgeving
12	de heer L. Brandes	Plan bureau voor de Leefomgeving
13	de heer M. de Jong	Gemeente Rotterdam
14	de heer R. van der Klip	Gemeente Rotterdam
15	de heer R. Hoekstra	VU
16	de heer R. Franz	Open Sahara
17	de heer A. van Leeuwen	Reed Elsevier
18	de heer R. Deijkers	Reed Elsevier

Bijlage IV: Linked Open Data: vijf sterren model van Tim-Berners-Lee

Tim Berners-Lee suggereert een vijf sterren schema voor Linked Open Data en Ed Summers verstrekt [nice rendering](#) ervan.

Open data en linked data



Linked Open Data: vijf sterren model met voorbeelden

- ★ Stel je data beschikbaar op het Web (in welk format dan ook) onder een open licentie¹
- ★★ Stel het beschikbaar als gestructureerde data (zoals Excel in plaats van een plaatje van een tabel)²
- ★★★ Gebruik niet-merkgebonden formats (zoals CSV in plaats van Excel)³
- ★★★★ gebruik URIs om dingen te identificeren, zodat mensen naar jouw data kunnen verwijzen⁴
- ★★★★★ link jouw data aan andere data om context te verstrekken⁵

Wat zijn de kosten en opbrengsten van ★ Web data?

Als consument ...

- ✓ Kun je er naar kijken.
- ✓ Kun je het printen.
- ✓ Kun je het lokaal opslaan (op je harde schijf of op een usb-stick).
- ✓ Kun je de data handmatig in een ander systeem zetten.

Als publicist ...

- ✓ Het is simpel om te publiceren.

”

It's great to have the data accessible on the Web under an open license (such as [PDDL](#), [ODC-by](#) or [CC0](#)), however, the data is locked-up in a document. Other than writing a custom scraper, it's hard to get the data out of the document.

Wat zijn de kosten en opbrengsten van ★★ Web data?

Als consument, kan je alles doen wat je wilt met ★ Web data en bovendien:

- ✓ Kun je het direct verwerken met merkgebonden software om het te aggregeren, berekeningen uit te voeren, het te visualiseren, etc.
- ✓ Kun je het exporteren naar een ander (gestructureerd) format.

Als publicist ...

- ✓ Is het nog steeds simpel om te publiceren.

”

Splendid! The data is accessible on the Web in a structured way (that is, machine-readable), however, the data is still locked-up in a document. To get the data out of the document you depend on proprietary software.

Wat zijn de kosten en opbrengsten van ★★★ Web data?

Als consument, kan je alles doen wat je wilt met ★★ Web data en bovendien:

- ✓ Kun je de data manipuleren op elke manier die je wilt, zonder dat je beperkt wordt door de mogelijkheden van specifieke software.

Als een publicist ...

- ✓ Heb je wellicht converters of plug-ins nodig om de data te exporteren van het merkgebonden format.
- ✓ Is het nog steeds vrij simpel om te publiceren.

”

Excellent! The data is not only available via the Web but now everyone can use the data easily. On the other hand, it's still data on the Web and not data in the Web.

Wat zijn de kosten en opbrengsten van ★★★★★ Web data?

Als consument, kan je alles doen wat je wilt met ★★ Web data en bovendien:

- ✓ Kun je er naartoe linken vanaf elke andere plaats (op het Web of lokaal).
- ✓ Kun je het bookmarken.
- ✓ Kun je delen van de data hergebruiken.

Als een publicist ...

- ✓ Investeer je wat tijd in het goed zetten van je data.

- ✓ Je moet URIs toewijzen aan data items en denken over een manier om de data te representeren.
- ✓ Heb je fine-granular controle over de data items en kun je de toegang ertoe optimaliseren (load balancing, caching, etc.)

»

Wonderful! Now it's data in the Web. The (most important) data items have a URI and can be shared on the Web. A native way to represent the data is using RDF, however other formats such as Atom can be converted/mapped, if required.

Wat zijn de kosten en opbrengsten van ★★★★★ Web data?

Als consument, kan je alles doen wat je wilt met ★★★★★ Web data en bovendien:

- ✓ Kun je meer gerelateerde data ontdekken, terwijl je data consumeert.
- ✓ Kun je direct leren over het data schema.

Als een publicist ...

- ✓ Moet je investeren in resources om je data te kunnen linken aan andere data op het Web.
- ✓ Kun je data vindbaar maken.
- ✓ Kun je de waarde van je data doen toenemen.

»

Brilliant! Now it's data, in the Web linked to other data. Both the consumer and the publisher benefit from the network effect.