

# Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4

**Auteurs** Jasper Soetendal (Gemeente Amsterdam)  
Renée van den Bremer (Gemeente Utrecht)  
Donovan Karamat Ali (Gemeente Utrecht)  
Antoine Gribnau (Gemeente Den Haag)  
Remco Duivesteijn (Gemeente Den Haag)  
Ad Steenbakkers (Gemeente Eindhoven)

**Datum** 8 juni 2017

## Doel document

In dit document beschrijven we een gezamenlijke aanpak voor de G4 met betrekking tot het vastleggen van metadata bij het publiceren van (open) data. Door gebruik te maken van een uniforme structuur voor metadata, die geschikt is voor alle type datasets (open data, stelsel van basisregistraties, financiële data, geografische data, etc.) scheppen de G4-steden helderheid voor interne en externe gebruikers van de data; Zij ontvangen van de verschillende steden de metadata op dezelfde wijze, wat de bruikbaarheid ten goede komt.

## Samenvatting

Voor het vastleggen van de metadata bij publicatie is reeds een nationale standaard voorhanden: DCAT-AP-NL. Deze beschrijft alleen de hoogstnoodzakelijke metadata. Daarnaast is behoefte aan meer context van de dataset.

Deze context kan aan de hand van drie nieuwe templates worden ingevuld voor:

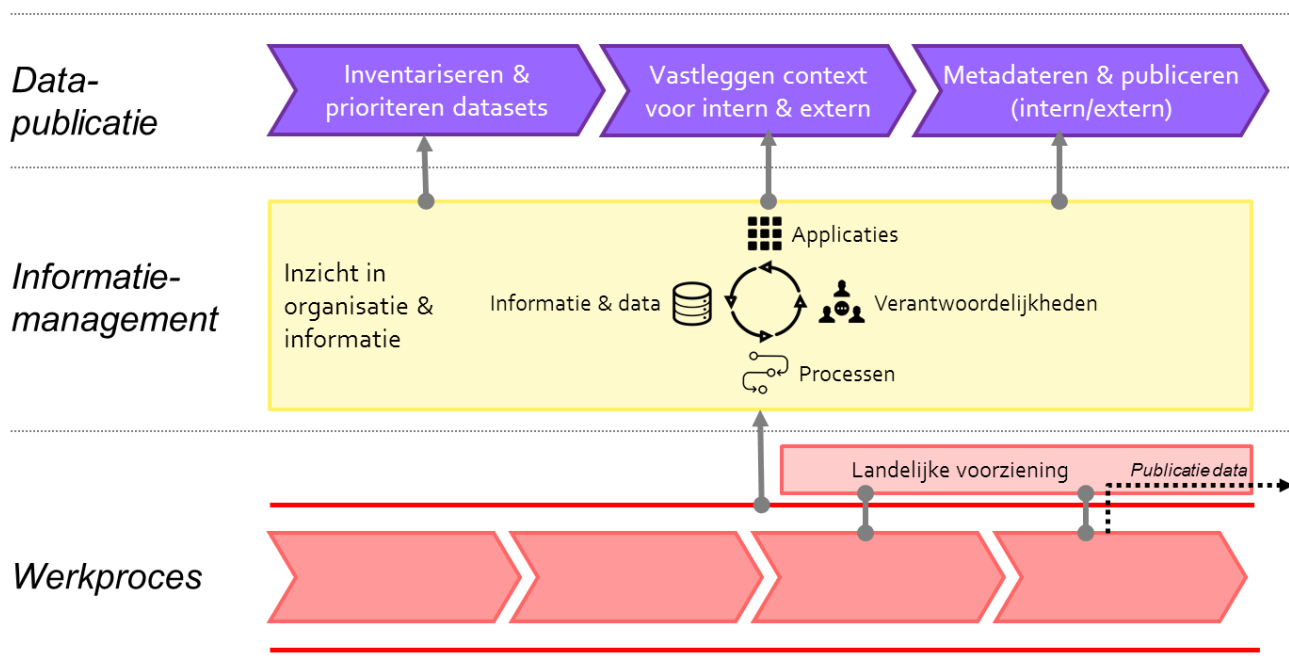
- de context voor bron & beheer, vooral voor intern gebruik,
- de context voor gebruik, voor iedereen die de data gaat gebruiken,
- en het 'proces publicatie' dat vooral van belang is voor de databeheerders.

## Vervolg

Het is aan de deelnemers van de G4 om de in dit document voorgestelde aanpak daadwerkelijk toe te passen. Idealiter wordt in de toekomst de metadata gedeeld in één gezamenlijk ingerichte datacatalogus.

## Metadatabeheer

### De ideale situatie



In een ideale situatie is de data gerelateerd aan een werkproces, dat data gebruikt en exporteert via een landelijke voorziening. 'Bovenop' het werkproces is vastgelegd welke processen er zijn, wie verantwoordelijk zijn, welke applicaties worden gebruikt en welke informatie & data er is. En vooral de onderlinge relaties tussen deze zaken. Kortom, er is inzicht in de organisatie en informatie middels Informatiemanagement.

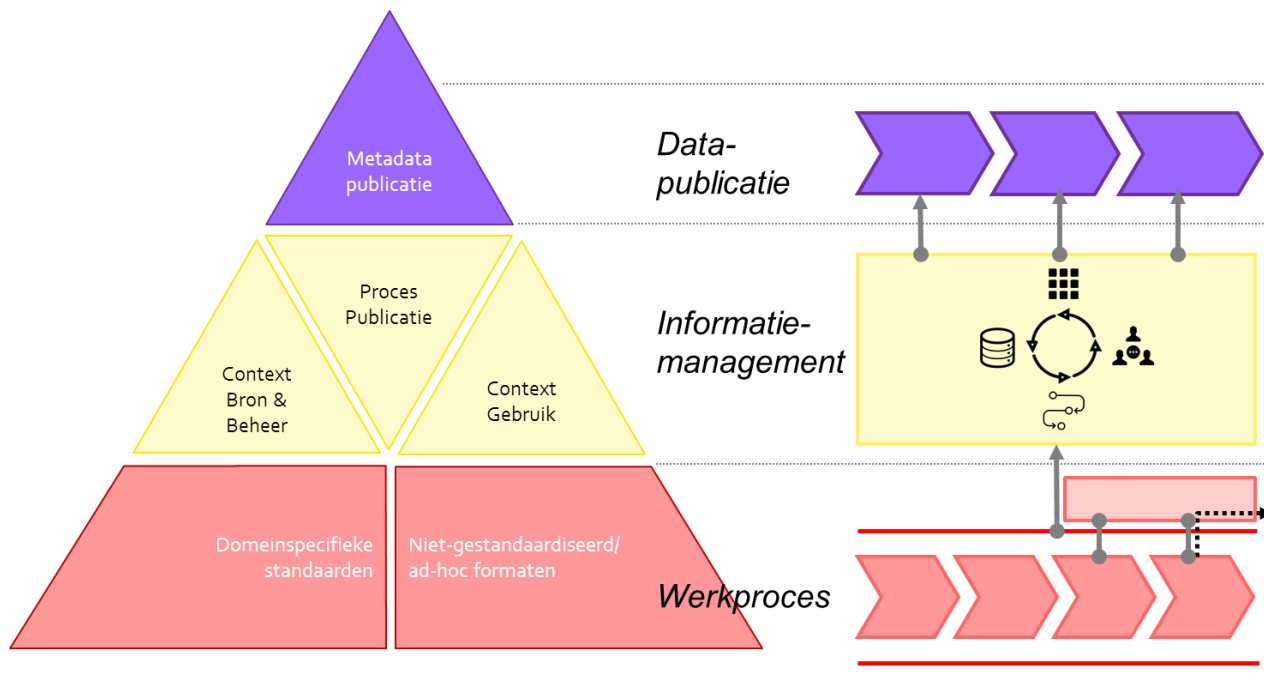
Als hierin is voorzien, zijn de stappen voor datapublicatie eigenlijk een simpele invuloefening. Alle zaken die we hiervoor nodig hebben kunnen we direct overnemen uit de Informatiemanagementlaag.

Helaas is deze laag van Informatiemanagement in de praktijk niet aanwezig of onvolledig. Er wordt vrij ad-hoc open data gepubliceerd vanuit het werkproces, zonder dat vooraf goed is nagedacht over de prioriteit en de context.

Omdat het inzicht in de organisatie & informatie ontbreekt, of in ieder geval niet expliciet is vastgesteld, moet er veel werk verricht worden om de informatie voor de metadatering boven water te krijgen. Vaak worden dan alleen de verplichte velden ingevuld en is het zowel voor interne als externe gebruikers moeilijk te achterhalen wat de context van de data is.

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017

### Metadata Piramide



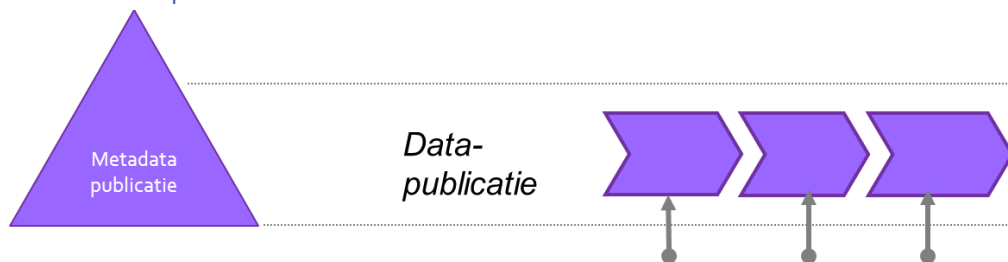
Om de metadata en context van gepubliceerde data zo duidelijk mogelijk vast te leggen, hebben we de 'Metadata Piramide' opgesteld. Deze is op dezelfde manier opgebouwd uit drie lagen.

Onderaan, binnen het werkproces worden al dan niet standaarden gebruikt. Deze zijn domeinspecifiek, de definitie of selectie van deze standaarden valt buiten de scope van dit document, de gebruikte standaarden worden enkel vermeld in de metadata.

Daarboven zijn er in geel drie driehoekjes die we verder invullen: de context voor bron & beheer, vooral voor intern gebruik, de context voor gebruik, voor iedereen die de data gaat gebruiken en het 'proces publicatie' dat vooral van belang is voor de databeheerders. Deze zaken vallen binnen 'Informatiemanagement' en zouden in de ideale Informatievoorziening al gevuld zijn.

Bovenaan staat de metadata publicatie, dit zijn de basismetadata die voor alle databronnen ingevuld kunnen worden en vooral gaan over wat het is, en wie het wanneer gepubliceerd heeft.

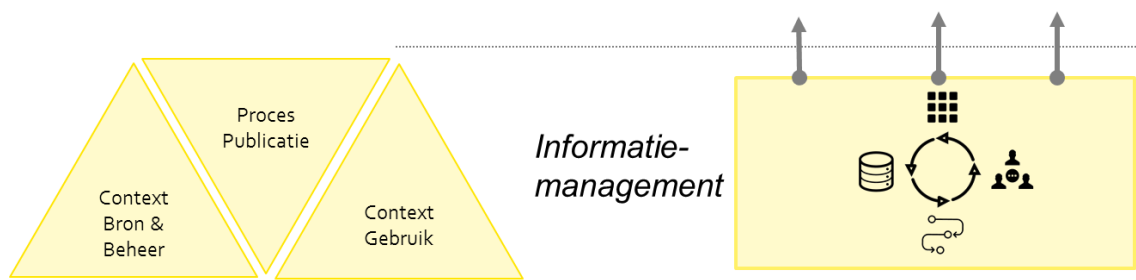
### Laag 1 - Metadata publicatie: DCAT-AP-NL



Voor de invulling van de 'Metadata publicatie' maken we gebruik van de DCAT-AP-NL standaard zoals deze door data.overheid.nl wordt gehanteerd in hun Open Data Catalogus.

Voor meer informatie over DCAT-AP-NL, zie: <https://data.overheid.nl/IPM-Datamodel>

## Laag 2 – Context



Voor informatiemanagement bestaan diverse standaarden, zoals BISL, NORA, GEMMA, etc. We zien informatiemanagement in de context van dit document als de samenhang van werkprocessen, voorzieningen en data en de wijze waarop dit wordt georganiseerd. Deze standaarden bieden echter geen duidelijke templates om de metadata en context van data vast te leggen. Ook bestaande standaarden voor metadata (bijv. NEN-3610, ISO 19115, INSPIRE/IMGEO, TMLO, RGBZ, etc.) focussen veel op de ‘harde’ metadata, al dan niet domeinspecifiek, maar veel minder op de ‘zachtere’ context van de data.

Daarom hebben we eerste een inventarisatie gemaakt van de templates, Excels en lijsten die hiervoor al gebruikt werden door de G4 en Eindhoven. Aangevuld met genoemde bestaande standaarden leverde dit een lijst op met ruim 150 verschillende velden.

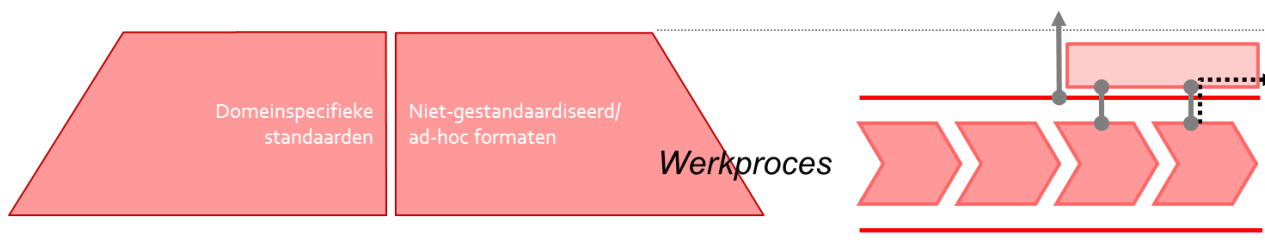
In de werkgroep hebben we deze velden gecategoriseerd en ingedeeld in drie delen:

- Proces publicatie
- Context bron & beheer
- Context gebruik

Onderstaande tabel geeft goed aan waarom en wanneer deze onderdelen relevant zijn:

Onderdeel	Waarom?	Wat?	Wanneer?	Voor wie?	Hoe?
	Vastleggen proces om tot open data te komen. Incl. prioriteren van datasets.	Wat moet er gedaan worden om data openbaar te maken, hoe moeilijk is het en hoe groot is verwacht gebruik.	Bij de start van (open) data activiteiten.	Interne (werk)groep+ beslissers Open Data	Aan de hand van template in dit document.
	Om inzicht te krijgen wat de databron is, hoe deze beheerd wordt en welke conversies er zijn.	Bron, proces, applicaties, beheer, verantwoordelijken, kwaliteit, borging, etc.	Tijdens het proces van beschikbaar maken van open data	Interne gebruikers data	Overnemen uit bestaande proces-beschrijvingen en informatiemodel, door template in dit document te vullen
	Om (externe) gebruikers inzicht te geven hoe zij de data kunnen gebruiken.	Toegang, toelichting, gebruikte standaarden, documentatie, waar wel/niet gebruiken	Bij publiceren van (open) data.	In- en externe gebruikers data	Overnemen uit bestaande proces-beschrijvingen en informatiemodel, door template in dit document te vullen
	Vastleggen publicatie + hoofdlijnen bovenstaande in gestandaardiseerd metadata formaat.	Gegevens publicatie + wat, wie, waar en wanneer, op hoofdlijnen.	Bij publiceren van (open) data. Minimale set metadata dient beschikbaar te zijn bij publicatie	In- en externe gebruikers data	Gegevens uit bovenstaande invullen in standaard <b>Standaard:</b> DCAT-AP-NL

### Laag 3 – Domeinspecifiek






In de onderste laag, in het werkproces, worden domeinspecifieke standaarden en/of niet-gestandaardiseerde ad-hoc formaten gebruikt. Voor de metadata van een dataset is vooral van belang of er standaarden gebruikt zijn en zo ja, een zo duidelijk mogelijke verwijzing naar deze standaard inclusief bijbehorende documentatie.

Dit document doet verder geen uitspraak over welke standaarden wel/niet toegepast moeten worden.

### Aan de slag met de templates

Na een inventarisatie van de genoemde 150 velden, hebben we aan de hand van een enquête naar data-beheerders, informatie managers en datagebruikers een inventarisatie gemaakt van welke velden men belangrijk vindt. Dit resulteert in een 'score' van prioriteit van alle velden: must have, should have en could have.

	<b>Must have</b> Alle velden met deze indicatie dienen ingevuld te zijn.
	<b>Should have</b> Bij voorkeur velden met deze indicatie invullen
	<b>Could have</b> Deze velden zijn optioneel



In de bijlagen zijn vier templates opgenomen voor de vier 'driehoekjes' uit de bovenste twee lagen van de piramide:

- Template A – Metadata Publicatie (DCAT-AP-NL)
- Template B – Proces publicatie
- Template B – Context bron & beheer
- Template C – Context gebruik


Deze templates kunnen ingezet worden bij de publicatie van open data zoals aangegeven in de tabel op de voorgaande pagina.

## Template A – Metadata Publicatie (o.b.v. DCAT-AP-NL)





## Inhoudelijk

	<b>Naam</b>	
		<i>Geef een titel van de dataset</i>
	<b>Beschrijving inhoud</b>	
		<i>Geef een samenvatting van de inhoud van de dataset is: welke gegevens zitten erin en wat is expliciet eruit gelaten</i>
	<b>Meer informatie</b>	
		<i>Geef een link naar een website waar meer informatie te vinden is over deze dataset</i>

## Verantwoordelijken

	<b>Eigenaar (Organisatie)</b>	
		<i>Eigenaar en verantwoordelijke voor de betreffende registratie</i>
	<b>Verstrekker (Organisatie)</b>	
		<i>Verantwoordelijke voor publiceren dataset</i>
	<b>Contactpersoon (Contact)</b>	
		<i>Persoon waarmee contact opgenomen kan worden met inhoudelijke vragen over de data in de dataset</i>



## Bereik &amp; granulariteit

	<b>Dekking in tijd (van/tot)</b>	
		<i>Geef de tijdsperiode aan (begindatum en einddatum), die de dataset beslaat</i>
	<b>Geografisch gebied</b>	
		<i>Beschrijf het gebied dat de dataset bestrijkt</i>
	<b>Geografische eenheid</b>	
		<i>Geef een eenheid van het gebied (bijv. gemeente, wijk, buurt, specifieke lijnen/vlakken/punten, etc.) waarin de dataset is uitgedrukt</i>
	<b>Wijzigingsfrequentie</b>	



## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017

	<i>Geef aan hoe vaak de gegevens worden geactualiseerd (bijv. real-time, dagelijks, wekelijks, jaarlijks, op afroep, etc.)</i>
--	--





### Datum

 <b>Uitgiftedatum</b>	
	<i>Geef de datum waarop de dataset beschikbaar is gesteld</i>
 <b>Wijzigingsdatum</b>	
	<i>Geef de datum waarop de dataset het laatst is gewijzigd</i>



### Categorisatie

 <b>Thema</b>	
	<i>Het hoofdthema waaronder de dataset valt</i>
 <b>Trefwoorden</b>	
	<i>Geef een aantal trefwoorden, die van toepassing zijn op de dataset</i>

### Gebruik

 <b>Licentie</b>	
	<i>Geef aan onder welke open data licentie de dataset gepubliceerd is, indien van toepassing</i>
 <b>Rechten</b>	
	<i>Beschrijf hier eventuele extra rechten die gelden bovenop de licentie</i>
 <b>Taal</b>	
	<i>Geeft de taal van de dataset</i>
 <b>Toegepaste standaard(en)</b>	
	<i>Geef aan welke (inter)nationale standaarden zijn gehanteerd in de dataset</i>


### Versie

 <b>Versie</b>	
	<i>Geef indien van toepassing de versie van de dataset aan</i>
 <b>Toelichting bij versie</b>	

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017









	<i>Geef indien van toepassing een toelichting op de vermelde versie</i>
--	---

### Wettelijk

 <b>Grondslag</b>	
	<i>Geef indien van toepassing aan wat is de oorspronkelijke juridische basis is van de dataset</i>

### Link/resource

Neem per link (dit kunnen er meerdere zijn) onderstaande velden op:

 <b>Titel</b>	
	<i>Geef de titel van de link</i>
 <b>Omschrijving</b>	
	<i>Geef een omschrijving van de link</i>
 <b>Status</b>	
	<i>Geef aan wat de huidige status van de dataset is (Bijv. 'actueel', 'historisch', 'concept', 'definitief', etc.)</i>
 <b>Formaat</b>	
	<i>Geef het dataformaat van de link aan (Bijv. CSV, XLSX of GeoJSON)</i>
 <b>Grootte bestand</b>	
	<i>Geef indien bekend de bestandsgrootte van het bestand aan (Bijv. 10.1 MB)</i>
 <b>Link/URL/Bestand</b>	
	<i>De daadwerkelijke link (URL) van de dataset</i>
 <b>Uitgiftedatum</b>	
	<i>Geef de datum waarop de resource beschikbaar is gesteld</i>
 <b>Wijzigingsdatum</b>	
	<i>Geef de datum waarop de resource het laatst is gewijzigd</i>





## Template B - Proces publicatie



### Status

	<b>Huidige status</b>	
		<i>Een korte samenvatting van de huidige status van de dataset. Denk hierbij aan of de dataset al online beschikbaar is, in welk systeem deze staat, of deze al geëxporteerd kan worden, etc.</i>
	<b>Wel/niet online</b>	
		<i>Geef aan of de dataset al online beschikbaar is en zo ja, voor wie (open, gedeeld, intern) deze data toegankelijk is.</i>
	<b>Locatie</b>	
		<i>Naam van de plek waarop de data (als open data, of als interne data) al beschikbaar is (bijvoorbeeld data.overheid.nl, opendata.cbs.nl, G-schijf mapje 'data', etc.).</i>
	<b>Contactpersoon</b>	
		<i>Naam + contactgegevens van de contactpersoon voor deze dataset</i>

### Technisch

	<b>Exporteerbaar</b>	
		<i>Geef aan welke mogelijkheden er zijn om de data te exporteren uit het systeem/proces/applicatie waar deze wordt opgeslagen.</i>
	<b>Geo-aanduiding</b>	
		<i>Geef aan of er al geografie is opgenomen in de dataset, en zo ja, in welke vorm (adres, buurten, punten, polygonen, etc.)</i>

### Metadatering


	<b>Metadatering aanwezig</b>	
		<i>Geef aan of er al metadata voor de dataset beschikbaar is. Zo niet, geef dan aan of er voldoende documentatie is om deze over te nemen, of dat deze nog door de broneigenaar ingevuld moet worden.</i>
	<b>url</b>	
		<i>Indien metadata reeds aanwezig is, geef hier een link naar de metadata</i>

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017

### Juridisch

	<b>Aanwezigheid persoonsgegevens</b>	
		<i>Geef aan of in deze datasets persoonsgegevens zijn opgenomen (persoonsgegevens als bedoeld in artikel 10 lid 1 sub d dan wel artikel 10 lid 2 sub e Wob)</i>
	<b>Weigeringsgronden</b>	
		<i>Geef aan of er weigeringsgronden zijn waardoor deze dataset niet openbaar gemaakt kan worden. Weigeringsgronden kunnen zijn: aanwezigheid persoonsgegevens, inbreuk persoonlijke levenssfeer, vertrouwelijke bedrijfs- en fabricagegegevens, schenden economische en financiële belangen, frustreren opsporing en vervolging, belemmeren toezicht, onevenredig bevoordelen of benadelen, eenheid van kroon in gevaar, schaden staatsveiligheid, schenden internationale betrekkingen, persoonlijke beleidsopvattingen van ambtenaren,</i>
	<b>Rechten van derden</b>	
		<i>Bepaling of er rechten van derden (auteursrecht, databankrecht) op de gegevens rusten. Bijvoorbeeld omdat commerciële bedrijven de gegevens hebben geleverd of eigenaar zijn van de structuur van de database.</i>

### Prioritering publicatie



	<b>Inschatting gebruik</b>	
		<i>Inschatting of deze data voor veel gebruikers relevant is. Bepaal zo goed als mogelijk in hoeverre deze informatie relevant is voor partijen binnen en, vooral, buiten de eigen organisatie.</i>
	<b>Eenvoud van publicatie</b>	
		<i>Maak een zo goed mogelijk inschatting van de eenvoud of de complexiteit om de data te publiceren. Denk hierbij aan de stappen die reeds genomen zijn of juist nog genomen moeten worden om tot publicatie te komen.</i>
	<b>Business impact</b>	
		<i>Geef een beschrijving van de business impact van de eigen organisatie voor het publiceren van deze data.</i>
	<b>Risico's</b>	
		<i>Benoem mogelijke risico's die ontstaan met het publiceren van de data</i>

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017



<b>i</b> <b>Prioritering publicatie</b>	
	<i>Bepaal de prioriteit voor publicatie, aan de hand van alle bovenstaande velden.</i>

## Template B - Context Bron & Beheer



### Bron

	<b>Bronhoudende organisatie</b>	
		<i>Officiële naam van de rechtspersoon bronhouder (rechtmatige eigenaar van deze dataset)</i>
	<b>Verantwoordelijken</b>	
		<i>Voor- en achternaam van de voor de data verantwoordelijke persoon, inclusief functie en organisatie/dienst/sector/afdeling</i>
	<b>Bronstelsysteem</b>	
		<i>Naam van het bronsysteem waar de data wordt vastgelegd.</i>
	<b>Locatie data</b>	
		<i>Waar staat de data (als de data niet in een applicatie staat)</i>

### Proces



	<b>Werkprocessen die de data verzamelen /genereren</b>	
		<i>Benoem de werkprocessen in de organisatie die de data verzamelen en/of genereren.</i>
	<b>Werkprocessen die de data gebruiken</b>	
		<i>Benoem de werkprocessen in de organisatie die de data gebruiken.</i>

### Applicatie






	<b>Applicatie(s)</b>	
		<i>Benoem de applicatie die deze gegevens gebruiken. Maak indien mogelijk onderscheid tussen applicaties die de data verzamelen/genereren, applicaties die de data wijzigen en applicaties die de data alleen lezen.</i>
	<b>Applicatiebeheerder</b>	
		<i>Persoon die verantwoordelijk is voor het dagelijks applicatiebeheer van de applicatie(s) die de data verzameld/genereerd, inclusief functie en organisatie/dienst/sector/afdeling.</i>

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017

### Beheer brondata

 <b>Verantwoordelijke beheer</b>	
	<i>Voor- en achternaam van de persoon verantwoordelijk voor het beheer van de brondata van deze dataset, inclusief functie en organisatie/dienst/sector/afdeling</i>
 <b>Uitvoering beheer</b>	
	<i>Voor- en achternaam van de uitvoerende van het beheer van de brondata van deze dataset, inclusief functie en organisatie/dienst/sector/afdeling</i>


### Beheer publicatie

 <b>Verantwoordelijke publicatie</b>	
	<i>Voor- en achternaam van de persoon verantwoordelijk voor het publiceren van deze dataset, inclusief functie en organisatie/dienst/sector/afdeling</i>
 <b>Leveringswijze</b>	
	<i>Beschrijving van de wijze waarop de data wordt aangeleverd voor publicatie.</i>
 <b>Conversie</b>	
	<i>Beschrijving van de conversie van de data die plaats heeft gevonden tussen het bronsysteem en de daadwerkelijke publicatie (denk aan aggregatie, verwijderen kolommen, toevoegen geografie, etc.)</i>
 <b>Wijze van actualisatie</b>	
	<i>Beschrijving hoe de gepubliceerde data geactualiseerd wordt. (bijv. 'volledig automatisch via API op live data' of 'jaarlijkse handmatige export')</i>
 <b>Frequentie van wijzigingen</b>	
	<i>Geef aan met welke frequentie de gepubliceerde dataset wordt gewijzigd. Indien deze frequentie afwijkt van de actualisatiefrequentie van de brondata, geeft dit dan expliciet aan.</i>

### Kwaliteit/borging




 <b>Gegevensleverovereenkomst</b>	
	<i>Indien de dataset beschikbaar wordt gesteld middels een</i>

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017







		<i>gegevensleverovereenkomst, neem hier dan een verwijzing naar dit document op.</i>
 <b>Datakwaliteit</b>		
		<i>Beschrijf de kwaliteit van de data aan de hand van in ieder geval de factoren: Volledigheid Juistheid Actualiteit</i>

## Template C - Context Gebruik

### Toegang

	<b>Toegangsrestricties</b>	
		<i>Wie hebben toegang tot de data en welke afspraken zijn er over de toegang tot de data.</i>
	<b>Registratie gebruik</b>	
		<i>Wordt het gebruik van de data handmatig of automatisch bijgehouden, gemeten?</i>
	<b>Beperkingen gebruik</b>	
		<i>Zijn er beperkingen opgelegd aan het gebruik van de data? Beschrijf de beperkingen of verwijst naar de licentie of overeenkomst waarin de beperkingen zijn vastgelegd.</i>



### Inhoudelijk

	<b>Inhoud &amp; toelichting</b>	
		<i>Uitgebreide toelichting van de inhoud van de data en de bestanden, welke objecten/zaken in de bestanden zijn opgenomen, de reden waarom de informatie verzameld wordt en evt. de relatie tot andere gerelateerde bestanden/datasets binnen en buiten de organisatie.</i>
	<b>Gebruikte standaarden</b>	
		<i>Beschrijf zo specifiek mogelijk welke lokale, nationale of internationale standaarden zijn gebruikt in de dataset. Voeg indien mogelijk een verwijzing naar deze standaard inclusief bijbehorende documentatie toe.</i>
	<b>Aandachtspunten</b>	
		<i>Indien er specifieke aandachtspunten zijn bij het hergebruik van de data, vermeld deze dan.</i>
	<b>Documentatie</b>	
		<i>Verwijzing naar de documentatie die beschikbaar is voor de dataset.</i>
	<b>Logisch datamodel</b>	
		<i>Indien een logisch datamodel beschikbaar is voor de data in de dataset, neem dan een verwijzing naar dit model op.</i>
	<b>Voorbeelden</b>	

## 5B Gezamenlijke aanpak Metadatabeheer (Open) Data in de G4, d.d. 8 juni 2017

	<i>Voorbeelden van gebruik. Dit kunnen echte voorbeelden/toepassingen zijn (incl. links) of beschrijving van hoe de data toegepast zou kunnen worden.</i>

### Contactgegevens (extern)

 <b>Contactgegevens inhoud</b>	
	<i>Naam, functie en contactgegevens van de persoon die direct vragen kan beantwoorden over de inhoud van de dataset.</i>
 <b>Contactgegevens levering</b>	
	<i>Naam, functie en contactgegevens van de persoon die direct vragen kan beantwoorden over de technische levering van de dataset.</i>