

Linked Open Data Pilot

Forum Standaardisatie



Programma stelsel van basisregistraties



gemeente Amersfoort



Interprovinciaal overleg



gemeente Nijmegen



Kadaster



Geobusiness Nederland



Kwaliteitsinstituut Nederlandse Gemeenten



Geonovum, programma standaarden



Ministerie van Infrastructuur en Milieu



Geonovum, programma Inspire



Management samenvatting

De op de voorpagina genoemde partijen onderschrijven het belang om onderzoek te doen naar de mogelijkheden om aan elkaar gerelateerde informatie uit verschillende overheidsregistraties uniform te ontsluiten op het web met betrekking tot open data.

Het doel van de Linked Open Data (LOD) pilot is om met een integrale aanpak van data preparatie, linken van data tot toepassing aan het werk te gaan en om al doende op basis van praktijkervaring inzicht te krijgen in de mogelijkheden en wensen. De verkregen inzichten kunnen vervolgens leiden tot algemene afspraken en afgesproken standaarden voor ontsluiting van open data op het web.

Open Data kent vele facetten; de scope in de pilot ligt op het linken van open data; veelal gezien als de vervolgstap na het beschikbaar stellen van Open Data. Buiten de scope van deze pilot valt de discussie rondom de juridische kant van open data. Alleen de data die zonder restricties door de deelnemende partijen van het grondgebied van de gemeenten Amersfoort en Nijmegen verstrekt kan worden is onderdeel van de pilot.

De uitvoering van de pilot ligt nadrukkelijk bij de deelnemende partijen uit zowel de wetenschappelijke, publieke als private sector. Het is een pilot die voor iedereen is opengesteld in deelname. De op de voorpagina genoemde partijen vormen het projectteam en zorgen voor de financiën, beslissingen, data en katalyseren van het proces. De deelnemers van de LOD pilot doen kosteloos mee en stellen uren van zichzelf of organisatie ter beschikking. Gedurende de pilot kunnen deelnemers aanhaken. Elke deelname is welkom, sommige deelnemers zullen actief aan een activiteit deelnemen (werken), anderen zullen vooral deelnemen om inzicht te krijgen in LOD (kennis halen / brengen).

Het plan voor de LOD pilot is geïnitieerd door de observatie dat op meerdere plaatsen open data gepubliceerd wordt maar dat er een gemis is aan een gemeenschappelijk referentiekader voor de ontsluiting.

Dit is aanwezig aan de technische kant van standaarden zodat verschillende formaten worden gebruikt vanuit de overheid dat het gemeenschappelijk gebruik van open data limiteert. De open data van de overheid worden als afzonderlijke silo's ontsloten en links tussen overheidsverzamelingen en verzamelingen op het web worden niet gelegd terwijl deze wel aanwezig (kunnen) zijn. Daarnaast leidt het ontbreken van een referentiekader tot drempelvrees voor overheden om open data te publiceren.

De standaard voor Linked Data biedt veel potentie en is daarmee een belangrijk onderdeel van de pilot. Andere ontsluitingsmechnismen zullen tevens onderwerp zijn van de LOD pilot.

Om richting te geven aan de toepassingen zullen communities worden gevormd. Hier komt de data (informatie) en vraag bij elkaar rondom bepaalde onderwerpen (bijvoorbeeld zorg, openbare orde en veiligheid, etc.). Dit zal voor een belangrijk deel de data en links tussen de data bepalen.

Omdat de deelnemende partijen zowel aanbieders als gebruikers zijn van open data en de bijbehorende ontsluiting, biedt de pilot de unieke kans om de aanbieders te voeden met nauwgezette wensen van gebruikers. Anderzijds zal door breed beschikbaar komen van open data voor deze gemeenten de data in meer toepassingen gebruikt gaan worden. Aanbod en vraag zullen door samen te werken in de pilot, al doende op elkaar worden afgestemd.

De pilot voor LOD heeft als uitvoeringsperiode september 2012- augustus 2013.

Inhoudsopgave

1	Inleiding en achtergrond project.....	5
2	Projectdefinitie.....	7
2.1	Doelstellingen.....	7
2.2	Aanpak en fasering	7
2.2.1	<i>Activiteitcluster 1: Aanbod van LOD</i>	8
2.2.2	<i>Activiteitcluster: Toepassingen</i>	10
2.2.3	<i>Activiteitcluster 3: Techniek en standaarden</i>	11
2.2.4	<i>Stuurgroep</i>	13
2.3	Resultaten.....	14
2.6	Relaties met andere projecten.....	15
3	Financien.....	16
4	Organisatiestructuur.....	17
5	Initiële Projectplanning	20
	Bijlage: Grafische weergave projectplanning	i

1 Inleiding en achtergrond project

Het plan voor de Linked Open Data (LOD) pilot is geïnitieerd door de observatie dat op meerdere plaatsen open data gepubliceerd wordt maar dat er een gemis is aan een gemeenschappelijk referentiekader voor de ontsluiting.

Dit is aanwezig aan de technische kant van standaarden zodat verschillende formaten worden gebruikt vanuit de overheid dat het gemeenschappelijk gebruik van open data limiteert. De open data van de overheid worden als afzonderlijke silo's ontsloten en links tussen overheidsverzamelingen en verzamelingen op het web worden niet gelegd terwijl deze wel aanwezig (kunnen) zijn. Daarnaast leidt het ontbreken van een referentiekader tot drempelvrees voor overheden om open data te publiceren.

Er is voor gekozen de LOD pilot in te richten rond een aantal activiteitclusters welke tussen- en eindresultaten zullen opleveren. Dit zijn:

1. Aanbod
2. Toepassingen
3. Techniek & standaarden

Hoe deze activiteitclusters gezamenlijk de inrichting van de pilot vormen is te zien in Figuur 2. In figuur 1 staat een schematisch overzicht van het speelveld van de LOD pilot.



Figuur 1: Schema



Figuur 2: Inrichting LOD pilot

Naast deze drie inhoudelijke activiteitclusters, is er tijdens de centrale bijeenkomsten de mogelijkheid om met elkaar te klankborden en kennis te delen.

In de volgende hoofdstukken komt achtereenvolgens aan de orde:

- de projectdefinitie (o.a. doelstellingen, aanpak en resultaten);
- de organisatiestructuur;
- de initiële projectplanning.

-

In de bijlagen zijn opgenomen:

- projectplanning met grafische weergaven.

2 Projectdefinitie

2.1 Doelstellingen

Het doel van de LOD pilot is om met een integrale aanpak van data preparatie, linken van data tot toepassing aan het werk te gaan en om al doende op basis van praktijkervaring inzicht te krijgen in de mogelijkheden en wensen. De verkregen inzichten kunnen vervolgens leiden tot algemene afspraken en afgesproken standaarden voor ontsluiting van open data op het web.

Na de uitvoering van de LOD pilot heeft Nederland een beter inzicht in de LOD mogelijkheden en kan het gericht aan de slag.

Doelstellingen zijn:

- Samenwerken met veel stakeholders voor de grondgebieden van de gemeente Nijmegen en Amersfoort waar toepassingen worden gerealiseerd met als doel om vast te kunnen stellen of LOD een techniek is om overheidsdata beschikbaar te stellen op het web
- Naast LOD de data in gangbare andere web-formaten beschikbaar te stellen om te komen tot een goed vergelijk
- Vaststellen van open data standaard(en) en referentiekader
- Het leggen van links tussen de 'formele' overheidsregistraties en de 'informele' registraties op het web
- Samenbrengen van kennis, technologie en behoefte
- Awareness en communicatie
- Vastlegging van de ervaringen en aanbevelingen als een open data referentiekader

Open Data kent vele facetten en initiatieven. Deze LOD pilot heeft een specifieke focus op "Linked"; veelal beschouwd als de vervolgstap na het open beschikbaar stellen van databronnen.

2.2 Aanpak en fasering

Het project wordt aangepakt volgens de Prince2 methodiek en bestaat uit de volgende fasen:

0. Voorbereiding en totstandkoming PID LOD
1. LOD pilot zoals beschreven in dit document
2. Doorontwikkeling van de resultaten uit fase 1, bij behoefte

Fase 0 is de voorbereidingsperiode in de totstandkoming van dit document. In fase 0 wordt het projectplan nader ingevuld en de specificaties vastgesteld voor de LOD pilot.

Fase 1 is beschreven in dit document.

Fase 2 valt buiten de scope van dit document. Fase 2 is opgenomen om inzichtelijk te maken dat deze pilot gezien kan worden als een eerste begrensde verkenning van LOD. Nadere verdieping of verbreding van het LOD referentiekader uit fase 1 door incidentele partijen, toepassingsgebieden, en samenwerkingen zullen volgen. De vraag hierbij is of enige vorm van nationale afstemming nodig is. Mocht dit nodig en gewenst zijn dan kan dit gezien worden als een fase 2.

Fase 1 heeft een als uitvoeringsperiode februari 2012 - december 2012.

Beschrijving van de activiteitclusters:

2.2.1 Activiteitcluster 1: Aanbod van LOD

De fasering van deze activiteitcluster is:

1.1. Bepalen relevante registraties grondgebied gemeente Amersfoort en Nijmegen

De gemeente Amersfoort en Nijmegen¹ zullen algemene data beschikbaar stellen en specifieke. Specifieke waar het bepaalde toepassingen betreft. Naast de gemeenten zijn er nog andere partijen die voor deze gemeenten data beschikbaar kunnen stellen. In de pilot is dat het Kadaster die energie-labels en erfdienstbaarheidgebieden beschikbaar stelt. Mogelijk dat ook andere partijen vanuit de pilot deelname open data beschikbaar willen stellen. Denk bijvoorbeeld aan de Rijksdienst voor Cultureel Erfgoed, KVK, Inspire-sets gerelateerd aan grondgebied gemeenten, etc.

1.2. Bepalen uit relevante registraties de set van gegevens die gepubliceerd gaan worden

Uit de verschillende relevante registraties die de gemeenten en andere bronhouders hebben dient bepaald te worden:

- a. Welke data uit de registratie verstrekt mag worden (bepaalde data kan bijvoorbeeld openbaar zijn en andere data mag bijvoorbeeld niet verstrekt worden op basis van privacy)

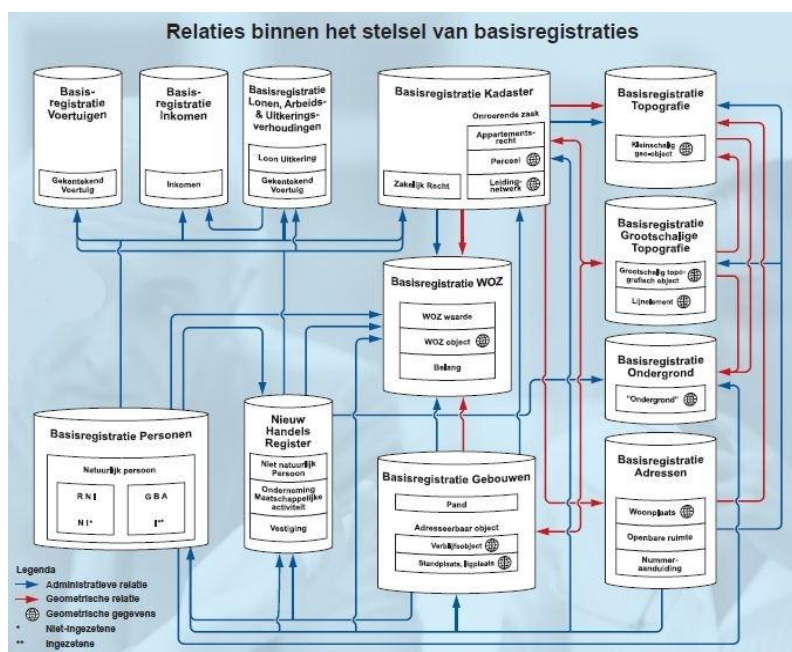
¹ Een eerste lijst van open data (geen Linked open data) is beschikbaar op: <http://www.nijmegen.nl/opendata/>

b. Welke data aansluit bij de toepassingen

b. dient in nauw overleg te gebeuren met de activiteitcluster toepassingen.

1.3. Bepalen extra links tussen data uit registraties

Op basis van overleg met activiteitcluster toepassingen en onderzoek naar interessante links (stelsel van basisregistraties, Inspire, RSGB, NEN 3610, vergelijk objecten uit verschillende registraties) wordt bepaald welke links worden aangebracht tussen data uit de verschillende registraties.



Figuur 3: Voorbeeld van relaties tussen overheidsregistraties

1.4. Data publicatie en aanbrengen extra links naar data op het web (proefomgeving, niet openbaar)

Waar in de vorige activiteiten vooral denkwerk is verricht gaat het in deze activiteit om doen.

De data uit de relevante registraties wordt als Linked data en mogelijk ook in andere web-formaten gepubliceerd. Hiervoor wordt gebruik gemaakt van technieken en standaarden die komen uit het activiteitcluster techniek & standaarden.

Waar tussen data uit registraties links nodig zijn worden deze gelegd. Dit kunnen behoorlijke inspanningen zijn. Deze links worden alleen aangebracht waar de inspanning afweegt tegen de baten. Mogelijk dat het ook voor een bepaald gebied gebeurt van het gemeentelijk grondgebied zodat gebruikt kan worden om aantoonbaar te maken dat deze links wel of geen toegevoegde waarde leveren in gebruik.

toepassingen extra gestimuleerd worden en dat de LOD pilot gebruik kan maken van de mechanismen die er al zijn.

Op basis van de ideeën uit het activiteitcluster kunnen thema's veranderen, andere focus krijgen, etc. Dit is een proces van alle deelnemers.

Hack de Overheid³ is een collectief van experts op het gebied van open data en open data toepassingen. Op de 'HackdeOverheid evenementen' kunnen programmeurs, ondernemers, journalisten en ambtenaren samenwerken aan innovatie op basis van open overheidsdata. Hack de Overheid houdt zich ook bezig met thema's zoals zorg, cultuur, geo, etc. Hack de Overheid ondersteunt de pilot om via haar kanalen te zorgen voor extra communicatie, vraag en aanbod te verbinden, mogelijkheden tot aanspreken kennis uit Hack de Overheid netwerk en organiseren events voor stimulatie apps en andere toepassingen.

Van de werkgroep toepassingen wordt naast allerlei toepassingen verwacht dat zij meedenkt met het aanbod zodat vraag en aanbod zoveel mogelijk op elkaar zijn afgestemd. Daarnaast zal zij toepassingen maken en vooral ook organiseren dat Nederland in beweging komt met het maken van toepassingen voor de grondgebieden van Amersfoort en Nijmegen.

Specifiek zal van de werkgroep toepassingen gevraagd worden of een toepassing gemaakt kan worden waarmee door de gegevens van een gemeente 'gelopen' kan worden. Hiermee wordt bedoeld dat: persoon a woont op adres b, op adres b is het bestemmingsplan c van kracht, geldt bouwbesluit d en heeft de kadastrale kenmerken e, etc. Hiermee kan antwoord gegeven worden op vragen over registraties heen die nu vanuit de Rijksoverheid per registratie beantwoord moeten worden, één ontsluitingsmechanisme.

2.2.3 Activiteitcluster 3: Techniek en standaarden

In dit activiteitcluster wordt praktisch onderzoek gedaan naar de techniek en standaarden. Een onderscheid in tijd dient gemaakt te worden tussen onderzoek dat nodig is voor de data-publicatie en onderzoek dat na data-publicatie dient te gebeuren.

Mogelijke onderzoeksonderwerpen:

3.1 Vergelijk Linked data standaarden en andere data standaarden / technieken voor publicatie op het web

³ <http://www.hackdeoverheid.nl/>

Linked data standaarden bieden veel mogelijkheden. Dit komt vooral door de mogelijkheid om de data te linken en hiermee een grote verzameling van data te maken. Andere geluiden zijn dat Linked data 'academisch' is (ingewikkeld en daarmee niet eenvoudig te implementeren). In dit onderzoek dient een vergelijk te worden gemaakt tussen XML/GML, CSV, JSON, JSON-Linked data, Linked data, en mogelijk andere standaarden of technische mogelijkheden (REST, API, en andere).

Geeft een overzicht van de voors en tegens van de verschillende standaarden/technieken op basis van verschillende criteria.

3.2 Technisch inzicht krijgen in mogelijkheden en onmogelijkheden van de standaarden RDF, SPARQL, GeoSPARQL, RDFS, SKOS, OWL

Hoe ziet de Linked data stack eruit. Hoe hangen de onderdelen samen en hoe werken ze op hoofdlijnen.

3.3 Publishing tools (Tools waarmee uit databases Linked data gegenereerd kan worden)

Er komen steeds meer tools beschikbaar waarmee Linked data automatisch gepubliceerd kan worden. Het testen van deze tools op geschiktheid om de data te publiceren uit activiteitcluster aanbod is onderdeel van deze activiteit.

3.4 Presentatie van Linked data aan gebruiker (user-interfaces zoals smart phones en websites)

Linked data is een techniek en niet zichtbaar voor eindgebruikers, het zit onder websites of apps. De vraag is of Linked data ook relevant is om te presenteren aan gebruikers? Lukt het om deze grote hoeveelheid data op een overzichtelijke manier te presenteren waarbij de trend is naar kleinere devices zoals tablets and smartphones?

3.5 Het linken zelf (Hoe link je informatie uit verschillende databases)

Best practises opdoen hoe links gemaakt kunnen worden tussen data van verschillende registraties. Hoe kan dit slim (zoveel mogelijk geautomatiseerd). Zijn er tools voor beschikbaar en welke zijn dit dan?

3.6 Hoe onderhoud je de links

Nadat de data is gepubliceerd zullen aan de bronbestanden bij de overheid of in collecties op het web aanpassingen gemaakt worden. Zodra deze weer gepubliceerd zijn ontstaan:

- 'dode' links (een link naar een gegeven dat inmiddels niet meer bestaat)
- objecten zonder link(s)

Best practises opdoen hoe op een slimme en consistente manier links kunnen worden onderhouden.

3.7 Linked data browsers en search engines

Er zijn een aantal Linked data browsers en search engines⁴ die verschillende mogelijkheden bieden. Wat bieden deze browsers en search engines en wat is het onderscheid van de Linked data browsers ten opzichte van de huidige browsers zoals Internet explorer en Chrome.

3.8 Geheugenbeperkingen bij laden triples

De Linked data bestanden zijn door de structuur groot in omvang. Bij grote registraties gaat het al gauw om miljoenen triples. Dit kan performance problemen geven. Het gaat hier om het verkrijgen van een beter inzicht naar wat de problemen zijn en welke mogelijkheden van oplossen mogelijk zijn.

3.9 Hoe monitor je wie wat met je data doet?

Is het mogelijk om te monitoren welk gebruik en welke applicaties ontwikkeld worden met de data van je organisatie? Onderzocht wordt of het gebruik van de data gemonitord kan worden en zo ja met welke tools (ondersteund met voorbeelden).

3.10 En andere

Dit kunnen meer of minder onderzoeken worden afhankelijk van de interesse van de deelnemers aan het activiteitcluster.

2.2.4 Stuurgroep

Naast de organisatorische aspecten heeft de stuurgroep de volgende activiteit:

4.1. Vastlegging van de ervaringen en aanbevelingen voor Nederland

De stuurgroep zorgt ervoor dat alle bevindingen in een coherente vorm beschikbaar komen; dat alles goed wordt vastgelegd en waar het kan tijdens en anders na het project openbaar is.

4.2. Specifieke adviezen

Naar aanleiding van de bevindingen en aanbevelingen worden specifieke adviezen geschreven voor:

- a. Forum standaardisatie

⁴ http://en.wikipedia.org/wiki/Linked_data#Browsers

- Een advies dat betrekking heeft op het wel of niet in procedure nemen van de geconstateerde standaarden voor de pas toe of leg uit lijst.
- b. GI-beraad
Een eindrapport over de LOD pilot met adviezen voor de GI-sector.
 - c. Programma stelsel van basisregistraties
Een eindrapport over de LOD pilot met adviezen voor het programma stelsel van basisregistraties.
 - d. KING/gemeenten.
Een eindrapport over de LOD pilot met adviezen voor KING en de betrokken gemeenten.

Het eindrapport zal voor a. t/m d. min of meer gelijk zijn. De adviezen zijn toegesneden op de doelgroep. Mogelijk zal het één eindrapport worden met de specifieke adviezen in aparte paragrafen.

4.3. Eindcongres

Een open eindcongres met de eindresultaten van de LOD pilot wordt medio december 2012 georganiseerd.

2.3 Resultaten

Het project levert de volgende hoofdproducten op, die elk nader gespecificeerd kunnen worden per activiteitencluster.

Activiteitencluster 1: Aanbod van LOD

Rapport met daarin:

- 1.A Beschrijving van de te gebruiken registraties / datasets
- 1.B Objecttypen voor publicatie met bijbehorende motivatie
- 1.C Aanwezig en aan te brengen links tussen overheidsregistraties en registraties op het web (op niveau van objecttypen tussen registraties)
- 1.D Gepubliceerde data (LOD en mogelijk andere formaten)

Activiteitencluster 2: Toepassingen

- 2.A Bepalen thema's en beoogde doelen per thema
- 2.B Beschrijving van bijdrage data aan het thema
- 2.C Stimulatie van allerlei toepassingen door workshops, "hackathons", etc.
- 2.D Toepassing voor het kunnen doorlopen van de data

Activiteitencluster 3: Techniek en standaarden

Rapporten, beschrijvingen, best practises van de verschillende onderzoeksitems.

Stuurgroep

4.A Eindrapport referentiekader LOD met specifieke adviezen voor:

- I. Forum standaardisatie
- II. Programma stelsel van basisregistraties
- III. GI-beraad
- IV. En andere

2.6 Relaties met andere projecten

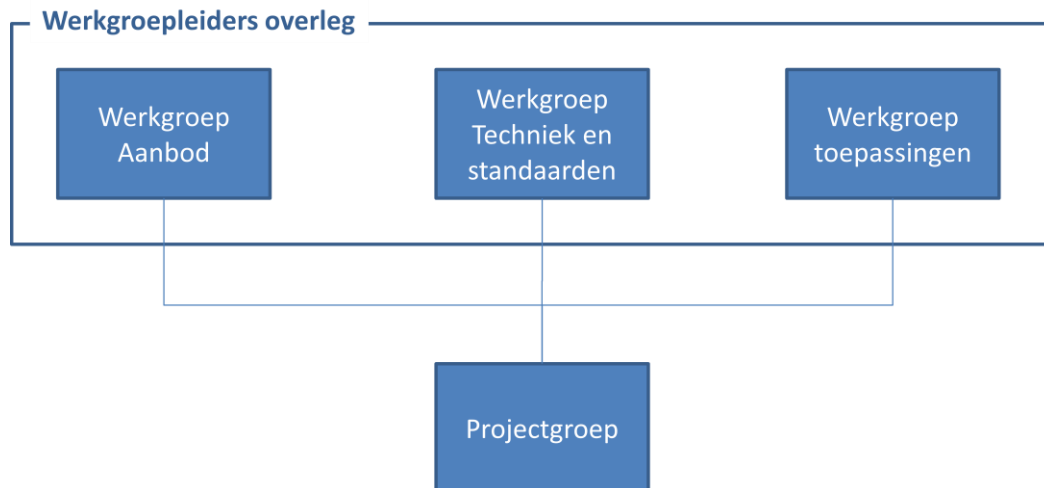
- Programma Inspire
- Programma stelsel van basisregistraties
- Referentiestelsel Gemeentelijke Basisgegevens (RSGB), KING
- Open data initiatieven gemeente Amersfoort en Nijmegen
- Onderzoek forum standaardisatie met de vraag welke technische standaarden het best kunnen worden gebruikt om open data te ontsluiten

3 Financien

Een belangrijk basisprincipe van de pilot is om te sturen op directe belangen van deelnemende partijen met betrekking tot LOD. Dit principe in combinatie met flexibiliteit (d.w.z. de pilot gaande weg richting kunnen geven op basis van wensen en input van deelnemende partijen) heeft er toe geleid dat de pilot zonder financiële compensatie wordt uitgevoerd.

4 Organisatiestructuur

Voor de duur van het project wordt de volgende tijdelijke organisatie opgezet:



Hierna worden de verantwoordelijkheden van de diverse betrokken groepen algemeen toegelicht.

Projectgroep	
Na(a)m(en)	Rol
Yvette Ellenkamp (Ministerie van I en M)	Projectgroeplid
Paul Francissen (gemeente Amersfoort)	Projectgroeplid
Paul Geurts (gemeente Nijmegen)	Projectgroeplid
Camille van der Harten (geo-business Nederland)	Projectgroeplid
Peter Klaver (KING)	Projectgroeplid
Arianne de Man (IPO)	Projectgroeplid
Marcel Reuvers (Geonovum, programma standaardisatie)	Projectgroeplid
Marijke Salters (Forum Standaardisatie)	Projectgroeplid
Arjen Santema (Kadaster)	Projectgroeplid
Imke Vrijling (programma Stelsel van basisregistraties)	Projectgroeplid
Sandra van Wijngaarden (Geonovum, programma Inspire)	Projectgroeplid

Hack de Overheid zal indien nodig afhankelijk van het onderwerp aanschuiven bij de projectgroep.

De projectgroep is eindverantwoordelijk voor het overall eindresultaat van het project en voor de coördinatie en communicatie buiten de pilot. De projectgroep is het

formele aanspreekpunt, faciliteert en borgt dat de afzonderlijke activiteiten resulteren in behalen van de LOD doelstellingen.

Projectleider & Project Support	
Na(a)m(en)	Rol
Erwin Folmer	Projectleider
Allen	

De projectleider krijgt van de projectgroep de verantwoordelijkheid en bevoegdheid voor de dagelijkse gang van zaken in het project. De projectleider kan beslissingen nemen binnen de grenzen zoals aangegeven door de projectgroep. De belangrijkste verantwoordelijkheid van de projectleider is het zekerstellen dat het project de gewenste producten oplevert, binnen tijd, binnen budget en volgens de afgesproken kwaliteit.

Samenstelling Clusters

De bemensing van onderstaande clusters worden tijdens de startbijeenkomst op woensdag 19 september besproken en ingevuld. Echter gedurende de looptijd kan de invulling van de clusters dynamisch zijn.

Activiteitencluster 1: Aanbod	
Na(a)m(en)	e-mail adres
Arjen Santema (leider)	

Activiteitencluster 2: Toepassingen	
Na(a)m(en)	e-mail adres
Paul Geurts (leider)	
Paul Francissen (leider)	

Activiteitencluster 3: Techniek en Standaarden	
Na(a)m(en)	e-mail adres
Wilko Quak (leider)	w.quak@tudelft.nl

De teams zijn verantwoordelijk voor het opleveren van producten. De werkgroep leider kan beslissingen nemen binnen de grenzen zoals aangegeven door de projectleider. De werkgroep leiders rapporteren regelmatig aan de projectleider tussentijds indien nodig en middels het periodieke overleg van de werkgroep leiders en de projectleider.

De groep deelnemers die niet deelneemt aan de activiteitclusters maar wel willen deelnemen aan de plenaire sessies zijn uiteraard van harte welkom. Hoe meer awareness, kennisdeling, hoe beter.

5 Initiële Projectplanning

De projectplanning is opgenomen in de bijlage.

