



Managementsamenvatting Verkenning PDOK *light*

1 Over de Verkenning

Deze Verkenning is uitgevoerd door het PDOK Adviesteam van Geonovum in opdracht van de Regiegroep PDOK, op verzoek van het GI Beraad. Input van gebruikers, data-aanbieders en marktpartijen is verzameld en meegenomen in deze Verkenning. Speciale aandacht gaat uit naar de samenhang met PDOK en de mogelijkheden tot hergebruik van bestaande voorzieningen (o.b.v. suggesties uit de marktconsultatie). De Verkenning is opgebouwd uit twee delen: in het eerste deel worden de functionele eisen aan PDOK *light* uitgewerkt (deze zijn vastgesteld door de PDOK Regiegroep), in het tweede deel wordt een mogelijke oplossingsrichting uitgewerkt om deze Verkenning te concretiseren. Over de te kiezen oplossingsrichting heeft nog geen besluitvorming plaatsgevonden. In de samenvatting gaan paragrafen 2 en 3 over het eerste deel, paragrafen 4 t/m 8 over het tweede deel.

2 Inhoud PDOK *light*

PDOK *light* is bedoeld als laagdrempelige, betaalbare oplossing voor het aanbieden van open geodata door overheidspartijen (incl. lagere overheden). PDOK *light* bestaat minimaal uit vier componenten:

- Self-service uploadvoorziening (waarin data-aanbieders zelf hun data uploaden en hun metadata invullen: eenvoudig tabblad met 10 metadata velden, aangevuld met een uploadknop)
- Publicatie functie (die na validatie van de metadata automatisch een Atom download aanmaakt van de dataset)
- Controle functie (die controleert of verwijzingen tussen metadata en data nog werken, of data nog voldoet aan de vooraf opgegeven 'houdbaarheidsdatum')
- Data opslag

Op deze wijze wordt de beheerlast (en daarmee de kosten) geminimaliseerd. De verantwoordelijkheden voor de data (correctheid van de data, aanbieden in open dataformaten) blijven nadrukkelijk decentraal bij de data-aanbieder zelf. De keuze voor downloads boven het aanbieden van webservices is zowel gebaseerd op de kosten als op het signaal dat veel gebruikers (zeker ontwikkelaars, maar ook marktpartijen) graag de data willen downloaden om zelf te hanteren in hun eigen app of geo-portaal.

Het zoeken en vinden van de open geodata vindt plaats op het Nationaal Georegister (NGR). Omdat gebruikers van open data vaker van buiten het traditionele geo-werkveld komen, wordt ingezet op het opzetten van een gebruiksvriendelijke, specifieke PDOK *light* user interface. Zo ontstaat binnen het NGR een lijn van doelgroepspecifieke user-interface, die enerzijds differentiëren in de specifieke interface, maar anderzijds ook duidelijk herkenbaar zijn als onderdeel van het NGR, zowel qua 'look and feel' als op technisch vlak.

Daarnaast wordt geëxperimenteerd met communityvorming: op deze wijze kunnen gebruikers (waaronder ook veel ontwikkelaars) elkaar helpen met *tips and tricks*, ervaringen en waarderingen. Denk voor het ideaalbeeld qua gebruikersvriendelijkheid aan een webwinkel als bol.com: daar vind je heel gemakkelijk niet alleen datgene wat je zoekt, maar zie je ook waarderingen (bijv. 4 v/d 5 sterren) en krijg je tips over mogelijk relevante andere informatie ("Andere mensen die dit product kochten, kochten ook...").

De doelgroep van PDOK *light* qua data-aanbieders wordt gevormd door alle overheden, waarmee PDOK *light* ook niet-landsdekkende data zal gaan bevatten.



3 Beleidskeuzes

Realisatie van PDOK *light* volgens de lijnen die in deze Verkenning worden geschetst, impliceert de volgende beleidskeuzes:

- Zowel het **aanbieden** als het **zoeken en vinden** van open geodata wordt structureel geborgd.
- PDOK *light* biedt alleen downloads aan, het aanbieden van services is aan PDOK en/of aan de markt.
- Het bieden van basale ontsluiting van open geodatasets is een overheidstaak.
- PDOK *light* richt zich op alle overheden en accepteert daarmee ook niet-landsdekkende data. Hiermee is de doelgroep van PDOK *light* breder dan de huidige doelgroep van PDOK.
- De verantwoordelijkheden voor correctheid van de data en de formaten waarin dit gebeurt, blijven decentraal bij de data-aanbieder. Belangrijkste argument hiervoor is de beperkte bijbehorende beheerinspanning, waarmee de kosten zo laag mogelijk blijven. Onvermijdelijk is dat de beheerder van PDOK *light* incidenteel (onterecht) aangesproken zal worden op de inhoud.

4 Techniek PDOK *light*

Het NGR werkt op basis van de open source software GeoNetwork. Veel van de functionaliteit die de bovengenoemde componenten vragen, biedt GeoNetwork al. Andere functionaliteit kan worden ontwikkeld op GeoNetwork en vervolgens aan deze community worden teruggegeven. Het aanbieden van Atom downloads kan gedaan worden met GeoServer. Zowel GeoNetwork als GeoServer worden nu al door PDOK gebruikt. Deze oplossingsrichting is ingebracht door één van de leden van GeoBusiness Nederland.

Met deze oplossing zal PDOK *light* dus zoveel mogelijk gebruik maken van functionaliteit van bestaande systemen. PDOK *light* is hiermee dus niet zozeer een op zichzelf staand technisch portaal, maar meer combinatie van een uitbreiding (data uploaden en doelgroepspecifieke interface) van het NGR met een organisatievorm waarin alle betrokken overheidspartijen zo hun open geodata kunnen aanbieden.

5 Realisatie PDOK *light*

Voorgesteld wordt om PDOK *light* projectmatig te ontwikkelen en te testen; om het vervolgens over te dragen aan de beheerorganisatie bij het Kadaster. Het opstellen van een projectplan voor een dergelijk project is de logische eerste vervolgstap.

6 Beheer PDOK *light*

PDOK *light* wordt grotendeels met GeoNetwork-functionaliteit gerealiseerd en komt functioneel dicht tegen het NGR aan te liggen. Voor het geautomatiseerd publiceren van datasets middels Atom downloads wordt GeoServer ingezet. Hiermee hergebruikt PDOK *light* maximaal bestaande PDOK functionaliteit. Het ligt hiermee voor de hand om het beheer van PDOK *light* bij het Kadaster onder te brengen.

Het communitydeel zal eerst op experimentele basis beheerd moeten worden. Daarom wordt geadviseerd het beheer hiervan projectmatig bij Geonovum onder te brengen. Dit kan als onderdeel van of in nauwe samenwerking met het PDOK Adviesteam worden ingericht.

7 Kosten en dekking

Kosten (nog op basis van ruwe schattingen, niet o.b.v. offertes):

- Realisatie PDOK *light* globaal 80 - 115 kEuro
- Beheerlast: 0,4 FTE + kosten data opslag (afhankelijk van keuze voor hosting in cloud of bij Kadaster intern)



Dekking (diverse partijen hebben aangegeven interesse te hebben, nog geen concrete toezeggingen):

- ICT Doorbraakproject Open geodata (bijdrage aan realisatie als impuls)
- Digitale Stedenagenda
- KING / VNG / Dataland
- PDOK partners

8 Vervolgtraject

- Instemming GI Beraad
- Opdracht tot opstellen projectplan Realisatie PDOK *light*
- Organiseren structurele dekking (Digitale Stedenagenda / KING / VNG / Dataland / Provincies / PDOK Partners)
- Goedkeuring projectplan, realiseren PDOK *light en* overdragen aan beheerorganisatie (bij Kadaster, bij of dicht tegen PDOK beheer)



Rapport

Verkenning PDOK *light*

Geonovum

datum

Oktober 2014

versie

versie 1.2



Inhoudsopgave

1	Inleiding	2
1.1	Aanleiding verkenning	2
1.2	Leeswijzer	2
2	Vraagstelling en uitgangspunten	3
2.1	Onderzoeksvraag	3
2.2	Context	3
3	Functionele eisen	4
3.1	Eisen vanuit het oogpunt van de gebruiker	4
3.2	Eisen vanuit het oogpunt van de data-aanbieder	5
3.3	Eisen vanuit het oogpunt van kwaliteit	5
3.4	Scenario PDOK <i>light</i>	5
4	Oplossingsrichting	7
4.1	Componenten van PDOK <i>light</i>	7
4.2	Analyse succesfactoren PDOK <i>light</i>	8
<i>Deel 2: mogelijke oplossing, financiën en vervolgtraject</i>		
<i>De functionele eisen zoals in deel 1 beschreven, zijn vastgesteld door de PDOK Regiegroep. Om deze Verkenning te concretiseren, is een oplossingsrichting verder uitgewerkt. Over deze uitwerking heeft nog geen besluitvorming plaatsgevonden.</i>		
4.3	Mogelijke oplossingsrichting op basis van GeoNetwork	10
5	Advies inrichting PDOK <i>light</i>	11
5.1	PDOK <i>light</i> algemeen	11
5.2	Techniek PDOK <i>light</i>	11
5.3	Beheer PDOK <i>light</i>	12
5.4	Realisatie PDOK <i>light</i>	12
6	Financiën	13
6.1	Kostenschattting realisatie PDOK <i>light</i>	13
6.2	Kostenschattting beheer PDOK <i>light</i>	13
6.3	Mogelijke dekking realisatie PDOK <i>light</i>	13
6.4	Mogelijke dekking beheer PDOK <i>light</i>	13
Bijlage 1	Deelnemers oplooppjes en geïnterviewden / reviewers	16
Bijlage 2	Opbrengst sessie Verkenning oplossingsrichtingen	17



Hoofdstuk 1

Inleiding

1.1 Aanleiding verkenning

Het ICT Doorbraakproject 'Open geodata als grondstof voor groei en innovatie' is gestart omdat het gebruik van open geodata achterblijft bij de potentiële mogelijkheden om economische waarde te creëren. Eén van de speerpunten van dit project is het ontzorgen van overheden op het gebied van het aanbieden van open data. Door met een centrale ontsluiting overheden te ontzorgen, wordt het aanbieden van open data laagdrempeliger gemaakt. Dit moet leiden tot een groter aanbod aan open data, met als uiteindelijk doel toenemend gebruik van open data. Cruciaal punt hierbij is dat de publicatiekosten van open data zo laag mogelijk gehouden worden, voornamelijk omdat nog onvoldoende duidelijk is of de baten van open data binnen of buiten de overheid op zullen treden. Om deze redenen heeft het doorbraakproject een voorstel voor onderzoek naar een PDOK-light voorziening ingediend bij het GI-beraad van 20 maart 2014. Een dergelijke voorziening zou onderdeel moeten gaan uitmaken van de nationale geo-informatie infrastructuur. Het GI-beraad heeft besloten tot een dergelijk onderzoek en de opdracht binnen de governance van PDOK bij Geonovum neergelegd.

Vanuit het PDOK Adviesteam bij PDOK is deze verkenning gehouden. Dit rapport presenteert de resultaten.

1.2 Leeswijzer

In hoofdstuk 2 wordt eerst de oorspronkelijke onderzoeksvraag weergegeven. Op basis van de aanvankelijke ideeën en een gebruikerssessie zijn functionele eisen aan PDOK *light* afgeleid. Deze eisen worden in hoofdstuk 3 behandeld en in context geplaatst in een gebruiksscenario. In hoofdstuk 4 worden uit deze eisen de functionele componenten en succesfactoren van PDOK *light* afgeleid. Deze zijn mede door een marktconsultatie tot stand gekomen. Deze sessie heeft ook tot de beoogde oplossingsrichting geleid. Op basis van deze resultaten wordt in hoofdstuk 5 een advies gegeven over de inrichting van PDOK *light*. Hoofdstuk 6 gaat specifiek in op de financiële aspecten van PDOK *light*.



Hoofdstuk 2

Vraagstelling en uitgangspunten

In dit hoofdstuk worden de onderzoeksvraag en onderzoekscontext beschreven. De onderzoeksopzet is geschreven op basis van een intakegesprek met vertegenwoordigers van het ICT Doorbraakproject Open geodata en vervolgens afgestemd met de PDOK Regiegroep, waarvan de voorzitter als opdrachtgever heeft gefungeerd.

2.1 Onderzoeksvraag

De centrale opdracht voor dit onderzoek luidt:

Verken hoe onder de vlag van PDOK een portaal voor open geo (overheids)data kan worden gerealiseerd?

Bij aanvang van deze verkenning was reeds een aantal uitgangspunten bekend:

- Centrale plek voor open overheidsdata, buiten de eigen infrastructuur van de dataproviders
- Kernwaarden: eenvoud, centraal, goedkoop, continuïteit, betrouwbaarheid
- Primair gericht op downloaden van data, geen uitgebreide functionaliteit zoals API's e.d. Het creëren van meerwaarde (o.a. in de vorm van hogere serviceniveaus) dient aan PDOK of de markt overgelaten te worden
- Upload- en beheerfunctionaliteit via portaal (self-service), om zo de beheerlast zo klein mogelijk te houden
- Positionering: als PDOK *light*, want PDOK is een sterk merk
- Doelgroep: gehele overheid, zowel centraal als decentraal (discrepantie t.o.v. huidige doelgroep PDOK)
- Toekomstvast: duurzaam portaal, voorbereid op ontwikkelingen als linked data, json, self-servcie, etc.
- Breed gedragen oplossing (gedragen door leden GI-beraad)
- Aanbieden en gebruiken van datasets is kosteloos
- Voor implementatie en beheer van PDOK *light* dient een financieringsvoorstel te worden vastgesteld.
- Gebruikers faciliteren, bijv. via forum, community-vorming door bijv. taggenen waarderen van data
- Verantwoordelijkheden (voor metadata, voor actualiteit, voor formaten, etc.) moeten decentraal bij de data-providers liggen en niet centraal bij een beheerorganisatie, om zo de beheerlast beperkt te houden.

2.2 Context

Bij de opdrachtverlening is duidelijk de behoefte uitgesproken het open data portaal onder de vlag van PDOK te positioneren, omdat PDOK inmiddels een sterk merk is geworden. Alhoewel de huidige inrichting van PDOK niet direct richtinggevend is voor de inrichting van PDOK *light*, is wel meegegeven om scherp te zijn op kansen om componenten in synergie in te zetten of te ontwikkelen. Ook de financiering van ontwikkeling en beheer van PDOK *light* wordt meegenomen.



Hoofdstuk 3

Functionele eisen

In meerdere sessies met gebruikers, data aanbieders en marktpartijen is verkend welke functionele eisen er gesteld kunnen worden aan PDOK *light*. Vervolgens is deze lijst uitgebreid op basis van best practices en consultaties van experts. Bij deze verkenning zijn zowel de gemeentelijke, provinciale en rijksoverheid betrokken om zo zoveel mogelijk draagvlak, ideeën en initiatieven te verzamelen.

3.1 Eisen vanuit het oogpunt van de gebruiker

Bij een open data portaal kunnen er twee typen gebruikers worden onderscheiden: de eindgebruiker en de gebruiker die de data vervolgens wil verrijken of op andere wijze ontsluiten, om zo de eindgebruiker beter te bedienen.

- Gebruik is gratis
- Centraal portaal i.p.v. decentrale portalen (bij overheden), want geen enkele gebruiker zit te wachten op 500 afzonderlijke open data portalen.
- Continuïteit: portaal bestaat over 5 jaar nog steeds. Dit biedt zowel de data-aanbieder als de gebruikers zekerheid over het aanbod, waarmee gebruik gestimuleerd wordt.
- Beknopte set metadata (minimaal titel, contactgegevens, licentievorm, bronhouder, houdbaarheidsdatum)
- Datasets vindbaar in Nationaal GeoRegister: gebruikers hebben behoefte aan een duidelijke plek waar zij data kunnen zoeken en vinden.
- Vereenvoudiging zoeken en downloaden: idealiter is het zoeken naar data net zo gebruiksvriendelijk en intuïtief als het zoeken van producten in een webwinkel, denk bijvoorbeeld aan bol.com of booking.com
- Community functies voor communiceren, delen en 'liken' datasets (laat gebruiker de kwaliteit bepalen en elkaar helpen).
- Zekerheid dat data actueel is, evt. aangevuld met goed herkenbare tijdreeksen voor historie
- Data(updates) ook geautomatiseerd te downloaden
- Bladerfunctie naast zoekfunctie voor downloads
- Atom downloads, zodat data ook geautomatiseerd opgehaald kan worden (ook door zwaardere portalen, bijv. CitySDK van Waag Society)
- Geen belemmeringen als identificatie voor gebruikers: gebruik van open data is zo laagdrempelig mogelijk, bovendien beperkt het de beheerlast.
- Open dataformaten
- Ook niet-landsdekkende bestanden zijn toegestaan, want juist bij de lagere overheden is een enorme behoefte geconstateerd aan ontzorging op dit punt.
- Persistente URL's
- Wens: mutatiemelding op datasets
- Wens: betekenisvolle URL

Het is een bewuste en nadrukkelijke keuze geweest om PDOK *light* alleen te richten op data downloads en niet op het inrichten van webservices. Een enkeling heeft wel aangegeven dergelijke services als wenselijk te zien, maar de ruime meerderheid is van mening dat het aanbieden van services een taak is die kan worden opgepakt door PDOK of de markt. Daarnaast maakt het aanbieden van services de voorziening complexer, zowel in technisch als in beheeropzicht. Die complexiteit staat op gespannen voet met het uitgangspunt van een goedkoop en laagdrempelig portaal.



3.2 Eisen vanuit het oogpunt van de data-aanbieder

- Portaal buiten eigen infrastructuur
- Aanbieden van data kosteloosom zo geen belemmeringen te creëren voor het aanbieden van meer datasets als open data.
- Laagdrempelige real time self-service uploadvoorziening (inloggen, zelf metadata invullen of uploaden, zelf dataset uploaden). Bij voorkeur worden metadata velden die automatisch gevuld kunnen worden (o.b.v. gebruikersprofiel) reeds vooraf ingevuld, om hierdoor zo laagdrempelig mogelijk te werken.
- Automatische upload mogelijk voor o.a. periodieke datareeksen (API).
- Mogelijkheid voor opbouwen tijdreeksen, denk bijvoorbeeld aan jaarlijkse luchtfoto's van een bepaald gebied.
- Onweerlegbaarheid van de data: identificatie en authenticatie voor uploadvoorziening

3.3 Eisen vanuit het oogpunt van kwaliteit

- Alle data is voorzien van (beknopte) metadata, minimaal verplichte velden invullen (groot deel kan standaard worden ingevuld, < 10 velden handmatig)
- Metadata gevalideerd voor publicatie van data
- Automatische controle op 'broken links' (van download link in metadata): broken link betekent niet langer zichtbaar maken van data en metadata en na bepaalde periode verwijdering.
- Geen handmatige kwaliteitscontroles, wel vooraf aanbieden van kwaliteitsrichtlijnen: PDOK *light* biedt hoogstens handvaten (in de vorm van richtlijnen / documentatie) om data goed aan te bieden, maar de verantwoordelijkheid voor de kwaliteit blijft volledig bij de aanbiedende partij liggen.
- Optie om gebruikers data te laten beoordelen: kwaliteitsindicatie, vergelijkbaar met bijvoorbeeld waardering voor hotels op booking.com

3.4 Scenario PDOK *light*

Om alle functionele eisen in de juiste context te kunnen plaatsen, is op basis van de input van de rondetafeldiscussies onderstaand scenario geschreven. In dit scenario wordt duidelijk hoe PDOK *light* zou moeten functioneren.

Aanbieder A (een ministerie) is eigenaar van een dataset die eenmalig is ingewonnen, maar niet wordt geactualiseerd. Volgens het open databeleid van de aanbieder wordt deze dataset als open data geclassificeerd. Aanbieder A besluit vervolgens om de dataset op PDOK light te publiceren. Hiertoe logt Aanbieder A in op de uploadvoorziening van PDOK light. Dankzij deze login wordt alleen data gepubliceerd onder de naam van Aanbieder A die ook echt van de aanbieder is. Op het portaal vult de aanbieder een aantal gegevens in over de dataset (de metadata). De aanbieder had ook kunnen kiezen om deze metadata te uploaden vanuit zijn eigen metadata systeem. Een deel van de metadata is reeds ingevuld op basis van het gebruikersprofiel (defaultwaarden voor o.a. aanbiedende organisatie, contactgegevens, etc.), waardoor de aanbieder nog slechts een paar velden handmatig hoeft in te vullen. Zo geeft de aanbieder aan dat de dataset nog drie jaar relevant is, maar daarna eigenlijk achterhaald is. Vervolgens uploadt de aanbieder de dataset. Op basis van het document met aanbevelingen voor data-aanbieders heeft de aanbieder gekozen voor een open formaat. Tot slot klikt de aanbieder op Publiceer.

Zodra de aanbieder zijn stappen heeft doorlopen en op Publiceer heeft geklikt, controleert het systeem als eerste of de metadata volledig en valide is. Dit blijkt het geval te zijn er wordt vervolgens automatisch een download link en Atom feed aangemaakt. Nu worden de metadata en



de URL van de download in het Nationaal Georegister (NGR) geregistreerd. Als de metadata niet volledig of valide was geweest, had het systeem de bijbehorende dataset ook niet gepubliceerd. KOOP, de partij achter data.overheid.nl, harvest de metadata uit het NGR en zo komt de metadata van de open datasets ook in data.overheid.nl beschikbaar.

Gebruiker 1 (een waterschap) is op zoek naar data. In het NGR vindt deze gebruiker de dataset van aanbieder A. Dankzij de metadata is het direct duidelijk wat de gebruiker van de dataset kan verwachten. Gebruiker 1 gebruikt de download URL om de data te downloaden en kan aan de slag.

Gebruiker 2 (een softwareleverancier) ontdekt dezelfde dataset en schat in dat veel klanten met deze data in zijn software aan de slag willen gaan. Om dat nog verder te vereenvoudigen, downloadt gebruiker 2 deze data, converteert deze naar het eigen opslagformaat en ontsluit deze gebruiker de data vervolgens voor alle gebruikers van de software.

Aanbieder B (een gemeente) beheert een dataset die in de eigen werkprocessen doorlopend wordt bijgewerkt. De gemeente wil deze dataset jaarlijks publiceren als open data, waarbij de versies van de voorgaande jaren beschikbaar blijven. In het eerste jaar vult aanbieder B de metadata in het portaal en uploadt de eerste versie. Vervolgens wordt elk jaar de nieuwe, actuele versie toegevoegd en de oude versie met een jaartal gelabeld.

Gebruiker 3 (een wetenschapper) vindt jaren later deze data in het NGR en kan direct de verschillende jaargangen downloaden. Deze tijdreeks vormt een interessante bron van analyses voor de wetenschapper.

Het systeem voert periodiek consistentiecontroles uit op de metadata en controleert hierbij ook op broken links. Wanneer broken links worden geconstateerd, wordt de data of de metadata op niet-zichtbaar geplaatst en de data-aanbieder automatisch gemaild. Na bepaalde tijd wordt de data of de metadata verwijderd. Zo wordt voorkomen dat het NGR en PDOK licht vol komen te staan met onbruikbare data.



Hoofdstuk 4

Oplossingsrichting

Uit de functionele eisen aan PDOK *light* worden in dit hoofdstuk de componenten van PDOK *light* afgeleid. Vervolgens wordt geanalyseerd welke succesfactoren PDOK *light* kent en wordt een oplossingsrichting gepresenteerd die maximaal gebruik maakt van bestaande voorzieningen.

4.1 Componenten van PDOK *light*

Op basis van de functionele eisen uit het voorgaande hoofdstuk, kunnen de componenten worden afgeleid waar PDOK *light* minimaal uit zou moeten bestaan.

- Self-service uploadvoorziening
- Publicatie functie
- Controle functie
- Data opslag

Met component wordt hier niet per se een technische component bedoeld, maar vooral een geclusterde hoeveelheid functionaliteit. Op basis van hun functionaliteit, kunnen deze vier componenten als volgt worden omschreven:

Self-service uploadvoorziening

Plek waar data-aanbieders zich minimaal kunnen identificeren middels een login, hun metadata invullen dan wel uploaden en hun datasets kunnen uploaden en beheren. Ook geautomatiseerde uploads (periodieke actualisaties) zijn mogelijk, dankzij een API. De voorziening is ook in staat tijdreeksen van bestanden op te bouwen, om zo gebruikers ook historische data te bieden.

Publicatiefunctie

Nadat de data-aanbieder een dataset heeft geupload en valide metadata heeft aangeleverd, wordt automatisch een Atom download aangemaakt en de metadata met daarin de link gepubliceerd in het NGR.

Controlefunctie

Periodiek wordt gecontroleerd of aangeboden data nog binnen de opgegeven houdbaarheidsdatum valt en of de URL's naar de data en metadata nog werken. Broken links of verouderde data leiden tot het verwijderen / niet zichtbaar maken van de Atom download op PDOK *light* en de metadata in het NGR.

Data opslag

Alle open data wordt in een data store opgeslagen. Het is voor deze verkenning niet relevant waar en hoe deze data store wordt ingericht, zolang de data store voldoende schaalbaar en betaalbaar is. Schaalbaar, zodat met minimale beheerinspanning voldoende opslagcapaciteit beschikbaar is en betaalbaar, zodat de kosten voor PDOK *light* niet hoger oplopen dan strikt noodzakelijk.

Naast deze vier minimaal vereiste componenten is er een vijfde component te identificeren:

Communityvoorziening

Het vrijgeven van data kan leiden tot breder gebruik, dat vaak ook buiten de traditionele gebruikersgroepen komt te liggen. Juist onder die nieuwe gebruikers (vaak ontwikkelaars) bestaat behoefte om elkaar op weg te kunnen helpen met *tips and tricks*, ervaringen en aanbevelingen. Het faciliteren van communityvorming vergroot het oplossend vermogen van gebruikers, waarmee de noodzaak voor 'traditionele' ondersteuning (zoals helpdesk e.d.) afneemt. Een actieve community is de beste garantie dat voldoende data wordt aangeboden en de juiste



kwaliteit krijgt. Dit moet ondersteund worden door een goed portaal, waarbij zoals bij social media en webshops plaats is voor opmerkingen, liken en delen van ervaringen.

4.2 Analyse succesfactoren PDOK *light*

Vanuit de functionele eisen en de randvoorwaarden aan / beoogde kenmerken van PDOK *light* is een aantal criteria te formuleren die richting kunnen geven aan de te kiezen oplossing. Een van de essentiële succesfactoren van PDOK *light* zijn de **kosten**. Het gaat hierbij zowel om realisatiekosten en beheerkosten van PDOK *light* (en de functionele eis dat het aanbieden van data kosteloos is). Om die kosten zo laag mogelijk te houden, dient de oplossing zoveel mogelijk te streven naar hergebruik van systemen (hard- en software), functionaliteit (bijv. door identificatie van NGR te gebruiken) en organisatie (synergie tussen PDOK en PDOK *light* beheer). Communityvorming helpt niet alleen bij het betrekken van gebruikers, maar leidt ook tot een kostenbesparing op functioneel beheer: binnen de community zullen gebruikers elkaar kunnen helpen en vragen beantwoorden.

Een andere succesfactor is de mate waarin PDOK *light* voorziet in de **behoefte** van de deelnemende partijen. Juist ook onder decentrale overheden leeft de behoefte aan een eenvoudige voorziening, waarbij het er op lijkt dat de gemeenten die tot nu toe in eigen portalen investeren, daar een veelvoud van het benodigde PDOK *light* budget aan uitgeven.

De derde succesfactor is **gemak**. Dit omvat zowel het gemak voor de data-aanbieder (eenvoudig data toe te voegen, metadata zoveel mogelijk invullen op basis van het gebruikersprofiel, mogelijkheid om automatisch nieuwe versies (via een API) up te loaden) als voor de gebruiker (eenvoudig data vinden, gebruiken, zekerheden over actualiteit van data). Om aanbieders en gebruikers dat gemak te bieden, zal PDOK *light* zo min mogelijk als nieuwe, extra omgeving gezien moeten worden. Op basis van de functionele eisen is het mogelijk dat de gebruikersinterface van PDOK *light* bij wijze van spreken niets meer is dan een upload-knop in het NGR of op de website www.pdok.nl. Functionaliteit voor het editen of uploaden van metadata en voor het zoeken van data is immers reeds beschikbaar. PDOK *light* is daarmee niet zozeer een op zichzelf staand product (technisch platform), maar eerder een organisatievorm om met een beperkt aantal technische aanpassingen het gewenste open data portaal te realiseren binnen bestaande omgevingen. Om de gepubliceerde geodata te zoeken en te vinden is een gebruikersvriendelijk portaal een onderdeel van de succesfactor 'gemak'. Daarbij is er geen behoefte aan een extra portaal, alleen aan een nieuwe user interface specifiek voor PDOK *light* (naast de huidige interface).

Vanuit de succesfactoren gemak en kosten, is ook een passend beveiligingsniveau nodig met de juiste kwaliteit. Aangezien PDOK *light* bedoeld is voor open data, volstaat conform de methodiek van E-herkenning username/password als beveiligingsmaatregel voor data aanbieders. Data-afnemers (die data zoeken en downloaden) hoeven zich niet te identificeren. Dit houdt de toegang laagdrempelig en de beheerlasten laag. Om toch voeling te houden met de data-afnemers, kan communityvorming worden ingezet.



Deel 2: mogelijk oplossing, financiën en vervolgtraject

In het voorgaande deel zijn de functionele eisen voor PDOK *light* afgeleid. Deze zijn door de PDOK Regiegroep vastgesteld in september 2014. Om ook in deze Verkenning al uitspraken te kunnen doen over kosten, wordt een mogelijke oplossingsrichting op basis van GeoNetwork verder uitgewerkt. Deze oplossingsrichting helpt om deze Verkenning concreter te maken. Definitieve keuzes over de uitwerking van PDOK *light* zijn nog niet genomen.



4.3 Mogelijke oplossingsrichting op basis van GeoNetwork

Tijdens één van de sessies rond PDOK *light* is de suggestie gedaan om de functionaliteiten van GeoNetwork (de software waarop het NGR draait) beter te benutten, omdat men stelde dat de meeste functionaliteit van PDOK *light* al gerealiseerd kan worden op basis van GeoNetwork. Dat maakt dat deze oplossing een beperkte ontwikkelinspanning vraagt. De functionele eisen aan de uploadvoorziening voor PDOK *light* vertonen veel overeenkomsten met de huidige NGR-functionaliteit, waarbij data-aanbieders inloggen en hun metadata kunnen invullen of uploaden. Het is nu al mogelijk om bij die metadata ook een dataset te uploaden. Ook voor geautomatiseerde uploads biedt GeoNetwork via een aantal API's al functionaliteit, al is hier vermoedelijk nog een beperkte ontwikkelinspanning nodig om exact te voldoen aan de eisen vanuit PDOK *light*. Eind 2014 komt GeoNetwork 3 uit, waarin o.a. een verbeterde engine zit voor het uitvoeren van consistency checks. Hiermee wordt aan de belangrijkste eisen aan de beheer- en controlecomponent voldaan. Ook het publiceren is mogelijk, GeoNetwork kan zowel Atom downloads aanmaken als (via een gekoppelde Geoserver – ook in gebruik bij PDOK) WMS of WFS publiceren, mocht daar in een later stadium behoefte aan ontstaan.

Tevens bieden de huidige ontwikkelingen van GeoNetwork een interessant perspectief voor doorontwikkeling van PDOK *light* op de langere termijn. Nu reeds is de community rond GeoNetwork bezig met het ontwikkelen van functionaliteit om de datacatalogus doorzoekbaar te maken voor Linked Data zoekmachines en om data die geupload wordt in bepaalde gesloten formaten –na validatie– om te vormen tot open formaten als CSV, GeoJSON en KML. Deze formaten zijn voor veel app-ontwikkelaars buiten de traditionele geo-community beter hanteerbaar dan de huidige geo-bestandsformaten.



Hoofdstuk 5

Advies inrichting PDOK *light*

In dit hoofdstuk wordt een aantal adviezen uitgebracht over hoe de benodigde PDOK *light* functionaliteit gerealiseerd en beheerd kan worden. Op basis van deze adviezen dient na instemming door alle betrokken partijen een projectplan voor de daadwerkelijke bouw van PDOK *light* opgesteld te worden.

5.1 PDOK *light* algemeen

Los van technische vragen zijn er op een aantal algemene punten advies te geven:

Doelgroep PDOK light

PDOK *light* moet zich richten op alle open geodata van overheden. Juist ook gemeenten werken hard aan open data en dit dreigt nu te leiden tot een kluwen aan lokale open data-portalen. Deze kluwen belemmert de vindbaarheid en is onnodig duur. PDOK *light* zal hiermee dus ook veel niet-landsdekkende data gaan bevatten. Dit is een andere lijn dan PDOK op dit moment hanteert, maar wel weer in lijn met het NGR, dat ook voor lokale datasets bedoeld is. Heldere communicatie en het goed inzetten van de merken PDOK (voor landsdekkende services) en PDOK *light* (voor al dan niet landsdekkende downloads) kan verwarring voorkomen. Tegelijk is het te voorzien dat PDOK vroeg of laat ook de vraag krijgt of er ook niet-landsdekkende services aangeboden kunnen gaan worden.

Taak voor overheid of overlaten aan markt?

Het structureel borgen van de ontsluiting van open overheids geodata is een taak die bij de overheid thuishoort: de overheid borgt zelf de transparantie en kwaliteit van de data. Er zijn marktpartijen die aangeboden hebben om de open data zelf te ontsluiten; van dit aanbod wordt geen gebruik gemaakt. Marktpartijen moeten wel in staat worden gesteld om de data automatisch over te nemen in eigen portalen en/of applicaties (en er meerwaarde op aan te bieden), ook wanneer er nieuwere versies van datasets beschikbaar komen.

Verantwoordelijkheden

De keuze om zoveel mogelijk functies te automatiseren en de beheerlast zo laag mogelijk te maken, leidt ertoe dat de verantwoordelijkheden voor de data altijd bij de data-aanbieder zullen blijven liggen en niet bij PDOK *light* terecht komen. De data-aanbieder moet zorgen voor valide data en bij voorkeur ook voor het aanbieden in open formaten. PDOK *light* kan hier niet op controleren, maar slechts aanbevelingen doen vooraf aan data-aanbieders.

5.2 Techniek PDOK *light*

In het vorige hoofdstuk zijn vijf componenten van PDOK *light* geïdentificeerd, nl. de self-service uploadvoorziening, publicatiefunctie, controlefunctie, dataopslag en community. De eerste drie componenten zijn te realiseren op basis van GeoNetwork. De huidige identificatie en authenticatie van data-aanbieders van het NGR kan worden hergebruikt voor PDOK *light* data-aanbieders: zij kunnen met de huidige NGR login voortaan niet alleen metadata invullen, maar ook direct hun open data uploaden naar PDOK *light*. Publicatie kan het beste in de vorm van Atom downloads. Het grote voordeel hiervan is dat andere portalen geautomatiseerd de data en eventuele updates kunnen binnenhalen. Deze Atom downloads kunnen vanuit GeoNetwork geautomatiseerd worden aangemaakt binnen GeoServer.



Waar de huidige versie van GeoNetwork nog functionaliteit mist, dient deze functionaliteit te worden ontwikkeld en beschikbaar gesteld te worden aan de GeoNetwork community. Op deze wijze komt de benodigde functionaliteit duurzaam beschikbaar binnen GeoNetwork.

De component data-opslag is eenvoudig realiseerbaar, al dient er voldoende aandacht te zijn voor schaalbaarheid en monitoring. Binnen deze Verkenning wordt geen advies gegeven over de te kiezen opslagoplossing. Er kan voor gekozen worden om de data-opslag van PDOK te gebruiken, maar ook cloudoplossingen zijn mogelijk. Afhankelijk van de gekozen oplossing, is er meer of minder controle over waar de data opgeslagen wordt en zullen de kosten hoger of lager uitvallen. Dit is een beleidsoverweging en geen keuze die puur uit de inhoud volgt.

Met het realiseren van de eerste vier componenten vormt PDOK *light* een laagdrempelige en betaalbare voorziening voor het **aanbieden** van open geodata. Voor het gebruik is het **zoeken en vinden** van open data belangrijk. Het is in principe mogelijk om het huidige NGR te gebruiken voor het zoeken en vinden van de gewenste data. Om het gebruiksgemak en de herkenbaarheid te vergroten en daarmee ook het gebruik te stimuleren, is het goed om een specifieke PDOK *light* user interface te bouwen.

De user interface valt uiteen in twee onderdelen. Ten eerste wordt geadviseerd om bij de ontwikkeling van de PDOK *light* user interface specifiek aandacht te schenken aan de gebruiksvriendelijkheid. Hiervoor dient een specialist op het gebied van user interaction betrokken te worden bij de ontwikkeling van de interface. Zijn input is idealiter ook herbruikbaar voor de huidige user interface. Ten tweede wordt geadviseerd om ook de vijfde component, de community, te realiseren. Hiervoor kan bijvoorbeeld GitHub ingezet worden. Omdat communityvorming nog relatief nieuw is, wordt geadviseerd om de communityvorming rond PDOK *light* eerst voor een periode van twee jaar op experimentele basis uit te voeren en onder te brengen bij het PDOK Adviesteam van Geonovum. Geonovum is in de positie om de juiste expertise (van o.a. de VU Amsterdam) gericht in te zetten en de community verder te ontwikkelen.

5.3 Beheer PDOK *light*

PDOK *light* wordt grotendeels met GeoNetwork-functionaliteit gerealiseerd en komt functioneel dicht tegen het NGR aan te liggen. Voor het geautomatiseerd publiceren van datasets middels Atom downloads wordt GeoServer ingezet. Hiermee hergebruikt PDOK *light* maximaal bestaande PDOK functionaliteit. Het ligt hiermee voor de hand om het beheer van PDOK *light* bij het Kadaster onder te brengen. Gezien de verschillen in doelgroep (ook niet-landsdekkende data van lokale overheden) zal PDOK *light* mogelijk (deels) andere foundingfathers hebben dan PDOK zelf. Het beheer van PDOK *light* zou daarom formeel gescheiden kunnen blijven van het beheer van PDOK, waarbij de afstand tussen beide onderdelen in de praktijk minimaal is (m.n. op sturingsniveau zou er verschil kunnen zijn). Indien voor een dergelijke formele scheiding wordt gekozen, kan men overwegen om de beheertaken voor het NGR over te hevelen van PDOK beheer naar PDOK *light* beheer, aangezien de bredere doelgroep van PDOK *light* beter aansluit bij de doelgroep van het NGR.

Aangezien het voorstel is om het communitydeel eerst op experimentele basis uit te voeren, wordt geadviseerd het beheer hiervan projectmatig bij Geonovum onder te brengen. Dit kan als onderdeel van of in nauwe samenwerking met het PDOK Adviesteam worden ingericht.

5.4 Realisatie PDOK *light*

Deze Verkenning levert nog geen volledig plan voor realisatie van PDOK *light* op. Geadviseerd wordt om PDOK *light* projectmatig te ontwikkelen en te testen. Het opstellen van een projectplan voor een dergelijk project is de logische eerste vervolgstap.



Hoofdstuk 6

Financiën

In deze Verkenning is dit hoofdstuk gebaseerd op schattingen (o.b.v. overleg met experts). Het nog op te stellen projectplan Realisatie PDOK *light* zal zich baseren op offertes en concrete toezeggingen.

6.1 Kostenschatting realisatie PDOK *light*

PDOK *light* maakt zoveel mogelijk gebruik van bestaande voorzieningen. Hierdoor blijven de ontwikkelingskosten beperkt. De globale kostenschatting is als volgt:

- Uitbreiden functionaliteit GeoNetwork, incl. inzet user interaction expert en verbeteren interface € 45.000 - 65.000
- Testen en acceptatie € 25.000 - 35.000
- Realisatie community € 10.000 - 15.000

Kostenrange: € 80.000 - € 115.000.

De risico's worden voornamelijk bepaald door de deskundigheid van de uiteindelijke bouwer (mate van kennis en ervaring van GeoNetwork), de hoeveelheid functionaliteit die in reguliere nieuwe versies van GeoNetwork beschikbaar komen en de doorlooptijd van het testen en acceptatie.

6.2 Kostenschatting beheer PDOK *light*

De beheerlast van PDOK *light* wordt als volgt ingeschat (op jaarbasis):

- Functioneel en technisch beheer 0,4 FTE
- Experimenteel beheer community onderdeel maken van PDOK Advies
- Hosting data PDOK *light* P.M.
(voorstel: PDOK Beheer maakt offerte voor data-opslag bij PDOK en voor externe hosting via raamwerk / vrije aanbesteding (globale schatting cloudoplossing € 20.000))

De bandbreedte in de kosten wordt voor een aanzienlijk deel bepaald door de keuze van data opslag. In mindere mate is ook het succes van PDOK *light* een risico: hoe groter het succes, hoe meer datasets aangeboden zullen worden en hoe meer storage benodigd is.

6.3 Mogelijke dekking realisatie PDOK *light*

Te benaderen partijen (hebben deels wil uitgesproken om deel te financieren, maar nog geen concrete toezeggingen):

- ICT Doorbraakproject Open geodata
- Digitale Stedenagenda
- KING / VNG / Dataland
- PDOK partners
- Provincies

6.4 Mogelijke dekking beheer PDOK *light*

Te benaderen partijen (hebben deels wil uitgesproken om deel te financieren, maar nog geen concrete toezeggingen):



- Digitale Stedenagenda
- KING / VNG / Dataland
- PDOK partners
- Provincies



Bijlage 1

Deelnemers oploopjes en geïnterviewden / reviewers

Verkenning Oplossingsrichtingen:

- Reind van Olst – 2CoolMonkeys
- Agustin Matilla Sanz – Geodan
- Richard Duivenvoorde – QGIS community
- Dirk-Jan Bossink – Nieuwland
- Marcel Steenis – Grontmij
- Ronald van Lanen – Grontmij
- Jeroen van Winden – ESRI
- Camille van der Harten – GeoBusiness + Doorbraakteam
- Paul van Genuchten – Geocat
- Jeroen Ticheler – Geocat
- Cees Guikers – BridGis
- Jan-Maarten Verbree – Nelen & Schuurmans

Verkenning Functionele eisen

- Tom Demeyer – Waag Society
- Frans Lips – Doorbraakteam
- Jan van Sambeek – GBO Provincies
- Rob Bots – Digitale Stedenagenda
- Pieter van Teefelen – DataLand
- PDOK Beheer



Bijlage 2

Opbrengst sessie Verkenning oplossingsrichtingen

1 Hoofdpijnen van discussie

Ronde visie op PDOK *light* concept:

- Focus ligt op data, dat vereist metadata en performance (incl. continuïteit)
- Commerciële toepassingen o.b.v. open data alleen haalbaar als data landsdekkend is. Daarnaast aandacht voor continuïteit en kwaliteitsdocumentatie
- Focus op link met NGR, denk aan interfaces voor niet-specialisten
- Overweeg gebruik GitHub
- Markt-overheid: wie doet wat? Richt je op gebruik buiten de overheid.
- Aanbieden kan ook overgelaten worden aan marktpartijen (zeker partijen met veel klanten binnen overheid). Zo niet, dan eis dat data geautomatiseerd uit PDOK *light* gehaald moet kunnen worden.
- Wees terughoudend met omvang aanbod, anders sta je creatie van economische meerwaarde in de weg
- Laat gebruiker bepalen wat kwaliteit is
- Realiseer met standaard software
- Focus op kwaliteit, automatisch data binnen kunnen halen, NGR als zoekingang
- Services kunnen ook meerwaarde bieden, zeker bij basiskaarten
- Open formaten hebben de voorkeur
- Uploaden verdient aandacht, gaat vaak veel langzamer dan downloaden
- Kijk ook naar IGN: meerdere lokale interfaces op zelfde technische back-end

Discussie over metadata:

- Focus op kleine kernset, in elk geval licentievorm, titel, houdbaarheidsdatum, point of contact, bronhouders
- Liever minder ingevulde metadata elementen dan slecht ingevulde metadata elementen
- Dilemma: beginnen met kleine set die je later moeilijk uit kunt breiden, of direct iets grotere metadata set vragen

Ronde visie op oplossingsrichting:

- Metadata is belangrijk, maar verplicht niet alles. Suggestie: kijk naar product Robin
- Meest *light*-variant is WeTransfer
- Decentraliseren data store ook overwegen (elke gemeente eigen portaal)
- Zorg voor goede wizard om vullen PDOK *light* goed te ondersteunen
- Suggestie: kijk naar product Voyager
- Suggestie: kijk naar product ArcGIS Online
- Bouw geen nieuw systeem, maar gebruik functionaliteit van software die je al hebt: kies voor GeoNetwork
- Pas op voor lokale overheden: ze willen soms te veel toeters en bellen op data portaal
- Investeer juist in de voorkant (goede interface), minder in de achterkant (hergebruik)
- Positioneer duidelijk t.o.v. LinkedData en internationale ontwikkelingen
- Overschat de impact niet, de markt in Nederland is klein om nieuwe producten o.b.v. open data te ontwikkelen, zeker als die niet landsdekkend zijn
- Moet je nu wel iets doen? Kun je niet op de markt wachten?

2 Analyse opbrengst / verantwoording keuzes

In functionele zin worden de kenmerken laagdrempelig en goed doorzoekbaar als speerpunten aangehaald. Diverse producten worden gesuggereerd als oplossing, inclusief het volledig overlaten aan de



markt van het aanbieden van open data. Tegelijk wordt er ook krachtig op gewezen dat de overheid al veel benodigde functionaliteit in huis heeft, maar dit alleen niet gebruikt.

Niet alle suggesties uit deze sessie zijn verwerkt in de Verkenning, maar veel wel:

- Oplossingsrichting hergebruik GeoNetwork
- Investeren in laagdrempelige interface, gericht op niet-specialistische gebruiker
- Faciliteer automatisch harvesten van datasets in PDOK *light*
- Laat aanbieden van services en creatie van meerwaarde over aan de markt (/PDOK)
- Zorg voor laagdrempelige upload: wizard o.b.v. gebruikersprofiel om minimale set metadata zo makkelijk en goed mogelijk te vullen
- Zoeken op basis van NGR
- Ontwikkelen community mede op basis van GitHub

Twee suggesties zijn niet overgenomen:

- Overlaten aanbieden basale ontsluiting open overheidsdata aan markt: het is een expliciete beleidskeuze dat dit een overheidstaak is
- Laten ontstaan van tal van lokale dataportalen: de kosten hiervoor zijn onevenredig hoog