

**Hergebruikers van open data in beeld
onderzoek i.o.v. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties**

Welle Donker, Frederika; Braggaar, Rob; van Loenen, Bastiaan

Publication date

2019

Document Version

Final published version

Citation (APA)

Welle Donker, F., Braggaar, R., & van Loenen, B. (2019). *Hergebruikers van open data in beeld: onderzoek i.o.v. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties*. Kenniscentrum Open Data, TU Delft. <https://www.rijksoverheid.nl/documenten/rapporten/2019/07/24/hergebruikers-van-open-data-in-beeld>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Hergebruikers van open data in beeld

dr. ir. F. Welle Donker, ir. R.C. Braggaar, dr. ir. B. van Loenen

Dit onderzoek is uitgevoerd in opdracht van:
Het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Auteurs:

dr. ir. F. Welle Donker

ir. R.C Braggaar

dr. ir. B. van Loenen

31 juli 2019

Kenniscentrum Open Data
Faculteit Bouwkunde, Afdeling Urbanism
Julianalaan 134, 2628 BL Delft
E-mail: opendata-bk@tudelft.nl
<http://www.opendata.bk.tudelft.nl/>



Dit werk is gelicenseerd onder een Creative Commons Naamsvermelding 4.0 Nederland. Bezoek <https://creativecommons.org/licenses/by/4.0/deed.nl> om een kopie te zien van de licentie of stuur een brief naar Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Inhoudsopgave

Samenvatting	i
1 Inleiding.....	1
1.1 Onderzoeksmethodologie	2
1.2 Leeswijzer	2
2 Theorie en onderzoeksmethodologie	3
2.1.1 Onderzoeksmethodologie Vraag 1 – wie zijn open data hergebruikers?	3
2.1.2 Onderzoeksmethodologie Vraag 2 – op welke manier(en) gebruiken hergebruikers open data	3
2.1.3 Onderzoeksmethodologie Vraag 3 - belemmeringen	4
2.1.4 Onderzoeksmethodologie Vraag 4 – behoeften hergebruikers.....	4
3 Vraag 1: Wie zijn open data hergebruikers	5
3.1 Gebruikte methode analyse logfile gegevens	5
3.2 Deelvraag 1a: categorieën open data hergebruikers.....	8
3.2.1 Analyse CBS open data	8
3.2.2 Analyse voor PDOK open data gebruikers.....	12
3.2.3 Analyse data.overheid.nl.....	18
3.2.4 Open data platforms in samenhang	23
3.2.5 Conclusie kwantitatief onderzoek hergebruikers.....	25
3.3 Deelvraag 1b: Hoe zijn de gebruikers te categoriseren?	26
3.3.1 Rollen van open data hergebruikers in open data waardeketen	26
3.3.2 Categorisering op basis van de logfile analyse	28
3.3.3 Categorisering op basis van rol in de open data waardeketen	29
3.4 Deelvraag 1.c. Welke dwarsverbanden zijn zichtbaar voor het type gebruiker, en de soort en hoeveelheid data die gebruikt wordt?.....	29
4 Vraag 2: Op welke manier wordt open data hergebruikt	30
4.1 Hergebruikers in de praktijk	30
4.2 Op welke manier(en) gebruiken zij momenteel open data?.....	31
4.2.1 Deelvraag 2a. Voor welke doeleinden wordt open data gebruikt?	31
4.2.2 Deelvraag 2b. Op welke manieren vinden gebruikers open data?	31
4.2.3 Deelvraag 2c. Via welke weg krijgen hergebruikers toegang tot open data?	32
5 Vraag 3: tegen welke belemmeringen lopen hergebruikers aan?	33
5.1 Potentiele barrières voor open data hergebruikers	33
5.2 Social media open data sentimentanalyse	34
5.2.1 Inleiding social media sentimentanalyses	34
5.2.2 Resultaten Twitter sentimentanalyse.....	35
5.2.3 Resultaten classificatie	38
5.2.4 Conclusies social media analyse	39

5.3	Waar lopen hergebruikers tegenaan bij het gebruiken van open data?	40
5.3.1	Deelvraag 3a. Hoe ervaren zij de kwaliteit van de aangeboden data?	40
5.3.2	Deelvraag 3b: Welke belemmeringen ervaren open data hergebruikers?	41
5.3.3	Deelvraag 3c: Welke verbetermogelijkheden zien hergebruikers?	44
6	Vraag 4: Wat zijn behoeften van de hergebruikers?.....	47
6.1	Deelvraag 4a: Welke behoefte hebben hergebruikers m.b.t. het aanbod van open data?	47
6.2	Deelvraag 4b: Welke behoefte hebben hergebruikers m.b.t. de manier waarop data worden aangeboden?.....	47
6.3	Deelvraag 4c. Welke behoefte hebben hergebruikers m.b.t. de vindbaarheid van data?	49
6.4	Overige behoeften: governance van open data	50
7	Conclusies en aanbevelingen open data hergebruikers.....	51
7.1	Conclusies.....	51
7.2	Beperkingen van dit onderzoek	52
7.3	Aanbevelingen	52
7.3.1	Aanbevelingen op het gebied van data platforms	52
7.3.2	Aanbevelingen op het gebied van open data.....	53
7.3.3	Van data-aanbod naar data-gedreven.....	54
	Dankwoord	56
	Lijst van afkortingen	57
	Literatuurlijst	58
	Bijlage 1 Overzicht van gebruikte datasets door grootgebruikers per open data portaal	60
	Bijlage 2 Lijst van geïnterviewden voor onderzoek.....	63
	Bijlage 3 Overzicht van IP register velden.....	64
	Bijlage 4 Methodologie Sentiment Analyse.....	65
B.4.1	Verzamelen van data.....	65
B.4.2	Data voorbereiding.....	66
B.4.3	Sentimentanalyse.....	66
B.4.4	Classificatie Tweets	67
B.4.5	Trainen van classificatiemodel	69
	Bijlage 5 Python script Tweets ophalen en sentiment analyse	70
	Bijlage 6 Python script sentiment analyse assessments per woord	71
	Bijlage 7 Rapportage online enquête	72

Samenvatting

Het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) investeert veel in het verbeteren van de ontsluiting van open data. Met de NL DIGITAAL Data Agenda Overheid¹ laat het ministerie van BZK ook zien dat er een duidelijke ambitie is om de burgers en ondernemers centraal te stellen en te komen tot gebruiker gedreven overheidsdata. Een van de actiepunten uit NL DIGITAAL is het bevorderen van hergebruik van open overheidsdata. Om daar stappen in te kunnen zetten, is het van belang om beter inzicht te krijgen in de vraag wie de hergebruikers van open data zijn, en wat hun behoeften zijn. Dat inzicht ontbreekt het Ministerie van BZK op dit moment. Er zijn voldoende aanwijzingen dat open data gebruikt worden, zeker als men kijkt naar het aantal bezoekers van open data portalen. Echter aantallen bezoekers zeggen op zich niet zoveel over het hergebruik zelf. Bekijken de bezoekers alleen naar een dataset, wordt de dataset gedownload dan wel via een Application Programming Interface (API) aangeroepen? Komen gebruikers uit het bedrijfsleven of uit de publieke sector en waarvoor gebruiken zij welke open data en via welke wegen? En wat zijn hun behoeften dan wel problemen waar zij tegen aan lopen?

In dit geven wij antwoord op de volgende vier vragen:

1. Wie zijn open data hergebruikers?
2. Op welke manier(en) gebruiken hergebruikers open data?
3. Waar lopen hergebruikers tegenaan bij het gebruiken van open data?
4. Wat zijn de behoeften van de hergebruikers?

Voor dit onderzoek is gebruik gemaakt van verschillende onderzoeksmethoden. Een kwantitatieve analyse is uitgevoerd op basis van de IP-Adressen in de logfile gegevens voor het open data portaal van CBS (statistische data), PDOK (geodata), en data.overheid.nl. Deze kwalitatieve analyse geeft antwoord op de vraag welke typen hergebruikers open data gebruiken en in welke categorieën zij kunnen worden onderverdeeld; welke datasets het meest worden gebruikt en welke kruisverbanden kunnen worden gelegd; op welke manier hergebruikers zich toegang tot de data verschaffen, en via welke weg (centraal dataplatform of individuele organisaties). Ook hebben wij technieken ingezet om inzicht te geven in welke websites melding gemaakt wordt van het gebruik van open data. Ten slotte hebben wij een kwantitatieve analyse gemaakt van Twitterberichten tussen 2010 en 2018 met betrekking tot open data.

Daarnaast is er een kwalitatieve analyse uitgevoerd op basis van een online enquête, een workshop met wetenschappelijke medewerkers, diepte-interviews met open data hergebruikers en een online prioritering van de grootste belemmeringen en behoeften door de geïnterviewde hergebruikers. Deze kwalitatieve analyse verifieert de kwantitatieve analyse en geeft inzicht in voor welke doelen open data worden gebruikt, eventuele knelpunten en behoeften van de verschillende categorieën hergebruiker. En vooral welke aanbevelingen zij hebben voor het Ministerie van BZK om het aanbod van open data te verbeteren.

Uit dit onderzoek blijkt dat open data op grote schaal en door diverse groepen worden gebruikt voor allerlei doeleinden en via allerlei wegen. Uit de kwantitatieve analyse van logfile gegevens blijkt dat vooral overheden, de private sector, universiteiten en hogescholen, en het bank- en verzekeringswezen grootgebruikers zijn van open data. Opvallend is ook dat Nederlandse open data ook door buitenlandse organisaties worden gebruikt. Verder blijkt dat het open data portaal van CBS vaker op afzonderlijke dagen worden gebruikt vergeleken met PDOK en data.overheid.nl. PDOK datasets worden het meest aangeroepen, vooral ook omdat veel van de datasets gebruikt worden in andere

¹ Zie Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). NL DIGITAAL. Data Agenda Overheid. Den Haag: 31. kst-26643-597. <https://zoek.officielebekendmakingen.nl/blg-876545.pdf>, (geraadpleegd 30 juni 2019)

diensten zoals het WOZ-waardeloket. Data.overheid.nl lijkt vooral gebruikt te worden als een catalogusdienst voor andere dataplatforms.

Uit de kwalitatieve analyse blijkt dat hoewel er verschillende categorieën hergebruikers te onderscheiden zijn, zoals bedrijven, not-for-profit organisaties en overheid, is het niet zo dat een bepaald type hergebruiker open data voor een specifiek doel gebruiken. Het bedrijfsleven beperkt zich niet tot hergebruik van open data voor alleen commerciële producten en diensten, en not-for-profit organisaties beperken zich niet tot alleen maatschappelijke diensten. Alle categorieën hergebruikers gebruiken open data om zowel producten en diensten voor derden te ontwikkelen als voor interne bedrijfsvoering. Hergebruikers maken gebruik van meerdere zoekstrategieën en van meerdere platforms om benodigde data te verkrijgen. Het feit dat er nog steeds gebruik wordt gemaakt van verzoeken voor datasets in kader van Wet openbaar bestuur bij de data-verstrekkers zelf is een indicatie dat nog niet alle data waar behoefte aan zijn, al beschikbaar zijn als open data.

Uit dit onderzoek blijkt dat hergebruikers vooral belemmeringen ervaren met open data van lokale overheden. Het aanbod is versnipperd, er is geen eenduidige standaardisatie met betrekking tot de datastructuur en bestandsformaten, en niet alle lokale overheden maken gelijksoortige datasets beschikbaar als open data waardoor er geen landsdekking of regiodekking bereikt kan worden.

Uit dit onderzoek blijkt ook dat hergebruikers vooral behoefte hebben aan meer samenwerking op het gebied van open data governance. Hergebruikers willen graag betrokken worden bij de governance van open data proces, maar geven ook aan dat – vooral voor ondernemers – geen tijd hebben om deel te nemen aan formele bijeenkomsten. Liever hebben hergebruikers dat de overheid naar hun komt om een beter beeld te krijgen van hoe open data worden ingezet en hoe open data governance verbeterd zou kunnen worden. Op de korte termijn hebben hergebruikers de meeste behoefte aan betere feedback faciliteiten en een tijdige reactie op de geleverde feedback, meer regie op het beschikbaar stellen van open data van lokale overheden en die data land dekkend te maken, en gebruik van (inter)nationaal herkende standaarden voor API's.

Uit dit onderzoek blijkt ook dat open data nu een vaste waarde zijn voor veel hergebruikers. Er zijn nu bedrijven die zonder open data niet zouden kunnen bestaan met hun businessmodel geheel afhankelijk van open data. Ook ontdekken steeds meer hergebruikers uit andere sectoren de toegevoegde waarde van open data. Waren er een aantal jaren geleden nog vooral hergebruikers uit de geo-sector, nu ziet men breder hergebruik in andere sectoren. Uit dit onderzoek blijkt ook dat hergebruikers open data gebruiken om producten en diensten te ontwikkelen die een bijdrage kunnen leveren aan maatschappelijke vraagstukken.

De uitkomsten van dit onderzoek zullen bijdragen aan de verdere ontwikkeling van een gebruiker-gedreven opendatabeleid, dat kan aansluiten bij de ambities van de digitale overheid. Voor de overheid om datagedreven te werken, zijn er nog een aantal stappen nodig. Het is belangrijk dat het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties de regie van de governance van data oppakt, niet alleen op nationaal niveau maar vooral ook op lokaal niveau. Het beschikbaar stellen van open data is geen doel meer op zichzelf maar een onderdeel van een datastrategie om de transitie naar datagedreven organisaties te verwezenlijken. Er moet een datastrategie belegd worden zowel op macroniveau (nationaal) als op mesoniveau (regionaal) en microniveau (organisatie), waarbij het voor eenieder duidelijk waar de verantwoordelijkheden liggen. Datakwaliteit wordt ook steeds belangrijker naarmate meer geautomatiseerde besluitvorming gebaseerd is op algoritmes. En dat vergt dat overheden transparant zijn over de bron en de inhoud van de data. Maar van beleid naar uitvoering gaan, vergt een cultuurverandering binnen organisaties en een lange adem. Daarom is het essentieel dat er een duidelijke lange termijn visie aanwezig is en dat de geplande fases ook werkelijkheid worden.

1 Inleiding

Het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) investeert veel in het verbeteren van de ontsluiting van open data. Met de NL DIGITAAL Data Agenda Overheid² laat het ministerie van BZK ook zien dat er een duidelijke ambitie is om de burgers en ondernemers centraal te stellen en te komen tot gebruiker-gedreven overheidsdata. Een van de actiepunten uit NL DIGITAAL is het bevorderen van hergebruik van open overheidsdata. Om daar stappen in te kunnen zetten, is het van belang om beter inzicht te krijgen in de vraag wie de hergebruikers van open data zijn, en wat hun behoeften zijn. Dit inzicht ontbeert het Ministerie op dit moment. Er zijn voldoende aanwijzingen dat open data gebruikt worden, zeker als men kijkt naar het aantal bezoekers van de drie grootste dataportalen: data.overheid.nl, Publieke Dienstverlening op de Kaart (PDOK) en het Open Data portaal van het Centraal Bureau voor de Statistiek (CBS). Echter aantallen bezoekers zeggen op zich niet zoveel over het hergebruik zelf. Kijken mensen alleen naar een dataset, bijvoorbeeld via een catalogus of wordt de dataset daadwerkelijk gedownload dan wel via een Application Programming Interface (API) aangeroepen? Welke soort gebruikers gebruiken welke open data en via welk platform? En wat zijn hun voorkeuren of problemen waar zij tegen aan lopen?

In dit onderzoek staan de volgende vier vragen centraal:

1. Wie zijn open data hergebruikers?

Deelvragen zijn:

- Welke typen hergebruikers kunnen worden onderscheiden?
- Hoe zijn zij te categoriseren?
- Welke dwarsverbanden zijn zichtbaar voor het type gebruikers, en de soort en hoeveelheid data die gebruikt wordt?

2. Op welke manier(en) gebruiken hergebruikers open data?

Deelvragen zijn:

- Voor welke doeleinden wordt open data gebruikt?
- Op welke manieren vinden gebruikers open data?
- Via welke weg benutten hergebruikers open data?

3. Waar lopen hergebruikers tegenaan bij het gebruiken van open data?

Deelvragen zijn:

- Hoe ervaren zij de kwaliteit van de aangeboden data?
- Welke belemmeringen ervaren zij?
- Welke verbetermogelijkheden zien hergebruikers?

4. Wat zijn de behoeften van de hergebruikers?

Deelvragen zijn:

- Welke behoefte hebben hergebruikers m.b.t. het aanbod van open data?
- Welke behoefte hebben hergebruikers m.b.t. de manier waarop data wordt aangeboden?
- Welke behoefte hebben hergebruikers m.b.t. de vindbaarheid van data?

De uitkomsten aan dit onderzoek zullen bijdragen aan de verdere ontwikkeling van een gebruiker gedreven opendatabeleid, dat kan aansluiten bij de ambities van de digitale overheid.³

² Zie Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). NL DIGITAAL. Data Agenda Overheid. Den Haag: 31. kst-26643-597. <https://zoek.officielebekendmakingen.nl/blg-876545.pdf>, (geraadpleegd 30 juni 2019)

³ Zie Knops (2019). Aanbieding NL DIGITAAL: Data Agenda Overheid, 15 maart, kst-26643-597, beschikbaar via <https://zoek.officielebekendmakingen.nl/kst-26643-597.html> (geraadpleegd 30 juni 2019).

1.1 Onderzoeksmethodologie

In dit onderzoek is gebruik gemaakt van verschillende onderzoeksmethoden. Een kwantitatieve analyse is uitgevoerd voor het open data portaal van CBS, PDOK, en data.overheid.nl op basis van:

- IP-adressen en web-statistieken naar de typen hergebruikers en in welke categorieën zij onderverdeeld kunnen worden, welke data worden gebruikt, op welke manier zij zich toegang tot de data verschaffen, en via welke weg (data.overheid.nl, PDOK, CBS, of via individuele organisaties).
- De inzet van web crawling technieken zoals search API's om inzicht te geven in welke websites melding gemaakt wordt van het gebruik van open data.
- Een kwantitatieve analyse van Twitterberichten met betrekking tot open data.

Verder is er een analyse uitgevoerd op basis van een online enquête, diepte-interviews onder open data hergebruikers en een online prioritering van de grootste belemmeringen en behoeften door de geïnterviewde hergebruikers. Deze kwalitatieve analyse verifieert de kwantitatieve analyse en geeft inzicht in de doelen waarvoor open data wordt gebruikt, eventuele knelpunten en behoeften van de verschillende categorieën hergebruiker. En vooral welke aanbevelingen zij hebben voor het Ministerie van Binnenlandse Zaken en Koninkrijksrelatie om het aanbod van open data te verbeteren. In hoofdstuk 2 wordt per onderzoeksvraag ingegaan op de toegepaste onderzoeksmethode.

1.2 Leeswijzer

In hoofdstuk 2 beschrijven we de onderliggende theorie en de gehanteerde onderzoeksmethodologieën om de onderzoeksvragen te beantwoorden. In Hoofdstuk 3 geven wij antwoord op vraag 1: welke categorieën open data hergebruikers kunnen we onderscheiden. Wij doen dat aan de hand van onze kwantitatieve analyse van logfile gegevens voor statistische data beschikbaar via CBS Open Dataportaal, geodata beschikbaar via PDOK en open data beschikbaar via data.overheid.nl. Uit deze logfile analyse kunnen wij de typen hergebruikers vaststellen, en welke datasets het meest populair zijn bij welke typen hergebruikers. Op basis van deze kwantitatieve analyse maken wij een categorisatie van de typen hergebruikers op basis van de rollen die zij kunnen innemen in de open data waarde keten. Vraag 2, op welke manier worden open data herbruikt en waarvoor, wordt in Hoofdstuk 4 behandeld. Wij presenteren de uitkomsten van onze kwalitatieve analyse gebaseerd op de uitkomsten van een enquête uitgezet onder hergebruikers, een workshop gehouden met wetenschappelijke hergebruikers, en een ronde van diepte-interviews gehouden met hergebruikers. De analyse zal antwoord geven op de vraag voor welk doeleinde hergebruikers open data gebruiken, en via welke weg. In Hoofdstuk 5 gaan wij in op vraag 3: welke belemmeringen lopen open data hergebruikers tegenaan. Wij presenteren een sentimentsanalyse van Twitter over open data. Vervolgens laten we de resultaten van onze kwalitatieve analyse zien waarin de belemmeringen van open data hergebruikers worden beschreven, en de aanbevelingen van open data hergebruikers. Vraag 4: Wat zijn de behoeften van open data hergebruikers wordt in Hoofdstuk 6 beantwoord aan de hand van onze kwalitatieve analyse. Hoofdstuk 7 sluit af met onze conclusies en aanbevelingen. In Bijlage 1 geven wij een overzicht van de dwarsverbanden tussen de typen hergebruikers en welke datasets zij het meest gebruiken.

2 Theorie en onderzoeksmethodologie

In dit hoofdstuk zullen wij per vraag beschrijven welke onderzoeksmethodes zijn gebruikt om de vragen te beantwoorden.

2.1.1 Onderzoeksmethodologie Vraag 1 – wie zijn open data hergebruikers?

Voor het beantwoorden van vraag 1 (wie zijn de hergebruikers en hoe kunnen zij gecategoriseerd worden?) is een combinatie van kwantitatieve en kwalitatieve onderzoeksmethodes ingezet.

Het kwantitatieve element maakt op basis van IP-adressen een onderverdeling naar type gebruikers en aantallen hits. Met deze methode is een inschatting gemaakt van het type gebruiker van de datasets (IP-adres behorend bij een bedrijf, overheid, onderwijsinstelling en overig; binnenland en buitenland) en een aantal van deze gebruikers te identificeren (o.a. via bedrijfsnaam) als grootgebruikers van open data. Deze analyse hebben wij uitgevoerd voor open statistische data beschikbaar via het open data portaal van CBS, voor open geografische data via PDOK, en open data beschikbaar via data.overheid.nl.

De analyses geven een eerste overzicht van de typen hergebruikers, welke data, welke hoeveelheden en of dit eenmalige of meer frequente bezoekers zijn. Deze typen hergebruikers zijn gecategoriseerd en daarmee het uitgangspunt voor het beantwoorden van deelvraag 1.b (hoe zijn de typen hergebruikers te categoriseren) en deelvraag 1.c (welke dwarsverbanden zijn er zichtbaar voor het type hergebruiker en de soort en hoeveelheid data die gebruikt wordt).

Daarnaast is er door middel van een literatuurstudie een analyse van de verschillende typeringen van hergebruikers gemaakt.

Op basis van de uitgevoerde analyses is er een kwalitatieve verdiepingsslag uitgevoerd via interviews met verschillende typen hergebruikers. De uitkomsten van de interviews zijn gebruikt om de antwoorden op deelvraag 1.b. en deelvraag 1.c. te verifiëren, en als input voor het beantwoorden van deelvragen 2, 3 en 4.

2.1.2 Onderzoeksmethodologie Vraag 2 – op welke manier(en) gebruiken hergebruikers open data

Om de vraag te beantwoorden op welke manier open data momenteel wordt gevonden, via welke weg en voor welke doeleinden open data worden gebruikt, hebben wij verschillende analyses gebruikt. Ten eerste hebben wij tussen 23 november 2018 en 31 januari 2019 een online enquête via social media (Twitter en LinkedIn) uitgezet om Vraag 2 - op welke manier en waarvoor hergebruikers open data gebruiken - te kunnen beantwoorden. Hoewel online vragenlijsten hun beperkingen hebben, is het een snelle manier om zo veel mogelijk mensen te bereiken.⁴ Tevens hebben we tijdens een workshop, gehouden op de Technische Universiteit Delft op 8 maart 2019, onder internationale wetenschappelijke gebruikers van open data gepeild welke open datasets zij gebruiken en via welke weg en diensten zij open data vinden. Ten slotte hebben wij een analyse gevoerd op basis van het inzetten van search API's. Dergelijke technieken gaven een eerste inzicht in welke websites meldingen maken van het gebruik van open data.

⁴ <https://surveyanyplace.com/nl/de-9-grootste-voordelen-van-enquetes-en-de-nadelen/>

Deze methode kan slechts een indicatie geven aangezien niet alle websites meldingen maken van het hergebruik van open data of van de licenties van de onderliggende data. Wel gaf deze analyse een beeld van nieuwsberichten waarin nieuwe toepassingen die ontwikkeld zijn op basis van open data werden gemeld. Deze analyse van nieuwe toepassingen vormde de basis van het opstellen van een longlist voor interviews van open data hergebruikers. Van de longlist zijn uiteindelijk een twintigtal hergebruikers van verschillende achtergronden (commerciële organisaties en organisaties zonder winstoogmerk, van kleine ondernemingen tot grote bedrijven, en van de overheid en de wetenschap) geselecteerd voor een shortlist voor diepte-interviews. Bij deze selectie is er op gelet dat er een evenredige spreiding was tussen typen hergebruikers en de grootte van de onderneming. Er zijn uiteindelijk 12 interviews in juni 2019 via Skype of in persoon uitgevoerd door middel van gestructureerde open vragen. Deze interviews hebben ons een kwalitatief beeld gegeven van voor welke doeleinden hergebruikers open data gebruiken, op welke manier de open data wordt gevonden, via welke weg (via welk (overheids)portaal, via individuele organisaties), via welke opties (download, API's, voorkeurformaten).

2.1.3 Onderzoeksmethodologie Vraag 3 - belemmeringen

Voor het beantwoorden van de vraag waar hergebruikers tegenaan lopen bij het gebruiken van open data, hebben wij een kwantitatieve analyse uitgevoerd van Twitter berichten tussen 2010 en 2018 met betrekking tot open data. De methode voor deze analyse bouwt op de methode ontwikkeld door een van de onderzoekers voor een eerder onderzoek.⁵ Deze analyse scheidt een beeld van de problemen waar open data gebruikers in het verleden tegen aan zijn gelopen met betrekking tot datakwaliteit, belemmeringen en eventuele suggesties hoe het beter zou kunnen. In de online enquête gehouden tussen 23 november 2018 en 31 januari 2019 is ook gevraagd naar belemmeringen van hergebruikers. Tevens hebben we tijdens de workshop op de Technische Universiteit Delft op 8 maart 2019 de belemmeringen waar wetenschappelijke gebruikers van open data tegenaan lopen gepeild. Daarnaast zijn wij tijdens de interviews in juni dieper ingegaan op deze belemmeringen. Ten slotte hebben wij in juli 2019 de meest genoemde belemmeringen gevalideerd met de geïnterviewde hergebruikers met een online enquête. In deze enquête werden de respondenten gevraagd om van de tien meest genoemde belemmeringen aan te geven welke belemmering de hoogste prioriteit voor hen hadden.

2.1.4 Onderzoeksmethodologie Vraag 4 – behoeften hergebruikers

De vraag wat de behoeften zijn van open data hergebruikers is beantwoord aan de hand van de uitkomsten van de analyse van de online enquête gehouden tussen 23 november 2018 en 31 januari 2019, peilingen tijdens de workshop van 8 maart, en diepte-interviews met hergebruikers in juni 2019. Bij het beantwoorden van deze vraag hebben wij ons niet beperkt tot de behoeften op het operationele niveau van open data (aanbod van open data, wijze van aanbieden van open data, vindbaarheid van data), maar ook op de governance van het open data ecosysteem. Hoe willen hergebruikers betrokken worden bij het opendatabeleid, welke behoeften hebben zij met betrekking tot bijvoorbeeld het bepalen van de datakwaliteit van open data, en hoe kan de communicatie met data-aanbieders gefaciliteerd worden. En vooral, welke aanbevelingen hergebruikers hebben voor de overheid. Ten slotte hebben wij in juli 2019 de meest genoemde behoeften gevalideerd door middel van een online enquête met de geïnterviewde hergebruikers. In deze enquête werden de respondenten gevraagd om van de tien meest genoemde behoeften aan te geven welke behoefte de hoogste prioriteit voor hen hadden.

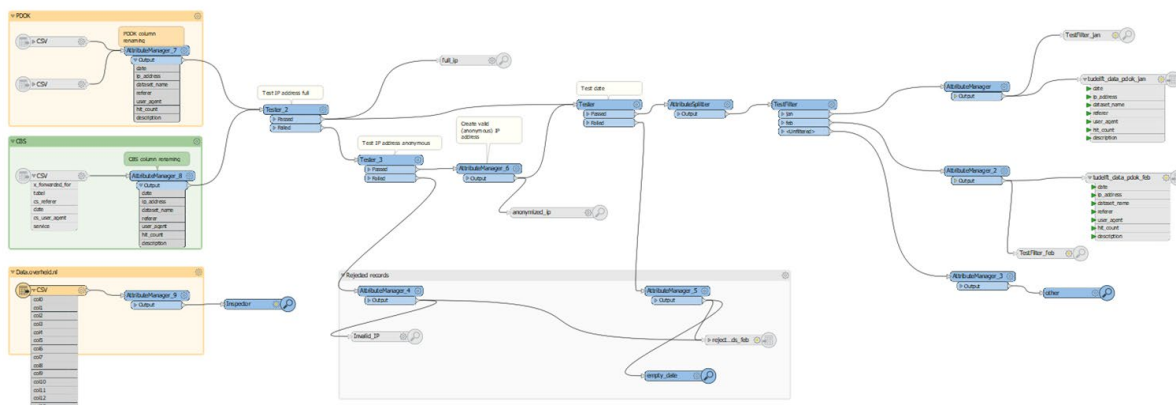
⁵ Zie Braggaa, R. (2018). Wi-Fi network-based indoor localisation: The case of the TU Delft campus. Fac. Architecture & the Built Environment. Delft, Delft University of Technology. MSc. in Geomatics for the Built Environment: 92. <https://repository.tudelft.nl/islandora/object/uuid:bb3009cb-605e-4b80-9cf6-85f7dbc9ace8?collection=education>.

3 Vraag 1: Wie zijn open data hergebruikers

Om deze vraag te beantwoorden worden er twee methodes ingezet. De eerste methode betreft een analyse van logfile bestanden van open data portalen. In deze logfile bestanden staan gegevens over welke gebruiker (via een IP-nummer) via welke ingang (referer) op welke datum welk deel van het portaal bezoekt. Dat kan zijn om een bepaalde dataset te zoeken of te bekijken, of door te klikken naar andere documenten die beschikbaar zijn via dat portaal. De logfile bestanden geven geen inzicht of een bezoeker een dataset downloadt of alleen maar bekijkt en daarna doorklikt. Daarom geven web-statistieken op zich alleen een indicatie van hoe vaak een website wordt bezocht maar niet met welk doel. Logfiles kunnen wel inzicht geven in wie de bezoekers zijn van een opendataportaal, dus wie de hergebruikers van open data zijn en hoe zij gecategoriseerd kunnen worden. Voor dit onderzoek is een kwantitatieve analyse uitgevoerd voor drie verschillende opendataportalen, te weten CBS Open Data Statline van Centraal Bureau voor de Statistiek PDOK voor actuele geo-informatie, en data.overheid.nl, het dataregister van de Nederlandse Overheid. De methode die voor die drie analyses is gebruikt is voor het merendeel identiek. Daarom wordt de methode hieronder voor de drie databronnen gezamenlijk in Sectie 3.1 besproken. De tweede methode betreft de kwalitatieve analyse uitgevoerd op basis van de online enquête, de workshop, en de diepte-interviews zoals beschreven in Hoofdstuk 2.

3.1 Gebruikte methode analyse logfile gegevens

Uitgangspunt voor de kwantitatieve analyse zijn in eerste instantie de IP-adressen van de gebruikers. De IP-adressen worden als web-statistieken bijgehouden op de servers van de dataportalen. Voor de voorbereiding was de eerste stap het structureren en opschonen van de bronbestanden. Om de dataset geschikt te maken voor analyse moesten de mislukte verzoeken, de ongeldige IP-adressen, de records waar een datum ontbrak, en de aanroepen via bots uit de dataset worden gehaald. Vervolgens is er structuur in de datasets aangebracht door de kolomnamen en de structuur uniform te maken. Ten slotte zijn de opgeschoonde bronbestanden naar CSV-formaat omgezet voor verdere analyse. Figuur 1 geeft het opschoningsproces grafisch weer.

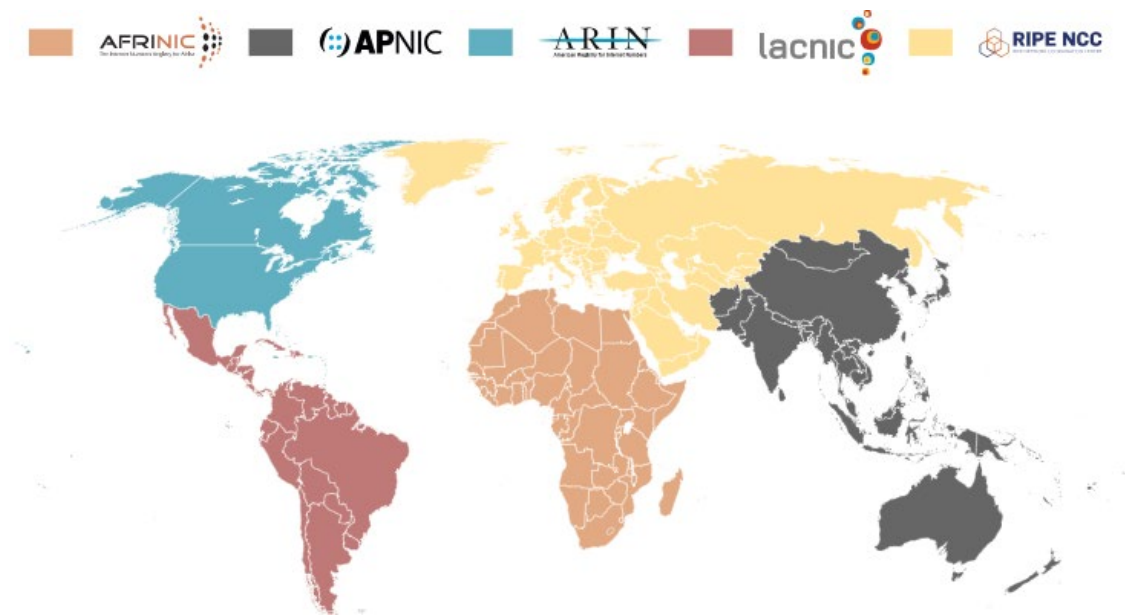


Figuur 1: van CSV bronbestand naar opgeschoonde CSV bestanden

De opgeschoonde datasets zijn geconverteerd naar een Bigquery database zodat de analyse van IP-nummers kon worden uitgevoerd. Door middel van registratie (Nederlands of buitenlands) wordt voor elk (blok van) IP-nummers een aantal vaste gegevens bijgehouden, waaronder informatie over

het land, een netnaam en een beschrijving. Deze informatie wordt bijgehouden in verschillende publieke databases, zie Figuur 2:

- RIPE (hier vallen de Nederlandse IP-blok registraties onder)
- APNIC
- AFRINIC
- LACNIC
- ARIN
- ARIN-RR
- RADb
- NTTCOM
- JPIRR



Figuur 2: de regionale gebieden per internetregister Bron:
<https://www.arin.net/knowledge/rirs.html> (geraadpleegd 4 februari 2019)

Met behulp van een zoekterm naar de registers kan de informatie worden geraadpleegd. Een query kan worden opgebouwd met behulp van flags welke bijvoorbeeld aangeven welke databases geraadpleegd moeten worden. Onderstaand een voorbeeld van een query die alle databases raadpleegt voor het geanonimiseerde IP-adres 49.52.97.0.⁶

```
https://rest.db.ripe.net/search.json?query-string=49.52.97.0&flags=no-referenced&flags=no-irt&flags=no-filtering&source=GRS
```

Per platform (CBS, PDOK en data.overheid.nl) is de ruwe informatie van de IP-adressen vervolgens verwerkt in de Bigquery database. Verdere analyses zijn vervolgens via SQL-queries gerealiseerd op de database. Relevante query's worden in dit rapport weergegeven in pseudocode, zie Figuur 3. Per platform is er gewerkt met een top 10 van datasets voor een nadere analyse gezien het grote aantal datasets op de platformen. De top 10 is vastgesteld aan de hand van de gebruiksstatistieken die op de platformen worden bijgehouden.

⁶ Het IP-adres in dit voorbeeld is geanonimiseerd. Een IPv4-adres bestaat uit vier groepen getallen, elk tussen 0 en 255, gescheiden door een punt. De laatste groep is in dit voorbeeld vervangen met een 0.

```

1 filter type: contains col: platform_name contains:
  'CBS' action: Keep
2 groupby group: dataset_name value: SUM(hit_count)
  type: agg
3 filterrownumber type: top orderType: custom rows:
  10 start: 1 order: sum_hit_count action: Keep

```

Figuur 3: pseudocode vanuit SQL

Verder is er gezocht op uit welke landen de gebruikers van de open data platforms komen, zie Figuur 4 voor de gebruikte pseudocode.

```

1 join table: current_dataset with: tudelft_ip_data -
2 col: dataset_name = current_dataset.dataset_name,
  date = current_dataset.date, ip_address =
  current_dataset.ip_address, referer =
  current_dataset.referer, user_agent =
  current_dataset.user_agent, hit_count =
  current_dataset.hit_count, description =
  current_dataset.description, country =
  current_dataset.country, country_code =
  current_dataset.country_code, continent =
  current_dataset.continent, ip_type =
  current_dataset.ip_type, ip_name =
  current_dataset.ip_name, isp = current_dataset.isp,
  org = current_dataset.org, business_name =
  current_dataset.business_name, business_website =
  current_dataset.business_website, platform_name =
  current_dataset.platform_name, user_classification
  = current_dataset.user_classification row:
  current_dataset.dataset_name == tudelft_ip_data -
  2.dataset_name type: inner
-----
2 pivot col: country group: dataset_name value:
  SUM(hit_count) limit: 50

```

Figuur 4: pseudocode Nationaliteiten gebruikers top 10 datasets

Per platform is er vervolgens van de top 10 datasets gekeken naar een onderscheid in gebruikersgroepen. Hierbij is gebruik gemaakt van een SQL-script waarbij de op naam van de organisatie aan wie het IP adres is toegekend, gezocht werd op (delen van) die naam zelf. Daarna kunnen die namen in categorieën worden ingedeeld. Het deel in het script voor 'overheid' wordt in Figuur 5 weergegeven:

```

case
when concat_name ilike '%gemeen%'
or concat_name ilike '%provin%'
--or concat_name ilike '%rijk%'
or concat_name ilike '%minist%'
or concat_name ilike '%waterschap%'
or concat_name ilike '%hoogheemraadschap%'
or concat_name ilike '%politie%'
or concat_name ilike '%veiligheid%'
or concat_name ilike '%rijksdienst%'
or concat_name ilike '%rijksbureau%'
or concat_name ilike '%rijksinstituut%'
      or netname ilike '%rijksinternet%'
or concat_name ilike '%waterstaat%'
or concat_name ilike '%agentschap%'
or concat_name ilike '%kadaster%'
or concat_name ilike '%government%'
or concat_name ilike '%verwaltung%'
or concat_name ilike '%Centraal Planbureau%'
or concat_name ilike '%Belastingdienst%'
then 'overheid'
when concat_name ilike '%europea%'
then 'overheid_europa'

```

Figuur 5: Deel van SQL-script om categorie 'overheid' te onderscheiden

3.2 Deelvraag 1a: categorieën open data hergebruikers

3.2.1 Analyse CBS open data

CBS hanteert een open data beleid en heeft sinds 18 juli 2014 zijn volledige databank met 3.800 dataverzamelingen als open data beschikbaar gesteld. De reden hiervoor was om statistische informatie beter bereikbaar en makkelijker herbruikbaar te maken (zonder beperkingen die verbonden zijn aan de rechtstreekse toegang tot registerdata), waardoor niet alleen wetenschappers maar ook het bedrijfsleven tijd en geld besparen en makkelijker CBS-data kunnen gebruiken om nieuwe toepassingen te maken.⁷ CBS investeert veel in het beschikbaar stellen van open data om hergebruik te stimuleren. In 2018 is Statline, de online databank van CBS, vernieuwd met een nieuw uiterlijk en meer mogelijkheden, waaronder een nieuwe zoekmachine. CBS open data worden door steeds meer gebruikers gevonden. Zo waren er in mei 2019 meer dan een half miljoen unieke gebruikers.⁸ Open data kunnen via een door CBS ontwikkelde API worden aangeroepen. CBS ontwikkelt samen met de statistiebureaus van Vlaanderen en Duitsland een grensoverschrijdend opendataportaal.⁹ CBS organiseert regelmatig gebruikersbijeenkomsten om ervaringen uit te wisselen en te peilen of het open data portaal voldoet aan de wensen van de gebruikers.

⁷ <https://www.cbs.nl/nl-nl/nieuws/2014/29/cbs-lanceert-open-data-portal-volledige-databank-nu-beschikbaar> (geraadpleegd 18 januari 2019)

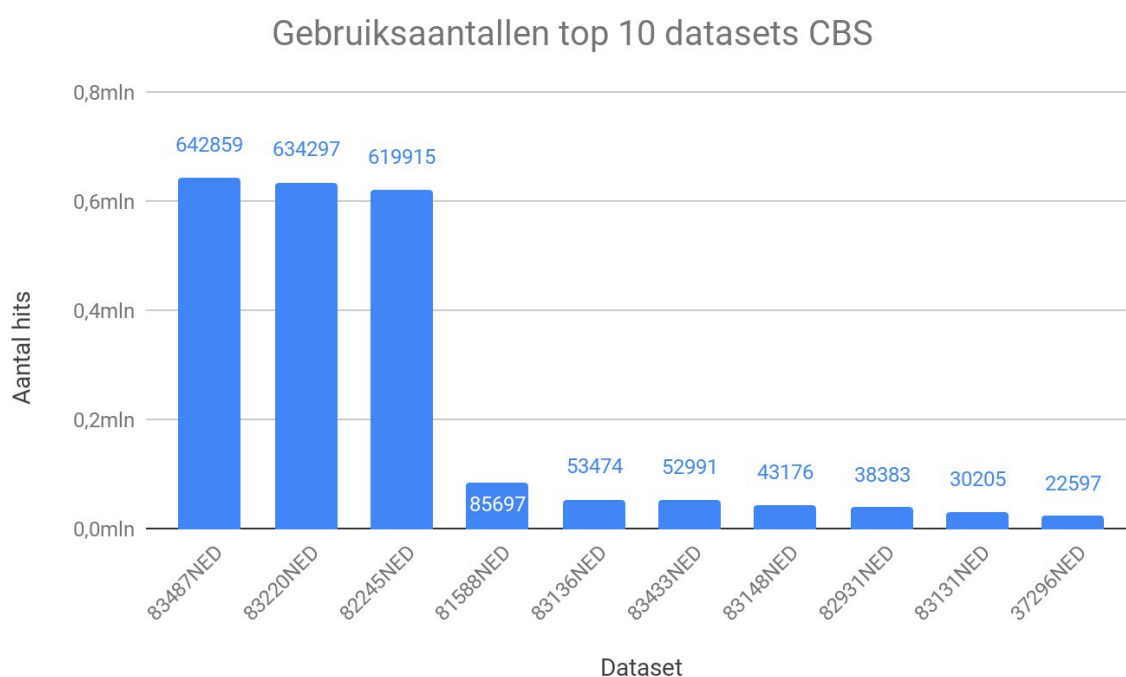
⁸ <https://ibestuur.nl/praktijk/cbs-plukt-vruchten-van-open-data-apis> (geraadpleegd 17 juli 2019)

⁹ <https://www.cbs.nl/nl-nl/corporate/2019/29/cbs-ontwikkelt-portal-met-grensoverschrijdende-data> (geraadpleegd 18 juli 2019)

Voor de analyse van CBS log bestanden zijn de logfile gegevens van januari en februari 2019 geanalyseerd. In de analyse zijn IP-adressen eerst naar een geldig 32-bits formaat en indien nodig geconverteerd. IP-adressen in een ander formaat leveren bij voorbaat foutmeldingen op bij het raadplegen van de registers. De analyse toonde aan dat onder andere de 'Kerncijfer' datasets en de postcode-4 dataset over bevolking en huishoudens populaire datasets zijn. Tabel 1 geeft aan welke datasets bij welke codes horen en Figuur 6 geeft een overzicht van de top 10 meest gebruikte datasets in de logfiles van het CBS:

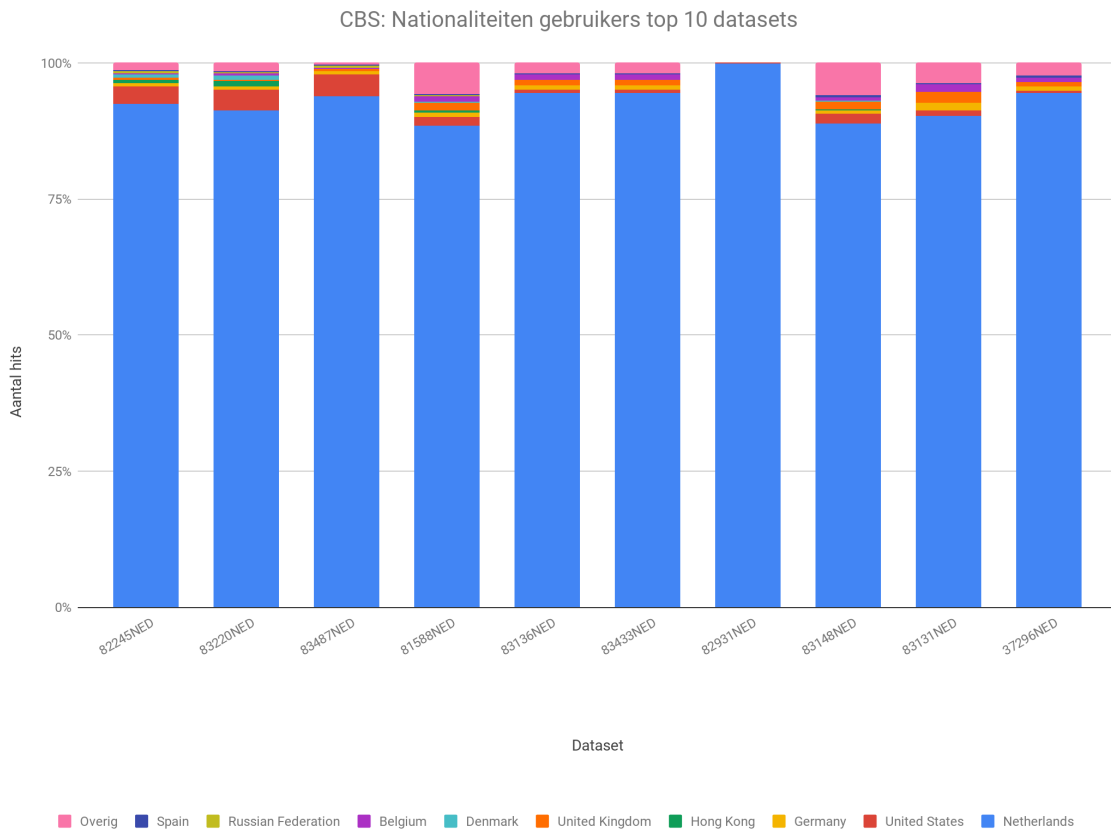
Tabel 1: CBS-codes en bijbehorende namen van datasets

83487NED	Kerncijfers wijken en buurten 2016
83220NED	Kerncijfers wijken en buurten 2015
82245NED	Bevolking en huishoudens; viercijferige postcode
81588NED	Bedrijven; bedrijfsgrootte en rechtsvorm
83136NED	Consumentenprijzen; basisjaren vanaf 1969
83433NED	Consumentenprijzen; werknemers laag, alle basisjaren 1969-1995
83148NED	Bedrijven; oprichtingen, bedrijfsgrootte, rechtsvorm, bedrijfstak
82931NED	Kerncijfers wijken en buurten 2014
83131NED	Consumentenprijzen; prijsindex 2015=100
37296NED	Bevolking; kerncijfers



Figuur 6: gebruiksaantallen top 10 meest gebruikte CBS datasets (januari en februari 2019)

Vervolgens hebben we in Figuur 7 een verdeling gemaakt van het gebruik van de top 10 datasets per land, waarbij wij de tien meest voorkomende landen hebben getoond.

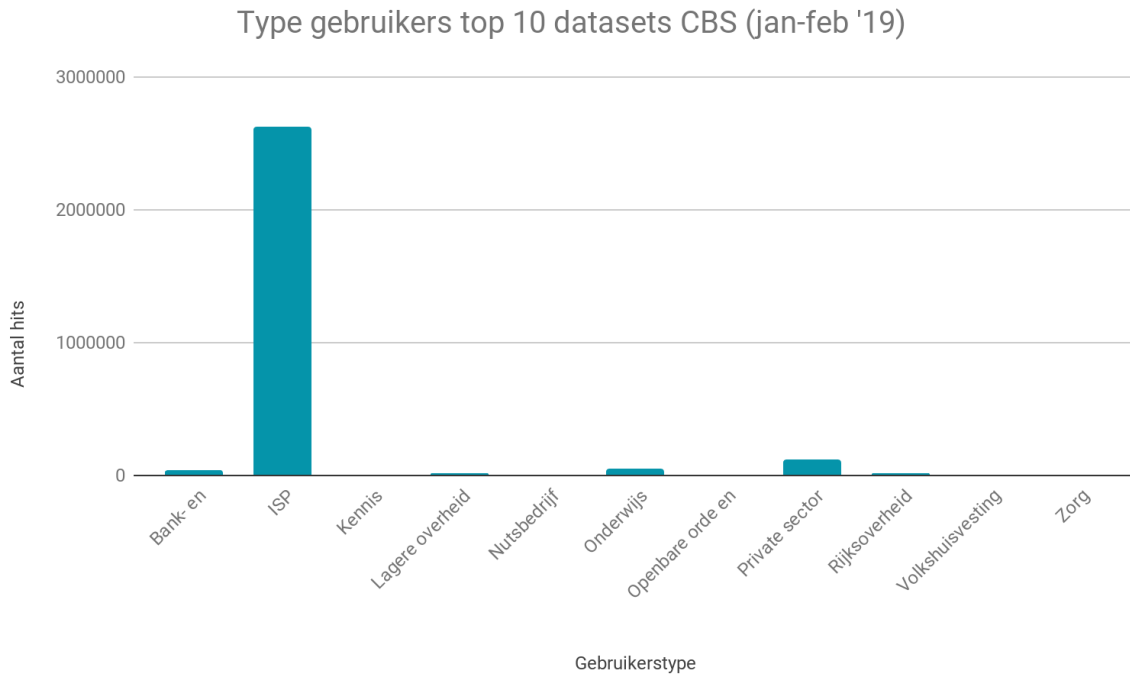


Figuur 7: verdeling gebruik top 10 CBS datasets per land (januari en februari 2019)

Het merendeel van de gebruikers van de top 10 datasets komt uit Nederland (blauw), gevolgd door 'Overig' (roze) en daarna de Verenigde Staten (donkerroze). Daarna zijn Duitsland (geel) en Hong Kong (groen) de grootste gebruikersgroepen.

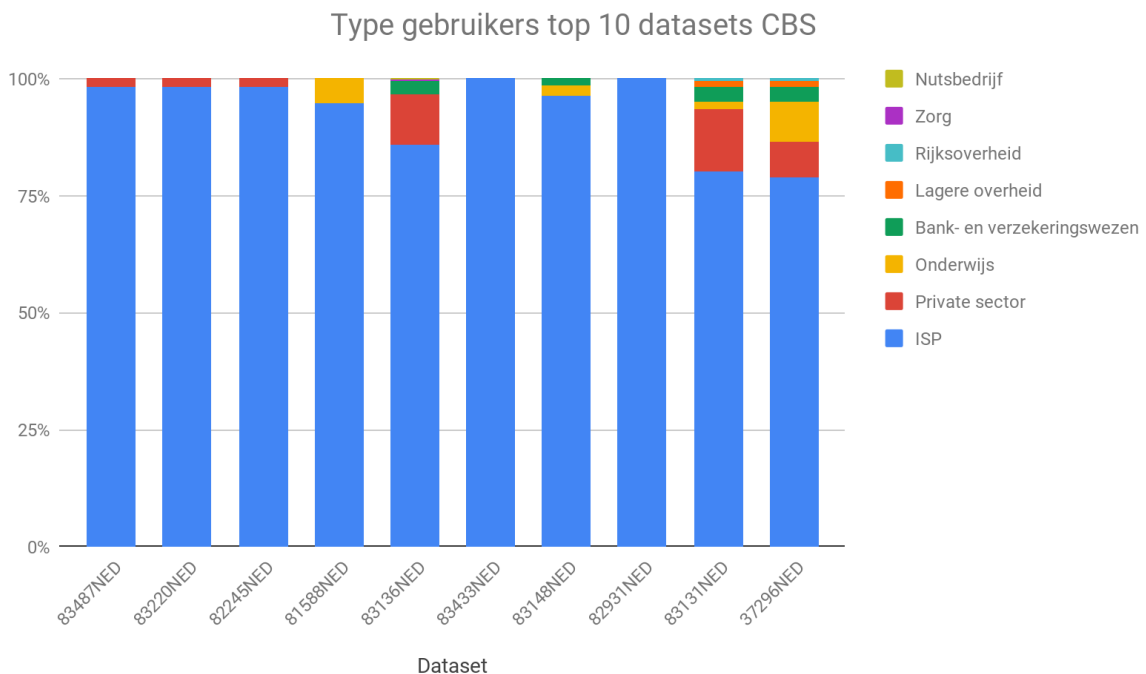
Van een circa 24% van de gebruikers kan de herkomst niet worden vastgesteld. Oorzaak hiervan is mogelijk dat deze IP-adressen niet zijn opgenomen in de online registers en daarom niet teruggeleid kunnen worden naar een land. Deze gebruikers zijn in de groep 'Overig' geplaatst, samen met de landen die vanaf positie 11 in de lijst voorkomen.

In Figuur 8 is een onderverdeling gemaakt naar het type gebruiker van de top 10 CBS datasets). Het merendeel van de gebruikers bevindt zich in de categorie 'Internet Service Provider' (ISP). Vaak zijn dit burgers of gebruikers zonder individuele IP-adressen, dat wil zeggen, het IP-nummer valt binnen een nummerblok van een ISP zoals KPN of Ziggo. In deze categorie vallen ook organisaties waarvan de organisatiernaam niet aan het IP-adres is gelinkt, bijvoorbeeld kleinere ondernemingen waarvan de bedrijfsnaam niet uit het IP-adres is te achterhalen. Tot de ISP categorie behoren ook de vastelijns of mobiele verbindingen naar de provider toe. Mobiele telefoons die gebruik maken van apps op basis van CBS open data zitten dus ook in deze categorie. In de categorie 'Private Sector' vallen alle bedrijven die niet in één van de apart genoemde categorieën vallen, zoals bank- en verzekeringswezen, maar waar wel een bedrijfsnaam uit het IP-adres is te achterhalen. Andere categorieën van hergebruikers die te onderscheiden zijn, zijn de lagere overheden en de onderwijssector.



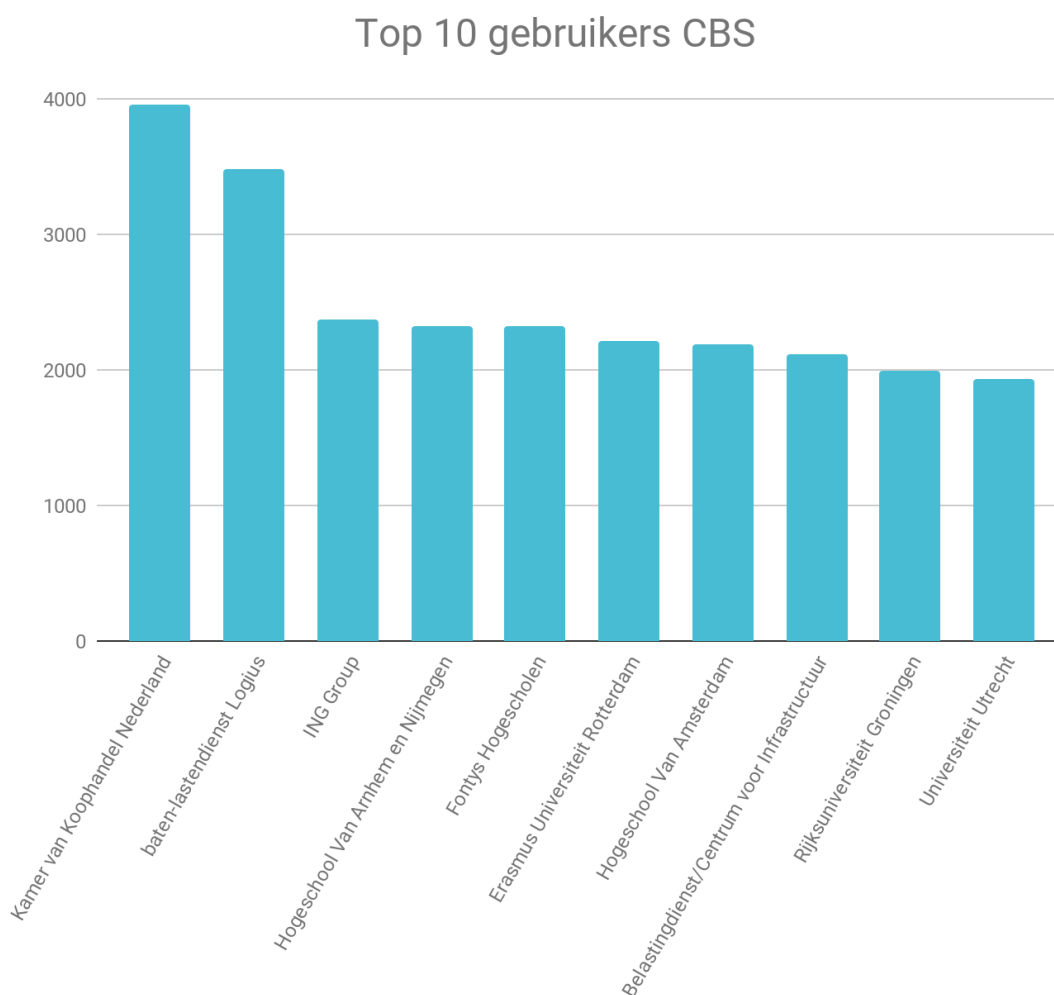
Figuur 8: verdeling gebruik van CBS data over gebruikersgroepen (januari en februari 2019)

In Figuur 9 wordt de verdeling van type gebruikers van de top 10 datasets CBS weergegeven. Hierbij valt op dat de datasets ‘Consumentenprijzen; prijsindex 2015=100’, ‘Bevolking; kerncijfers’ en ‘Consumentenprijzen; basisjaren vanaf 1969’ het populairst zijn bij meerdere gebruikerstypes, onder wie de private sector en het bank- en verzekeringswezen.



Figuur 9: verdeling type gebruikers per dataset CBS (januari en februari 2019)

Ten slotte toont Figuur 10 de tien grootgebruikers van CBS data in januari en februari van 2019.¹⁰ Hierbij zien we dat de meest frequente gebruikers van CBS data in de eerste twee maanden van 2019 vooral de overheid, universiteiten en hogescholen waren. De ING-Group is de enige private organisatie in de top tien van grootverbruikers. Hierbij moet worden opgemerkt dat deze CBS datasets waarschijnlijk ook regelmatig door makelaars en potentiële kopers via diensten zoals Funda.nl en Jaap.nl worden aangeroepen. Maar omdat veel van dergelijke diensten via mobiele applicaties worden aangeroepen, vallen deze onder de categorie 'ISP'.



Figuur 10: Top 10 gebruikers CBS data (januari en februari 2019)

3.2.2 Analyse voor PDOK open data gebruikers

Publieke Dienstverlening Op de Kaart (PDOK) is een platform voor het ontsluiten van geografische informatie van Nederlandse overheden. PDOK-diensten zijn gebaseerd op open data en voor iedereen vrij beschikbaar.

PDOK is een samenwerking tussen het Kadaster, de ministeries van Infrastructuur en Waterstaat, Binnenlandse Zaken en Koninkrijksrelaties en Economische Zaken en Klimaat, Rijkswaterstaat en Geonovum. PDOK is een open initiatief waarbij elke overheidsorganisatie die zijn geodata voor hergebruik beschikbaar wil stellen, zich tot PDOK kan wenden.

¹⁰ Voor zover gebruikers geïdentificeerd konden worden. Het is uiteraard mogelijk dat er meer grootgebruikers zijn waarvan de bedrijfsnaam niet op basis van IP-nummer geïdentificeerd kon worden, en dus tot de ISP categorie worden gerekend.

PDOK is opgezet als een voortvloeisel van de zogenaamde INSPIRE Richtlijn.¹¹ Deze richtlijn richt op de implementatie van een infrastructuur voor ruimtelijke informatie in de Europese Gemeenschap en is verankerd in de Implementatiewet EG-richtlijn infrastructuur ruimtelijke informatie van 2009. De implementatiewet beschrijft hoe de gegevens over de fysieke leefomgeving uitwisselen. Denk aan gegevens over de topografie, vervoersnetwerken, bevolkingssamenstelling, emissie van vervuilende of gevaarlijke stoffen, waterkwaliteit en energiebronnen. Deze gegevens zijn nodig voor het opstellen en monitoren van Europees milieubeleid.¹² De diensten die aan de ruimtelijke infrastructuur gekoppeld moeten worden, zijn zoekdiensten (catalogi), raadpleegdiensten, downloaddiensten, verwerkingsdiensten en diensten die het mogelijk maken diensten met betrekking tot ruimtelijke gegevens op te roepen.¹³ Hoewel de infrastructuur in de eerste plaats bedoeld is voor het uitwisselen en delen van ruimtelijke data tussen overheden, moeten de zoekdiensten en raadpleegdiensten publiekelijk en gratis beschikbaar zijn.

De PDOK-diensten voldoen aan nationale en internationale standaarden, waaronder de Europese INSPIRE standaard en de Nederlandse e-overheidstandaarden. Bij PDOK kan men open datasets van de overheid met actuele geo-informatie vinden. Deze datasets zijn benaderbaar via geo-webservices, RESTful API's en zijn beschikbaar als downloads en een aantal datasets ook beschikbaar als linked data.

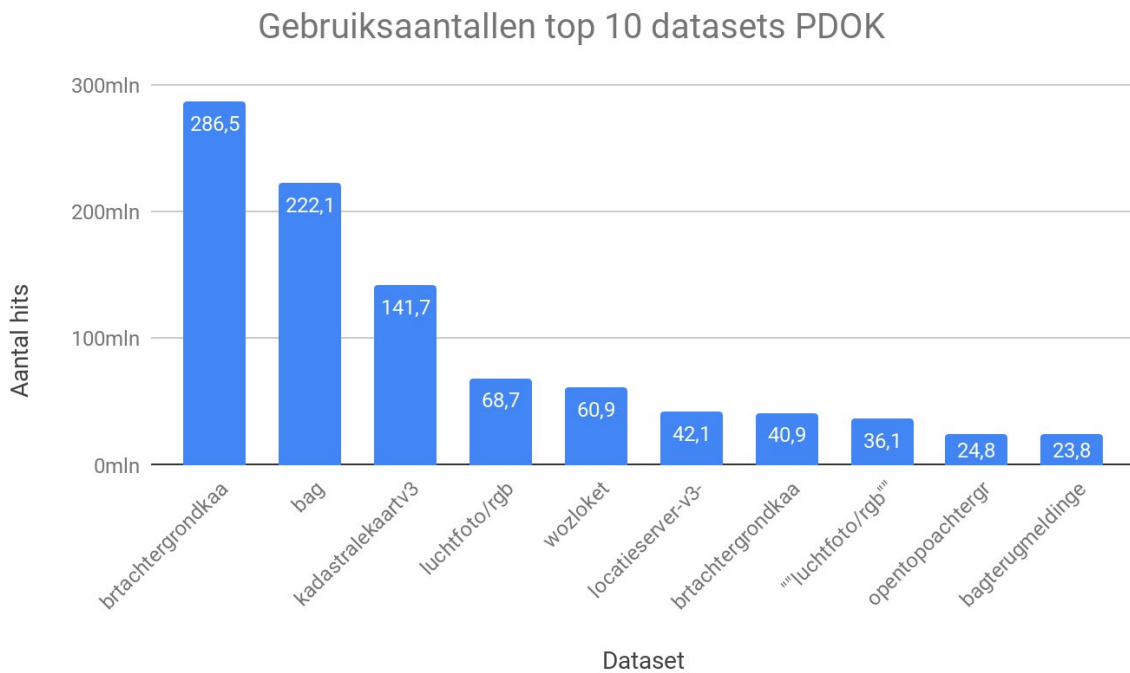
Voor de analyse van PDOK log bestanden zijn de logfile gegevens van januari en februari 2019 geanalyseerd. Populaire datasets zijn onder andere de Basisregistratie Topografie (BRT) achtergrondkaart en de Basisregistratie Adressen en Gebouwen (BAG). Figuur 11 geeft een overzicht van de top 10 meest voorkomende hits in de logfiles van PDOK. Locatieserver-v3-lookup is een geocodeerservice API van PDOK.¹⁴

¹¹ Richtlijn 2007/2/EG van het Europees Parlement en de Raad van 14 maart 2007 tot oprichting van een infrastructuur voor ruimtelijke informatie in de Gemeenschap (Inspire), PbEG L 108/50-63, p.1-14, beschikbaar via: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:NL:PDF> (geraadpleegd 30 juni 2019)

¹² Grothe, M. (2016). INSPIRE verankert zich in milieubeleid. Geo Info. 2016-2: 44-47. https://www.geonovum.nl/uploads/documents/201602GeoInfo_INSPIRE_0.pdf, (geraadpleegd 30 november 2018)

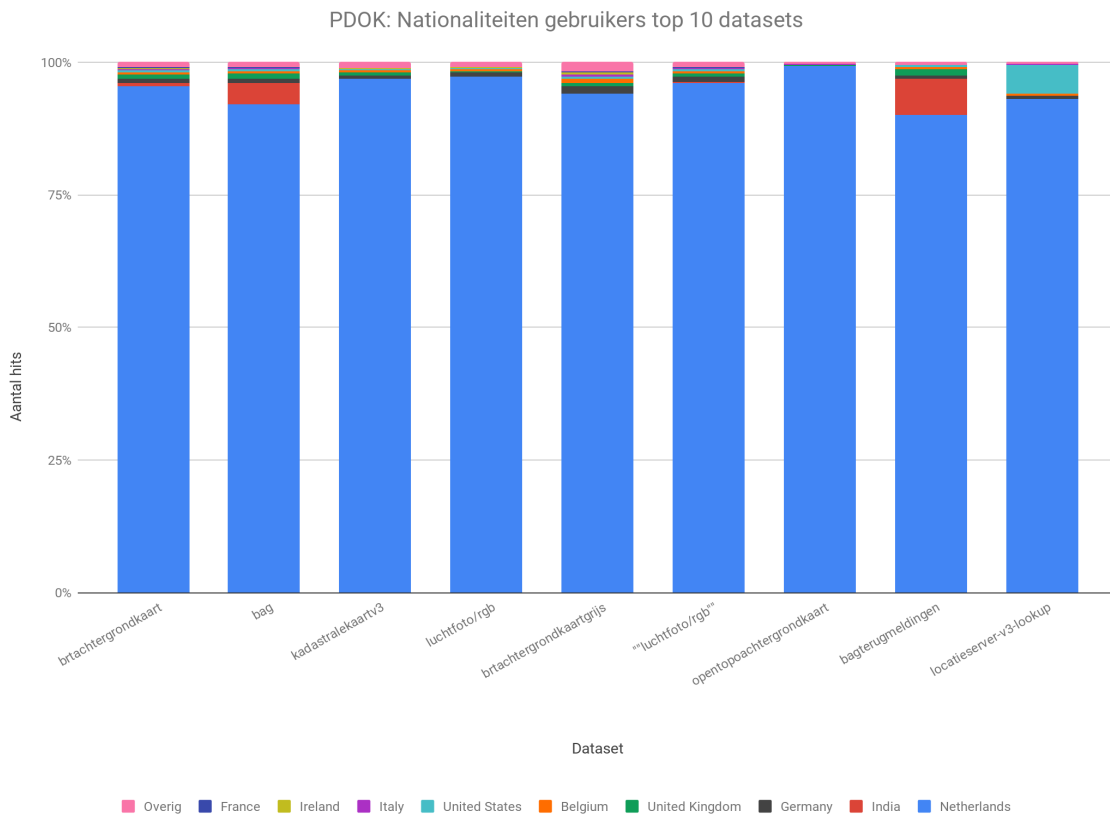
¹³ Implementatiewet EG-richtlijn infrastructuur ruimtelijke informatie (Besluit Inspire), Stb. 2009, 310, art. 9, lid 1. Beschikbaar via <https://wetten.overheid.nl/BWBR0026158/2009-09-01>

¹⁴ <https://github.com/PDOK/locatieserver/wiki/API-Locatieserver>



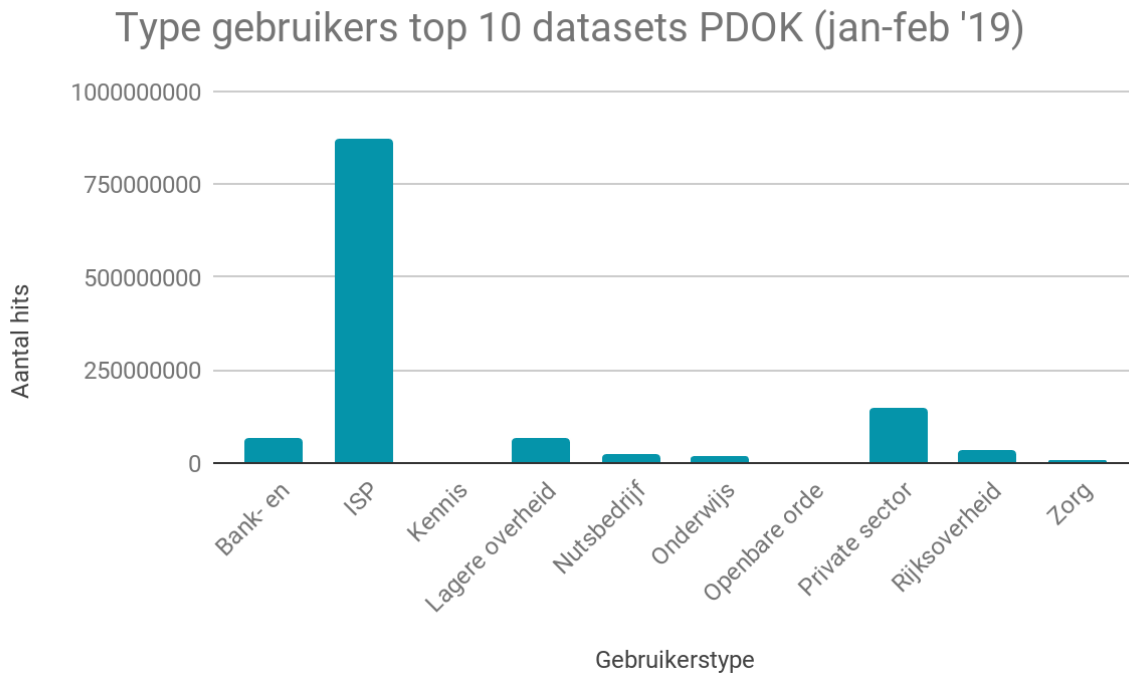
Figuur 11: Overzicht van de top 10 meest voorkomende hits in logfiles van PDOK (januari en februari 2019)

Vervolgens is gekeken naar de verdeling van het gebruik van de top 10 datasets per land, zie Figuur 12. Het merendeel van de gebruikers van de top 10 datasets komt uit Nederland (blauw), gevolgd door 'Overig' (circa 14%). Met 1,2% en 0,7% zijn India (donkerroze) en Duitsland (donkerblauw) daarna de grootste gebruikersgroepen. Van een circa 13,6% van de gebruikers kan de herkomst niet worden vastgesteld. Oorzaak hiervan is mogelijk dat deze IP-adressen niet zijn opgenomen in de online registers en daarom niet teruggeleid kunnen worden naar een land. Deze gebruikers zijn in de groep 'Overig' geplaatst, samen met de landen die vanaf positie 11 in de lijst voorkomen.



Figuur 12: top 10 verdeling PDOK gebruik per land (januari en februari 2019)

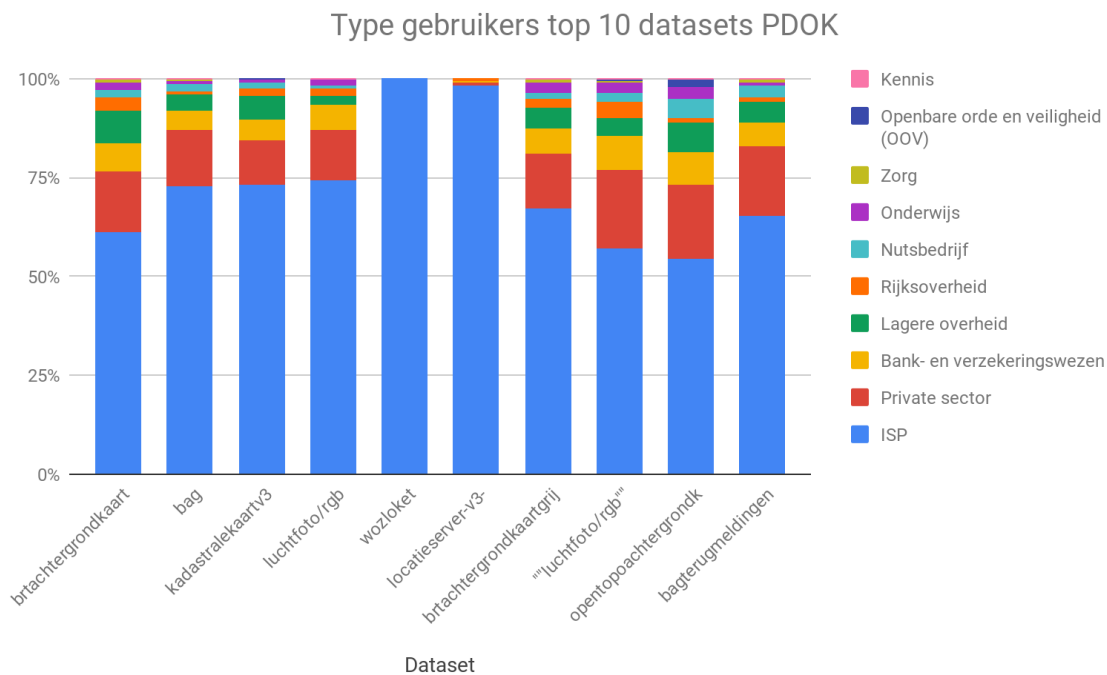
Tenslotte is gekeken naar het type gebruiker, zie Figuur 13. Het merendeel van de gebruikers bevindt zich in de categorie ISP. Vaak zijn dit burgers of gebruikers waarvan het IP-nummer is toegewezen aan een Internet Service Provider (ISP) zoals KPN of Ziggo. In deze categorie vallen ook organisaties waarvan de organisatiennaam niet aan het IP-adres is gelinkt, bijvoorbeeld kleinere ondernemingen waarvan de bedrijfsnaam niet uit het IP-adres is te achterhalen. Tot de ISP categorie behoren ook de vastelijns of mobiele verbindingen naar de provider toe. Mobiele telefoons die gebruik maken van apps op basis van PDOK open data zitten dus ook in deze categorie. . In de categorie 'Private Sector' vallen alle bedrijven die niet in één van de apart genoemde categorieën vallen, zoals bank- en verzekeringswezen, maar waar wel een bedrijfsnaam uit het IP-adres is te achterhalen. Andere categorieën van hergebruikers die te onderscheiden zijn, zijn de lagere overheden, de rijksoverheid, de nutsbedrijven en de onderwijssector.



Figuur 13: verdeling gebruik over gebruikersgroepen PDOK (januari en februari 2019)

Figuur 14 toont de verdeling van type gebruikers van de top 10 datasets PDOK in januari en februari 2019. Hierbij valt op dat - met uitzondering van wozloket - de tien meest populaire datasets door een variëteit aan gebruikerstypen worden gebruikt, en vooral ook door de private sector en het bank- en verzekeringswezen. Een verklaring voor de populariteit van wozloket data zou kunnen zijn dat in de eerste acht weken van het jaar de belastingaanslagbiljetten worden verstuurd zoals de aanslag onroerendezaakbelasting), die gebaseerd zijn op de zogenaamde WOZ-waarde¹⁵ van de woning of bedrijfspand. Via WOZ-waardeloket.nl kunnen de WOZ-waarden van woningen individueel worden geraadpleegd. Het ligt in de lijn der verwachting dat juist in de eerste twee maanden van het jaar de WOZ-beschikkingen worden gecontroleerd.

¹⁵ WOZ staat voor wet Waardering Onroerende Zaken. De WOZ-waarde wordt elk jaar door de gemeente bepaald en vormt de basis voor diverse belastingen en heffingen.



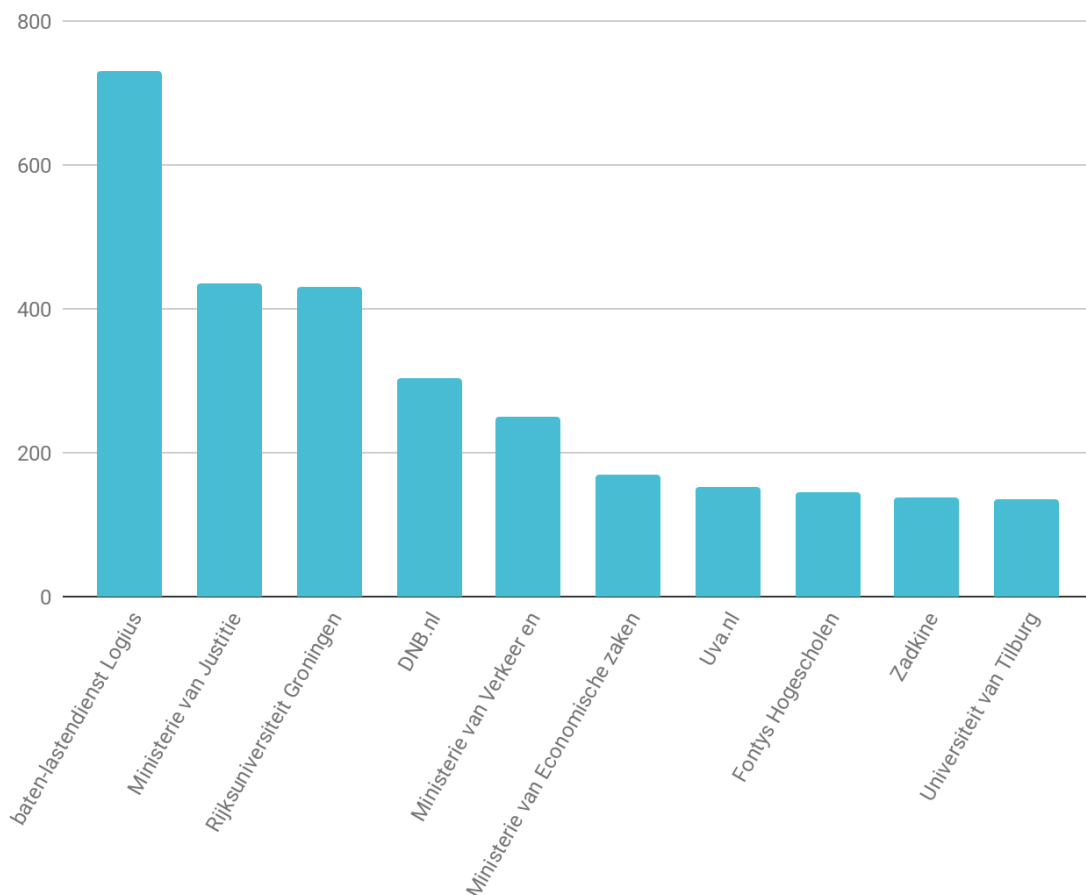
Figuur 14: verdeling type gebruikers per top 10 dataset PDOK (januari en februari 2019)

Ten slotte toont Figuur 15 de top 10 grootgebruikers van PDOK data in januari en februari 2019.¹⁶ Hierbij valt op dat de overheid, met name Logius,¹⁷ de universiteiten en hogescholen tot de meest frequente gebruikers van PDOK behoren. Maar ook de Nederlandse Bank is één van de grootgebruikers van PDOK.

¹⁶ Voor zover gebruikers geïdentificeerd konden worden. Het is uiteraard mogelijk dat er meer grootgebruikers zijn waarvan de bedrijfsnaam niet op basis van IP-nummer geïdentificeerd kon worden.

¹⁷ Logius is de dienst digitale overheid en onderdeel van het Ministerie van BZK. Logius beheert, onder meer, generieke ICT-voorzieningen zoals DigiD en MijnOverheid, en levert diensten aan andere overheidsorganisaties en organisaties met een publieke taak.

Top 10 gebruikers PDOK



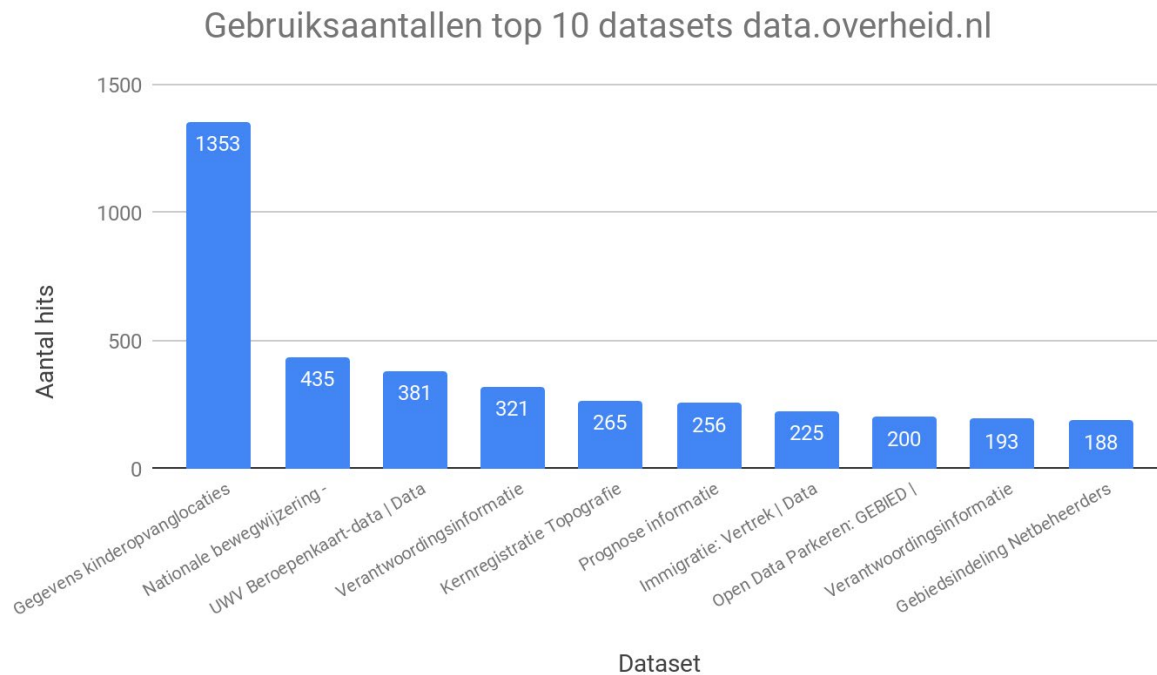
Figuur 15: Top 10 gebruikers van PDOK (januari en februari 2019)

3.2.3 Analyse data.overheid.nl

Het data.overheid.nl dataportaal bestaat uit een register met informatie over en verwijzingen naar datasets van Nederlandse overheden, en wordt beheerd door KOOP Kennis- en Exploitatiecentrum Officiële Overheidspublicaties. Data.overheid.nl biedt een actueel overzicht van alle beschikbare datasets binnen de Nederlandse overheid. Het dataportaal verwijst naar de vindplaats en/of naar de organisaties en contactpersonen achter een dataset. De functie wordt ingevuld door het dataregister en de synchronisatiefuncties met databronnen van overheidsorganisaties. Daarnaast kunnen via dit portaal ook verzoeken voor datasets worden ingediend.

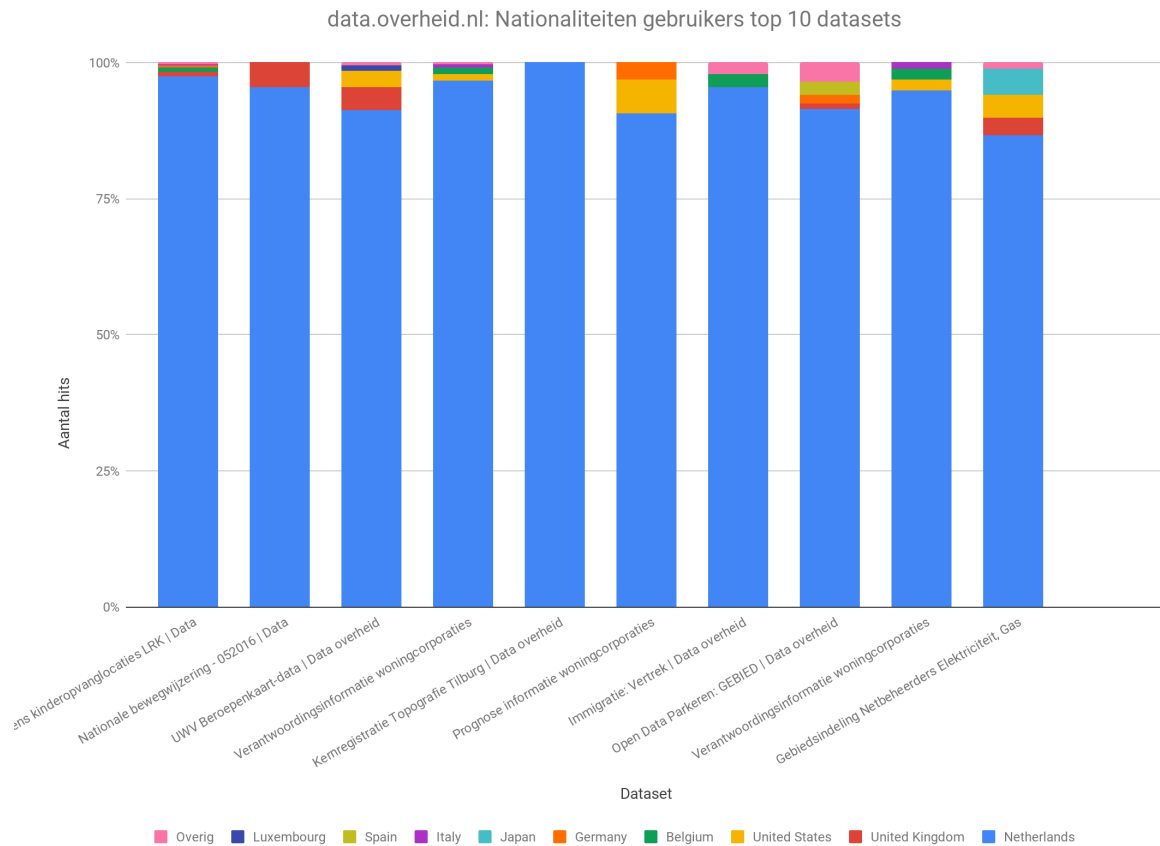
Data.overheid.nl biedt ondersteuning aan overheidsorganisaties bij het beschikbaar maken en houden van data, door bijvoorbeeld door het leveren van handreikingen en het ontwikkelen van standaarden, en ook door bijeenkomsten te organiseren. Data.overheid.nl werkt bijvoorbeeld samen met het Nationaal Georegister, het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, het Ministerie van Economische Zaken, het Ministerie van Infrastructuur en Waterstaat, CBS en met alle overheidsorganisaties die hun data aanbieden voor hergebruik.

Voor deze analyse is gebruik gemaakt van logfile gegevens over januari en februari 2019. Figuur 16 geeft een overzicht van de top 10 meest voorkomende rijen in de logfiles van data.overheid.nl. De meest bekeken dataset in die periode is: 'Verantwoordingsinformatie woningcorporaties ((dVi2017-hfd3)', en op de negende plaats is: 'Verantwoordingsinformatie woningcorporaties (dVi2017-hfd1)'.



Figuur 16: Overzicht van de top 10 meest voorkomende hits in logfiles van data.overheid.nl (januari en februari 2019)

Vervolgens is gekeken naar de verdeling van het gebruik van de top 10 datasets per land, zie Figuur 17. Het merendeel van de gebruikers van de top 10 datasets komt uit Nederland (blauw), gevolgd door 'Overig' (roze), het Verenigd Koninkrijk (rood) en de Verenigde Staten (geel). Van een circa 24% van de gebruikers kan de herkomst niet worden vastgesteld. Oorzaak hiervan is mogelijk dat deze IP-adressen niet zijn opgenomen in de online registers en daarom niet teruggeleid kunnen worden naar een land. Deze gebruikers zijn in de groep 'Overig' geplaatst, samen met de landen die vanaf positie 11 in de lijst voorkomen.

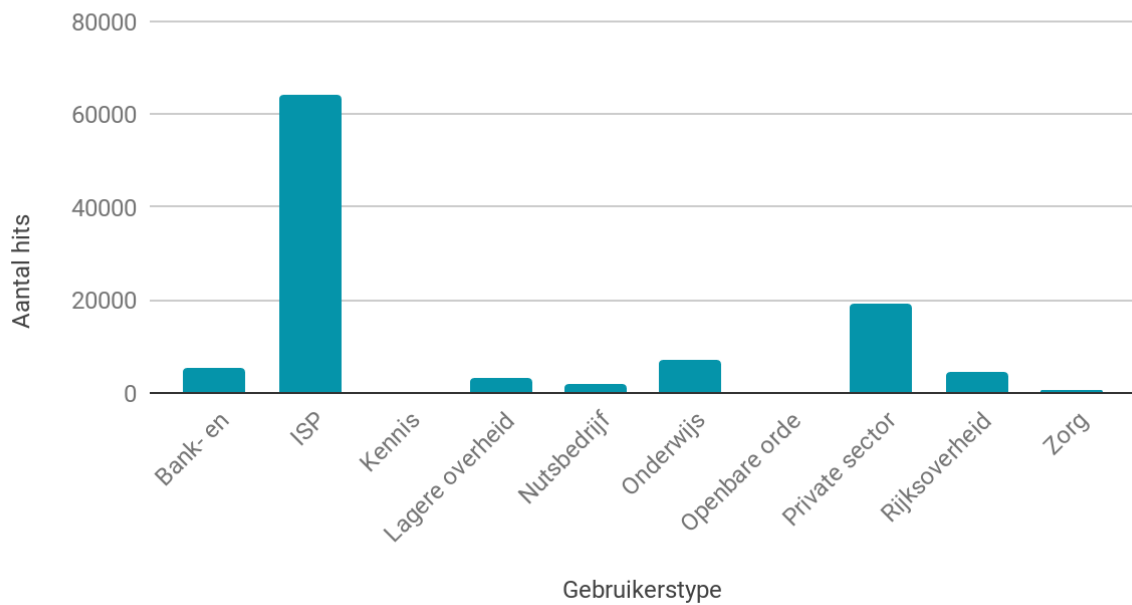


Figuur 17: verdeling gebruik top 10 datasets data.overheid.nl per land (januari en februari 2019)

Verder is gekeken naar het type gebruiker, zie Figuur 18. Het merendeel van de gebruikers bevindt zich in de ISP-categorie. Vaak zijn dit burgers of gebruikers waarvan het IP-nummer is toegewezen aan een ISP zoals KPN of Ziggo. In deze categorie vallen ook organisaties waarvan de organisatiename niet aan het IP-adres is gelinkt, bijvoorbeeld kleinere ondernemingen waarvan de bedrijfsnaam niet uit het IP-adres is te achterhalen. Tot de ISP-categorie behoren ook de vastelijns of mobiele verbindingen naar de provider toe. Mobiele telefoons zitten dus ook in deze categorie. In de categorie ‘Private Sector’ vallen alle bedrijven die niet in één van de apart genoemde categorieën vallen, zoals bank- en verzekeringswezen, maar waar wel een bedrijfsnaam uit het IP-adres is te achterhalen.

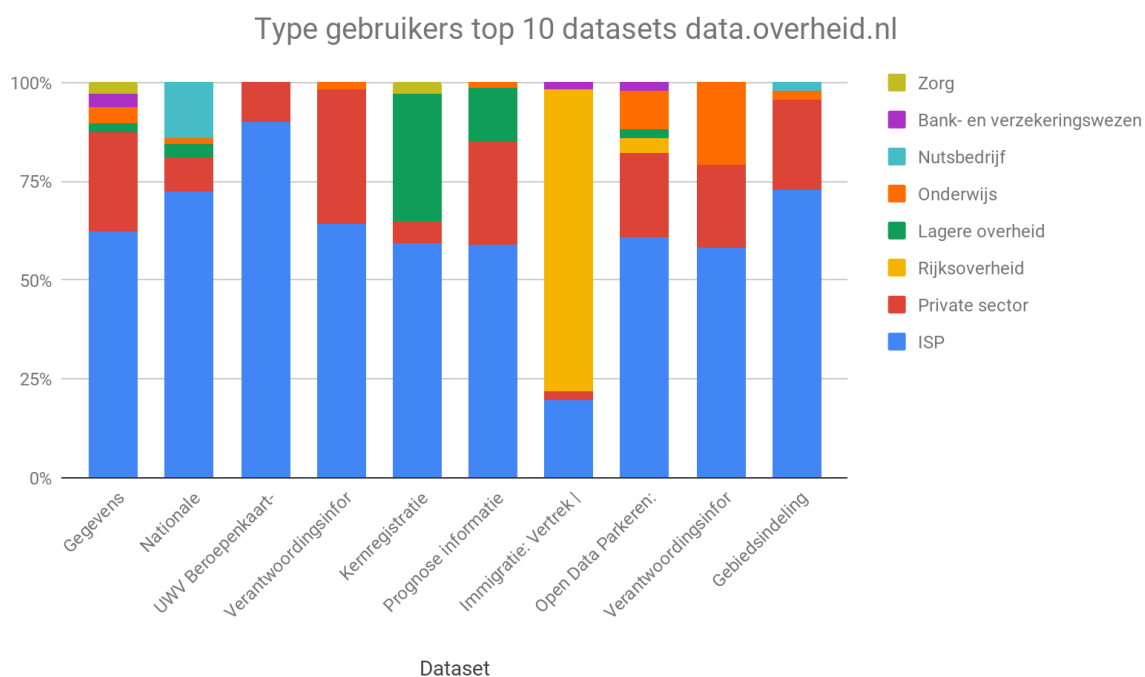
Ook bij data.overheid.nl vormt – na de ISP-categorie – de private sector, de overheid, organisaties met een publieke taak zoals nutsbedrijven, en het bank- en verzekeringswezen de grootste groep gebruikers van de top 10 datasets.

Type gebruikers top 10 datasets data.overheid.nl(jan-feb '19)

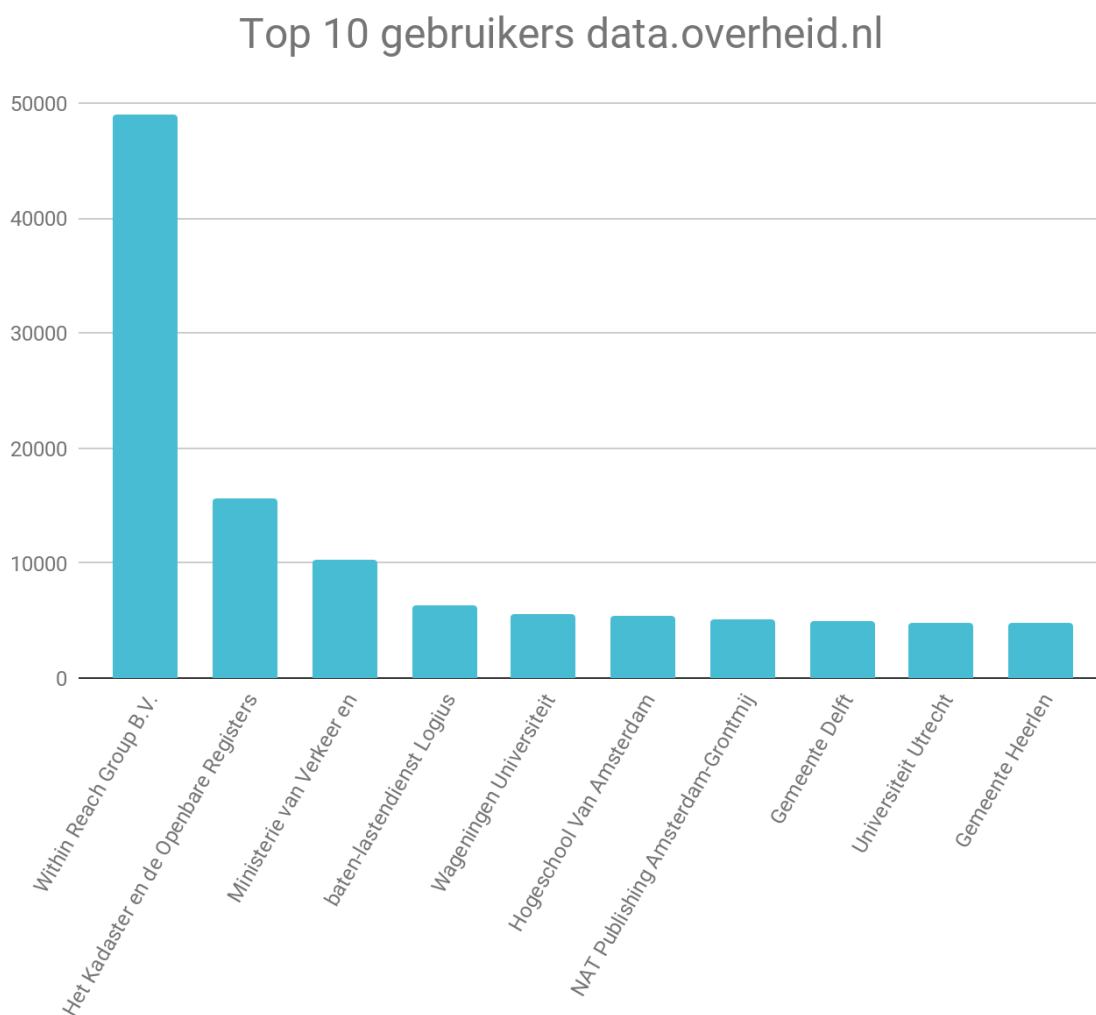


Figuur 18: verdeling gebruik over gebruikersgroepen data.overheid.nl (januari en februari 2019)

In Figuur 19 wordt de verdeling van type gebruikers van de top 10 datasets van data.overheid.nl weergegeven. Ook hier ziet men dat er een variëteit is aan type gebruikers voor de meeste van de top 10 meest aangeropen datasets, vooral voor de meest aangeropen dataset 'gegevens kinderopvanglocaties'. De uitzondering lijkt de 'Immigratie: Vertrek' die vooral door het Ministerie van Justitie en Veiligheid werd gebruikt in die periode. De kernregistraties worden veelal door de ISP-categorie en lagere overheden gebruikt, terwijl de verantwoordingsinformatie woningcorporatie datasets vooral door de ISP categorie en de private sector werden gebruikt.



Figuur 19: verdeling gebruik over gebruikersgroepen data.overheid.nl (januari en februari 2019)



Figuur 20: Top 10 (groot)gebruikers data.overheid.nl (januari en februari 2019)

In Figuur 20 wordt de tien grootgebruikers van data.overheid.nl in januari en februari 2019 weergegeven.¹⁸ Hierbij valt op dat een private holding bedrijf, Within Reach Group BV, met stip het vaakst data.overheid.nl heeft aangeroepen in januari en februari 2019. Gaat men echter inzoomen op de daadwerkelijk gebruikte datasets via data.overheid.nl, dan blijkt dit beperkt te zijn. De meeste datasets die daadwerkelijk door deze organisatie zijn gebruikt in de onderzochte periode zijn vooral datasets via PDOK, zie Bijlage 1. Behalve Within Reach Group BV zien we dat vooral overheden, universiteiten en hogescholen de meest frequente gebruikers van data.overheid.nl zijn.

Within Reach Group BV uit Almere is een moederbedrijf van onder meer 14IP en Voiceworks, en heeft sinds juli 2018 krachten gebundeld met een Duits en een Frans bedrijf om een marktleider te willen worden op het gebied van Unified Communications als een Service (UCaaS), d.w.z. een cloudgebaseerd communicatiesysteem voor meerdere kanalen en apparaten, in Europa.

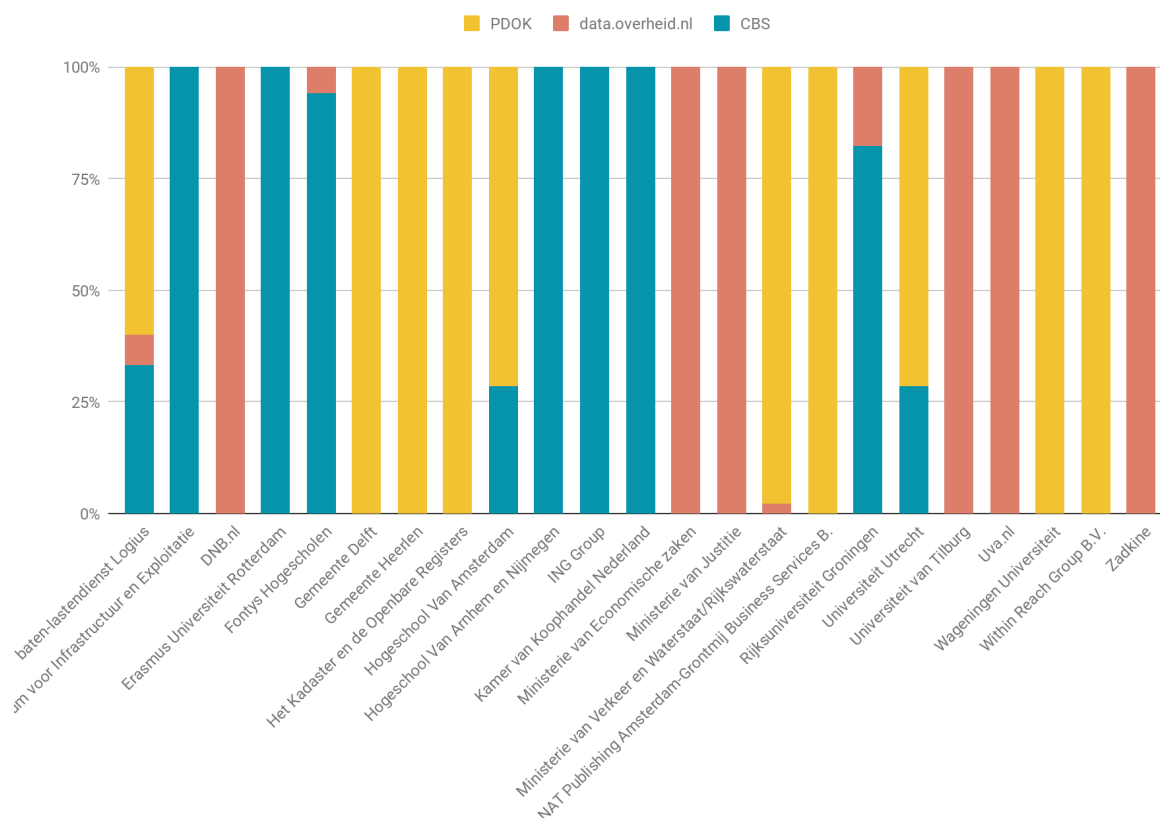
3.2.4 Open data platforms in samenhang

Beschouwt men de grootgebruikers van open data voor de drie onderzochte open data platforms, zie Figuur 21, dan ziet men dat een aantal van de grootgebruikers van meerdere dataplatforms gebruik maken, maar alleen Logius van alle drie platforms. De meeste grootgebruikers hebben een voorkeur voor één dataplatform, en de universiteiten en hogescholen van twee dataplatforms. Zo maken de gemeenten Delft en Heerlen vooral gebruik van PDOK, en de Ministeries van Economische Zaken en van Justitie en Veiligheid vooral gebruik van data.overheid.nl. Echter, als men dit vergelijkt met de grootgebruikers van data.overheid.nl in Figuur 20, ziet men dat data.overheid.nl vooral als catalogus voor open data wordt gebruikt. Zo wordt Within Reach Group als de grootste gebruiker van data.overheid.nl geïdentificeerd in Figuur 20, maar in Figuur 21 ziet men dat dezelfde organisatie vooral gebruik heeft gemaakt van PDOK.

In Bijlage 1 wordt een overzicht tussen welke specifieke datasets worden gebruikt door een grootgebruiker of de combinaties. Uit deze tabellen kunnen de dwarsverbanden worden gehaald. Zo blijkt uit deze tabellen dat Within Reach Group BV ook gebruik maakt van CBS data, maar niet tot de 10 grootste gebruikers behoort van CBS-data.

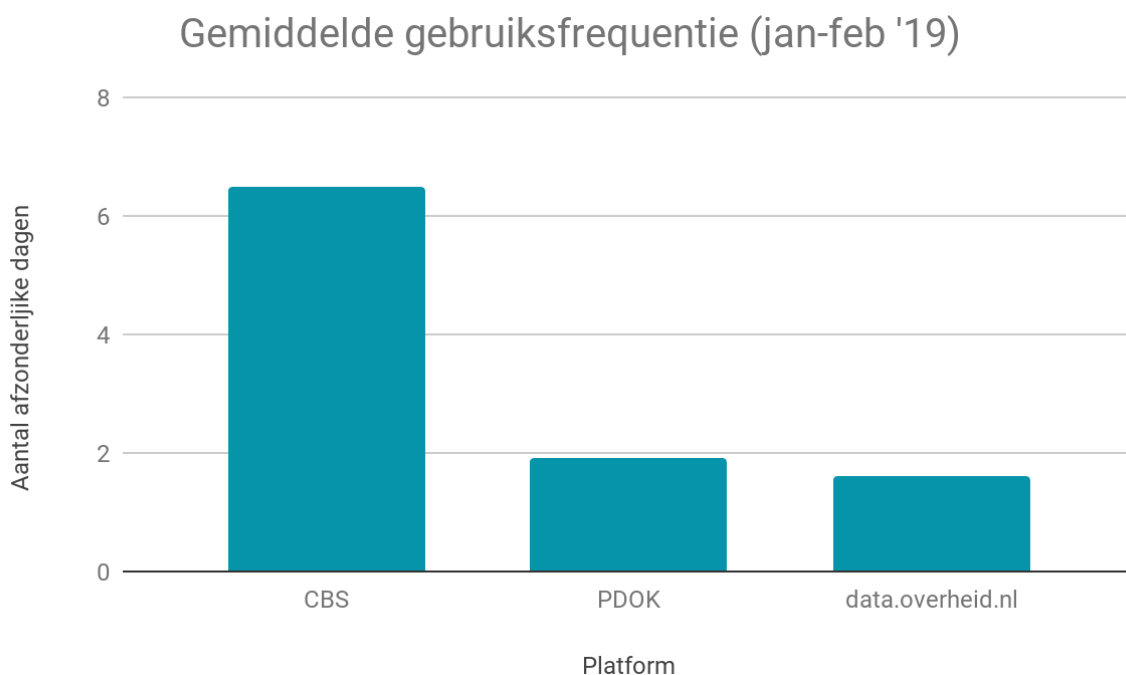
¹⁸ Voor zover gebruikers geïdentificeerd konden worden. Het is uiteraard mogelijk dat er meer grootgebruikers zijn waarvan de bedrijfsnaam niet op basis van IP-nummer geïdentificeerd kon worden.

Platformgebruik grootgebruikers open data



Figuur 21: platformgebruik grootgebruikers van de drie open data platforms (januari en februari 2019)

Verder kan men in Figuur 22 zien dat gebruikers van open data vaker terugkeren naar het open data platform van CBS in vergelijking met PDOK en data.overheid.nl. Bij het CBS zijn gebruikers gemiddeld 6,5 keer op afzonderlijke dagen langsgeweest, waarbij meerdere bezoeken op 1 dag niet meegeteld zijn. Een verklaring zou kunnen zijn dat gebruikers van CBS-data meer gebruik maken van het aanroepen van data via API's terwijl bij PDOK meer gebruik wordt gemaakt van download services.



Figuur 22: gemiddeld gebruiksfrequentie van de drie open data platforms (januari en februari 2019)

3.2.5 Conclusie kwantitatief onderzoek hergebruikers

Per platform zijn de statistieken van de webservers onderzocht om de hergebruiker van de open datasets te achterhalen. Uit analyse van de CBS-logfiles blijkt dat vooral de kerncijfer datasets en de data over bevolking en huishoudens populair waren onder hergebruikers in januari en februari 2019. Het gebruik van de overige datasets lag een stuk lager. Van een aanzienlijk deel van de IP-adressen (24% voor CBS IP-adressen en 14% van de PDOK IP-adressen) was er geen nadere informatie beschikbaar en kon geen land van herkomst van deze gebruikers achterhaald worden. Van de gebruikers van wie dit wel vastgesteld kon worden, kwamen de meesten uit het Nederland, gevolgd door de Verenigde Staten, het Verenigd Koninkrijk en Duitsland, waarvan het merendeel tot de ISP-categorie, de private sector en de overheid gerekend kunnen worden. De meest frequente gebruikers komen uit de overheid en van universiteiten en hogescholen, met de ING Group als enige private gebruiker in de top 10.

Via PDOK zijn vooral de basisregistraties populair zoals de Basisregistratie Topografie (BRT) en de Basisregistratie Adressen en Gebouwen (BAG). Ook zijn luchtfoto's en kadastrale gegevens vaak in die periode geraadpleegd. Van de IP-adressen die vanuit dit platform zijn aangeleverd kon een groot deel van de gebruikers worden geanalyseerd. Hierbij waren de bevindingen dat het voornamelijk ging om Nederlandse hergebruikers. De types gebruikers die vastgesteld konden worden waren voornamelijk particulieren en een klein aandeel zakelijke gebruikers (27,1%). De meest frequente gebruikers komen uit de overheid, de Nederlandse Bank, en van universiteiten en hogescholen.

Uit analyse van data.overheid.nl logfiles blijkt het gebruik fors lager dan van de andere twee open data platformen (orde grootte duizenden tegenover miljoenen). Het gebruik van de datasets lag dicht bij elkaar. De meeste gebruikers kwamen uit Nederland, waarvan het merendeel uit de ISP-categorie. Data.overheid.nl datasets kenden relatief het grootste aandeel gebruikers uit de onderwijssector. De meest frequente gebruiker in januari en februari was Within Reach Group BV, een privaat holding bedrijf, met de andere gebruikers afkomstig uit de overheid en van universiteiten

en hogescholen. Echter uit een verdere analyse blijkt dat data.overheid.nl vooral wordt gebruikt om naar andere platforms te gaan, en daar de datasets te gebruiken, een catalogusdienst dus.

3.3 Deelvraag 1b: Hoe zijn de gebruikers te categoriseren?

Uit ons literatuuronderzoek blijkt dat open data steeds meer hergebruikers uit verschillende sectoren aantrekken, en ook door steeds meer burgers worden gebruikt. Verder blijkt dat open data gebruikt worden door vooral nieuwe maar ook door langer bestaande bedrijven, en zowel door grote als door kleine bedrijven. In de rapporten waar de grootte van de organisatie onderzocht is, lijken het midden- en kleinbedrijf (inclusief microbedrijf)¹⁹ meer gebruik te maken van open data,²⁰ maar dat wordt waarschijnlijk veroorzaakt doordat deze bedrijven meer geneigd zijn aan deze onderzoeken deel te nemen dan grote bedrijven²¹. Uit een onderzoek uitgevoerd in het Verenigd Koninkrijk bleek dat grote ondernemingen zoals Google ook grootgebruikers zijn van geodata.²² Verder kunnen open data zowel voor commerciële doeleinden als voor niet-commerciële²³ en/of maatschappelijke doeleinden worden gebruikt.

3.3.1 Rollen van open data hergebruikers in open data waardeketen

Bij onze kwantitatieve analyse hebben we de hergebruikers kunnen categoriseren als ISP, overheid, private sector en de hoger onderwijs sector. Hoewel deze classificatie enig inzicht geeft in de hergebruiker en het hergebruik, is een verdere verfijning wenselijk om de hergebruiker en het hergebruik zodanig inzichtelijk te maken dat er beter inzicht wordt verkregen in de vraag wie de hergebruikers van open data zijn, en wat hun behoeften zijn. Een mogelijke verfijning zou kunnen worden gemaakt aan de hand van de rol die een hergebruiker van open data speelt in de open data waardeketen.

Deloitte LLP heeft in 2012 uit een lijst van meer dan 230 bedrijven en not-for-profit organisaties een indeling gemaakt op basis van drie rollen in de (open) data waardeketen: (1) open data aanbieder, (2) intermediair en (3) eindgebruiker, zie Figuur 23. Data aanbieders bieden open data aan in de vorm waarvoor de data oorspronkelijk zijn gebruikt zonder verdere bewerking. Intermediairs combineren de open data met andere (niet-open) data, verrijken de datasets, bieden platforms aan, bieden dienstverlening aan en/of ontwikkelen open data toepassingen. Eindgebruikers gebruiken de toegevoegde waarde data van de intermediairs, en kunnen zelf ook weer nieuwe producten/

¹⁹ In dit onderzoek maken wij gebruik van een onderverdeling van bedrijven naar grootte, waarbij microbedrijven minder dan medewerkers hebben, kleine bedrijven 10 tot 49 medewerkers, middelgrote bedrijven 50 tot 249 medewerkers en grote bedrijven meer dan 250 medewerkers.

²⁰ Zie bijv. Cipriano, P., C. Easton, E. Roglia & G. Vancauwenberghe (2013). A European Community of SMEs built on Environmental Digital Content and Languages. Final Report: 160. http://www.smespire.eu/wp-content/uploads/downloads/2014/03/D1.3_FinalReport_1.0.pdf, (geraadpleegd 1 december 2018)

²¹ Zie bijv. Berends, J., W. Carrara, W. Engberts & H. Vollers (2017a). Re-using Open Data. A study on companies transforming open data into economic & societal value. Brussels, Capgemini Consulting: 2016. https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf, (geraadpleegd 6 maart 2019)

²² ConsultingWhere Ltd & ACIL Tasman (2012). Assessing the value of Ordnance Survey OpenData™ to the economy of Great Britain. Full study interim report: 107

²³ Dit onderscheid is niet altijd scherp af te bakenen. Als een commercieel bedrijf open data gebruikt om de bedrijfsvoering effectiever te maken, dan kan men dat als een commercieel doeleinde bestempelen ook al worden er geen commerciële producten / diensten ontwikkeld.

diensten ontwikkelen. Deloitte maakt binnen de rol van intermediairs een onderscheid tussen verzamelaars, enablers, ontwikkelaars en verrijkers.²⁴

Verzamelaars zijn organisaties die open data verzamelen en aggregeren, al dan niet in combinatie met niet-open data. Dit verzamelen en/of aggregeren van data vindt vaak plaats op sectorniveau of geografisch gebied. Door het verzamelen en/of aggregeren van data uit verschillende bronnen kunnen de data worden gebruikt voor bijvoorbeeld het efficiënt aanbieden van data, of voor het uitvoeren van analyses (door verzamelaars zelf of door andere gebruikers). Een voorbeeld is het Nationaal Georegister²⁵ dat Nederlandse geodata toegankelijk maakt.

Enablers zijn organisaties die een platform en technologieën aanbieden zodat derden gebruik kunnen maken van open data. De enablers maken zelf geen gebruik van de data, maar bieden kosteneffectieve en efficiënte oplossingen aan, zowel aan aanbieders als aan (her)gebruikers van open data. Enablers verwerken de ruwe informatie in een dienstenplatform waar verzamelaars van meerdere domeinen worden samengebracht. Naast het beschikbaar stellen van open data, kan een enabler aanvullende diensten beschikbaar stellen zoals communicatie, advies, maatwerkdiensten en trainingen. Enablers kunnen ook als makelaars tussen data-aanbieders en gebruikers fungeren, bijv. door te lobbyen voor het publiceren van data die nog niet als open data beschikbaar zijn.

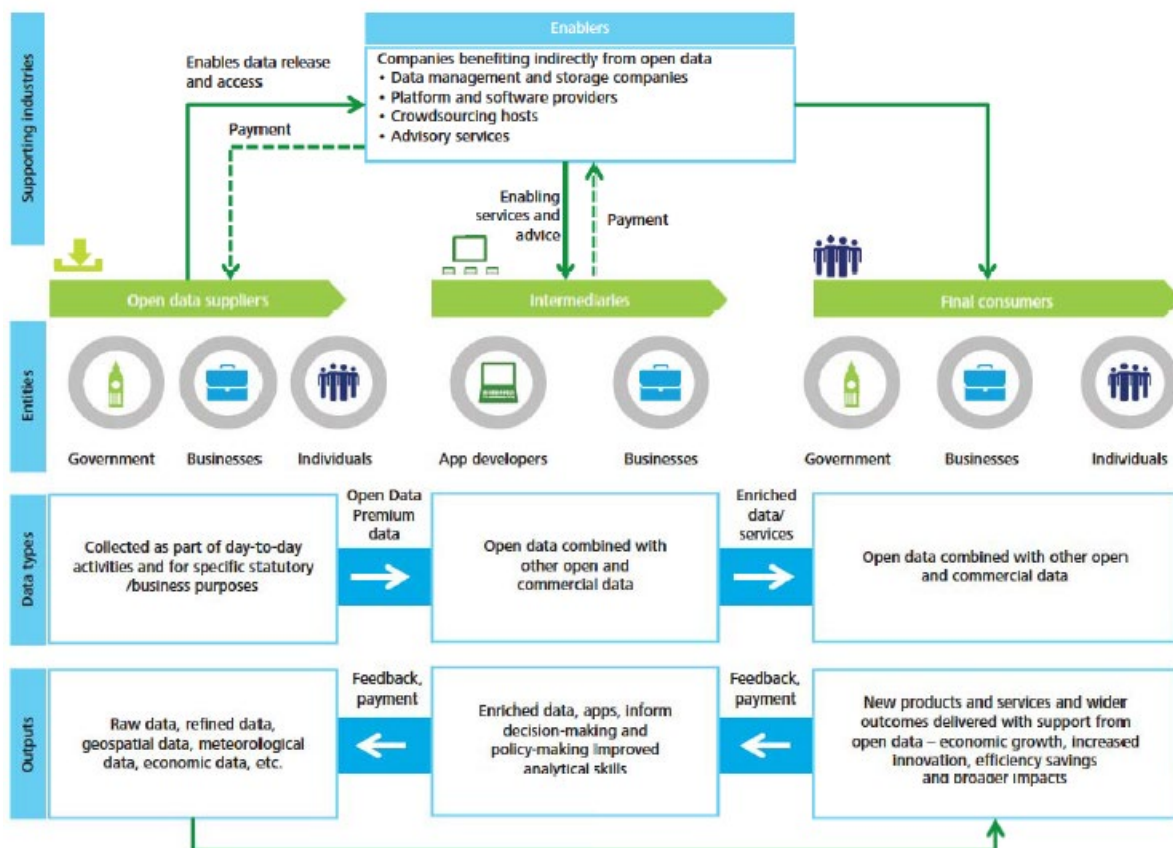
Ontwikkelaars zijn organisaties en softwareondernemers die tools en toepassingen ontwikkelen om de data nog beter te gebruiken. Ze ontwerpen, bouwen en verkopen bijvoorbeeld web- en slimme telefoontoepassingen (vaak gebaseerd op dynamische data). De afnemers van de diensten van de ontwikkelaars kunnen consumenten zijn maar ook overheden omdat open data aantrekkelijker en informatiever kan worden aangeboden. De ontwikkelaars onderscheiden zich van de enablers doordat zij wèl direct gebruik maken van open data.

Verrijkers verwerken de ruwe informatie tot nieuwe producten, al dan niet met behulp van ontwikkelaars. Dit zijn organisaties die open data veelal gebruiken om hun bestaande producten / diensten te verbeteren om nieuwe of betere inzichten kunnen leveren. Dergelijke producten / diensten zouden zonder open data ook kunnen bestaan, maar kunnen nu efficiënter en effectiever worden ingezet, bijvoorbeeld door verzekeraars of ingenieursbureaus.

Eindgebruikers zijn de gebruikers van de producten en diensten die op basis van open data zijn ontwikkeld. Dit kunnen zowel consumenten zijn als het bedrijfsleven en de overheid. Eindgebruikers kunnen ook weer aan het begin van een nieuwe waardeketen staan.

²⁴ Deloitte LLP (2012). Open Growth: Stimulating demand for open data in the UK. A briefing note from Deloitte Analytics. London, Deloitte Touche Tohmatsu Ltd: 12. <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf> (geraadpleegd 8 oktober 2018)

²⁵ <http://www.nationaalgeoregister.nl/> is de catalogus webservice voor geografische informatie opgezet in kader van de INSPIRE Richtlijn, en is gericht op de professionele gebruiker.



Figuur 23: de opkomst van de open data 'marktplaats'. Bron: Deloitte LLP/ODI 2012, p.4

Andere onderzoekers hebben andere onderverdelingen van de verschillende rollen gemaakt.²⁶ Dergelijke onderverdelingen kunnen gebaseerd zijn op het type product / dienst ontwikkeld door de hergebruiker, of betreft een verdere onderverdeling van de rollen gecategoriseerd door Deloitte. Voor dit onderzoek zijn wij van de onderverdeling van Deloitte uitgegaan.

In de voorgaande beschrijving hebben wij open data hergebruikers in algemene zin beschouwd. In de rest van dit hoofdstuk beschouwen wij als 'de hergebruiker' alle partijen die deelgenomen hebben aan dit onderzoek, d.w.z. de respondenten van de vragenlijsten en de geïnterviewde hergebruikers.

3.3.2 Categorisering op basis van de logfile analyse

Gebaseerd op de analyses uit de logfile data kunnen gebruikers ingedeeld worden in de categorieën, "overheid", "private sector" (met bedrijfsnaam in IP-adres), "ISP-categorie" (geen bedrijfsnaam in IP-adres), en "hoger onderwijs sector". Uit de analyses van de loggegevens blijkt dat er ook gebruikers uit het buitenland zijn. Deze voor de hand liggende indeling van hergebruikers geeft enige richting en inzicht maar verfijning van de categorieën is wenselijk. Voor een gewenste verfijning van categorieën van hergebruik kan gebruik worden gemaakt van de rol die een hergebruiken vervult in de open data waardeketen, zoals hierboven beschreven.

²⁶ Zie bijv. Janssen, M. & A. Zuiderwijk (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review* 32(5): 563-576 DOI: 10.1177/0894439314525902; den Haan, D. (2018). Intermediation as a dissolver of barriers. How intermediaries can help overcome barriers in open data use. GIMA. Utrecht, Utrecht University. M.Sc Thesis: 59. <https://dspace.library.uu.nl/handle/1874/368825>; Vancauwenberghe, G., F. Welle Donker, & B. van Loenen (2017). Business Models for Geographic Information. In J. B. Kruse, J. Crompvoets, & F. Pearlman (Eds.), *GEOValue: The Socioeconomic Value of Geospatial Information*. (pp. 85–115). CRC Press.

3.3.3 Categorisering op basis van rol in de open data waardeketen

Hoewel de categorisering van Deloitte LLP uit 2012 een goed aanknopingspunt geeft, blijkt uit later onderzoek²⁷ en ook uit dit onderzoek dat de verschillende rollen van verzamelaars, enablers, ontwikkelaars en verrijkers in de open data waardeketen in de praktijk vaak door elkaar lopen, vooral ook door de dynamiek van de markt. Verder kunnen de intermediaire organisaties zowel uit de publieke sector als uit de private sector komen. Intermediairs variëren van maatschappelijk (betrokken burgers, not-for-profit organisaties, kennisinstututen) tot commerciële organisaties (ZZP, MKB, multinational).²⁸ En commerciële organisaties kunnen eindproducten ook weer als open data beschikbaar stellen.

3.4 Deelvraag 1.c. Welke dwarsverbanden zijn zichtbaar voor het type gebruiker, en de soort en hoeveelheid data die gebruikt wordt?

In Bijlage 1 worden de datasets per geïdentificeerde grootgebruiker weergegeven. Echter, er is meer onderzoek nog om de dwarsverbanden te kunnen analyseren.

²⁷ Berends et al. (2017a)

4 Vraag 2: Op welke manier wordt open data hergebruikt

4.1 Hergebruikers in de praktijk

Om een beeld te krijgen van hergebruikers en het hergebruik van open data in de praktijk is er een online enquête uitgezet onder hergebruikers via social media. De enquête is door 567 mensen benaderd, en door 35 personen ingevuld (6% respons). Daarnaast zijn er met 12 hergebruikers aanvullende en verdiepende interviews gehouden. Ten slotte hebben wij onder de geïnterviewde hergebruikers een online evaluatie gehouden, waarin men aan kon geven welke aspecten als grootste belemmeringen werden ervaren. Uit de online enquête en uit de vraaggesprekken blijkt dat er een grote diversiteit onder hergebruikers is, zowel qua type hergebruiker en rol in de open data waardeketen als welke sector zij afkomstig zijn en hun mate van open data volwassenheid.

Uit de enquête blijkt dat bijna een derde van de respondenten zich tot de ‘verrijker’ categorie rekenen en 29% tot de ‘enabler’ categorie. De rest van de gebruikers zien zich als ontwikkelaar en als eindgebruiker. 11% van de respondenten gaven aan data voor onderzoek te gebruiken (zowel voor wetenschappelijk onderzoek als voor datajournalistiek). De overige deelnemers hebben niet aangegeven voor welk doel zij open data gebruiken.

De enquête is vooral ingevuld door hergebruikers uit het bedrijfsleven (18 respondenten). 10 van de 18 respondenten uit het bedrijfsleven behoren tot micro bedrijven (minder dan 10 medewerkers) en 5 respondenten behoren tot kleine bedrijven (10-49 medewerkers). Overige respondenten zijn afkomstig uit de overheid, not-for-profit organisaties, burgers, journalisten, en de wetenschappelijke/ onderwijs sector.

De meeste bedrijven die als intermediair opereren, gebruiken zowel open data als data waarvoor betaald moet worden. Een aantal van de micro-bedrijven en de not-for-profit organisaties geven aan alleen open data te gebruiken omdat hun businessmodel is gebaseerd op open data, bijvoorbeeld als ontwikkelaars van diensten of als aanjagers/ lobbyisten van open data.²⁹

De gebruikte categorieën van hergebruikers blijken in de praktijk veelal te overlappen. Veel hergebruikers blijken meerdere rollen aan te nemen en sommige hergebruikers opereren zowel met een commerciële als een non-profit doel. Een beeld dat tijdens de vraaggesprekken bevestigd werd.

Algemene bevindingen

Uit de enquête blijkt dat open data voor allerlei doeleinden worden hergebruikt. De meeste hergebruikers gaven aan dat ze de data voor meerdere doelen gebruiken. Bij het bedrijfsleven kunnen open data voor zowel commerciële (for-profit), voor maatschappelijke (not-for-profit) doelen, als voor het verbeteren van de interne bedrijfsvoering gebruikt worden. Bij overheden worden open data voor zowel het uitvoeren van publieke taken/ beleidsdoeleinden als voor maatschappelijke producten/ diensten, interne bedrijfsvoering en onderzoek gebruikt.

Uit de interviews blijkt eveneens dat een aantal van de hergebruikers zich tot verschillende categorieën rekenen en vanuit verschillende rollen in de waardeketen open data voor meerdere doeleinden gebruiken. Zo worden open data zowel voor maatschappelijke doelen ingezet als voor commerciële toepassingen binnen één organisatie. Door die ‘dubbele petten’ van een aantal hergebruikers (zowel hergebruik voor zowel maatschappelijke als voor commerciële doeleinden, of zowel werkend voor overheid/ kennisinstituut als zelf bijklussen) is de term ‘intermediair’ beter op zijn plaats dan de oorspronkelijke categorieën van verzamelaar, ontwikkelaar, enabler en enricher.

²⁹ Dit komt overeen met de rol van ‘communicator’ geïdentificeerd door den Haan (2018).

4.2 Op welke manier(en) gebruiken zij momenteel open data?

In deze sectie geven wij antwoord op de vraag waarvoor open data worden gebruikt, welke zoekstrategieën worden ingezet om open datasets te vinden, en via welke diensten open data worden gebruikt.

4.2.1 Deelvraag 2a. Voor welke doeleinden wordt open data gebruikt?

In de vragenlijst en de vraaggesprekken werden de volgende doeleinden genoemd:

- Om data van verschillende lagen of van lokale overheden via één centraal punt te ontsluiten, bijv. raadsinformatie of informatie voor specifieke sectoren
- Om modellen te ontwikkelen die zowel voor eigen bedrijfsvoering wordt gebruikt als voor externe klanten, onder wie vaak ook overheidsorganisaties
- Om software/ applicaties te ontwikkelen voor eigen gebruik
- Om toepassingen voor de overheid te ontwikkelen (zodat die effectiever gebruik kunnen maken van hun eigen data)
- Om overheidsdata in een gebruiksvriendelijker formaat te converteren en opnieuw beschikbaar te stellen.
- Voor onderzoek, zowel wetenschappelijk als voor journalistiek
- Voor het valideren van overheidsbeleid
- Om mee te experimenteren

Uit deze antwoorden blijkt dat open data vooral voor intermediaire toepassingen worden gebruikt, zowel voor commerciële toepassingen als voor maatschappelijke/ niet-commerciële toepassingen. Maar open data worden ook ingezet om de eigen bedrijfsvoering te verbeteren.

4.2.2 Deelvraag 2b. Op welke manieren vinden gebruikers open data?

Hergebruikers maken gebruik van verschillende zoekstrategieën, mede afhankelijk van hun skills, ervaring en achtergrond (open data volwassenheid).

Algemene zoekmachines als Google en Duckduckgo worden gebruikt voor het vinden van datasets waarvan de hergebruiker niet weet of deze bestaat (in een bepaald domein). Gedacht kan worden aan zoektermen als “gegevens over de weg”. Wanneer een zoekmachine geen bevredigend antwoord oplevert, dan worden de persoonlijke netwerken, forums/ community's, en/of sociale media ingezet om een vraag aan een breder publiek te stellen. Sommige hergebruikers gebruiken Twitter om een vraag aan iedereen te stellen. Andere hergebruikers gebruiken LinkedIn om meer (zakelijke) informatie over (open) data te vinden. Wikipedia werd ook genoemd als een bron om meer informatie over datasets te vinden, vooral over de typische eigennamen en bronhouders van specifieke datasets of over de geschiedenis van de datasets, zie als voorbeeld Figuur 24. Voor specifieke datasets waarvan de hergebruiker het bestaan kent, wordt er direct naar de bronhouder van de data gegaan. Wat meer ervaren hergebruikers maken meteen gebruik van hun eigen netwerk om direct contact op te kunnen nemen met de bronhouders.

Hergebruikers met een geo-informatie achtergrond maken vaker gebruik van PDOK en Nationaal GeoRegister (als catalogusdienst voor geodata) dan hergebruikers met een niet-geo achtergrond. Hergebruikers met een achtergrond in statistiek kennen hun weg naar CBS Open Data Statline (met name gebruikers van de oorspronkelijke versie van Statline). Data.overheid.nl wordt door de geïnterviewde hergebruikers voornamelijk gebruikt als catalogusdienst om te zien welke datasets nieuw zijn en minder als een portaal om de datasets daadwerkelijk te downloaden.



Figuur 24: screenshot met informatie over Nationaal Wegenbestand bron: https://nl.wikipedia.org/wiki/Nationaal_Wegenbestand (geraadpleegd 21 juni 2019)

4.2.3 Deelvraag 2c. Via welke weg krijgen hergebruikers toegang tot open data?

Hergebruikers verkrijgen open data veelal via de oorspronkelijke bronhouders (bijv. Gemeenten, Rijkswaterstaat, Kadaster, RDW), en via CBS en PDOK. Wetenschappelijke hergebruikers gebruiken CBS en PDOK veelal, maar hebben vaak al voldoende specialistische voorkennis. Voor sectorspecifieke data maken hergebruikers gebruik van dataportalen zoals NDW (verkeersinformatie) of NDOV (openbaar vervoer). Soms maken de hergebruikers gebruik van niet-overheidsplatforms en/of diensten zoals van Esri Nederland³⁰, OpenStreetMap, WAZE of NL Extract. Hoewel geïnterviewde hergebruikers aangaven ook data via data.overheid.nl te verkrijgen, gaven de geïnterviewde hergebruikers juist aan weinig gebruik te maken van data.overheid.nl om data te verkrijgen.

De respondenten van de vragenlijst en de geïnterviewde hergebruikers maken gebruik van zowel (bulk) download diensten als van API's, API's Light en ftp-servers voor hergebruik van de data. Hergebruikers die datasets regelmatig gebruiken hebben een voorkeur voor API's vooral voor (near) real-time. Andere hergebruikers downloaden liever de dataset zodat zij de benodigde bewerking offline kunnen uitvoeren, zeker als zij niet erg bedreven zijn met API's. Web viewing diensten worden door hergebruikers niet veel gebruikt, alleen als hulpmiddel om te zien welke data voor welke gebieden beschikbaar zijn.

³⁰ Esri is de ontwikkelaar van, één van de meest gebruikte software voor geografische informatie wereldwijd (ArcGIS). Esri Nederland converteert een aantal van de basisregistraties zoals de Basisregistratie Topografie van open source formaat naar ArcGIS formaat (shapefile). Vervolgens stellen zij die geconverteerde basisregistraties weer beschikbaar als open data via een download service en via ArcGIS Online (SaaS toepassing).

5 Vraag 3: tegen welke belemmeringen lopen hergebruikers aan?

5.1 Potentiele barrières voor open data hergebruikers

Een literatuuronderzoek uitgevoerd voor dit onderzoek wijst uit dat hergebruikers van open data tegen een aantal barrières aanlopen. De meest voorkomende belemmeringen die uit het literatuuronderzoek naar voren komen zijn onder te verdelen in drie categorieën: bekendheid, bereikbaarheid en bruikbaarheid.

Onder bekendheid vallen belemmeringen zoals onbekendheid met open data in algemene zin, maar ook bij hergebruikers die meer bekend zijn met open data de kenbaarheid en vindbaarheid van open data (hoe heet de dataset en waar kan die worden gevonden). Verder is een veel genoemde belemmering het versnipperd aanbod van open data. Niet alle overheidsdata zijn via centrale portalen te vinden, en niet alle gelijksoortige datasets zijn beschikbaar. Zo maken niet alle lokale overheden bepaalde datasets beschikbaar als open data die wel door andere lokale overheden beschikbaar worden gesteld.

Onder bereikbaarheid vallen organisatorische aspecten, zoals datasets die onder beperkende gebruikersvoorwaarden en/of tegen financiële vergoedingen beschikbaar worden gesteld, of waar het ontbreken van open data licenties/ publiek domein verklaringen zodat het voor hergebruikers niet duidelijk is of de data hergebruikt mogen worden voor verwerking met andere data. Onder bereikbaarheid valt ook technische bereikbaarheid, zoals via welke diensten data gebruikt kunnen worden. Kunnen data alleen worden ingezien of kunnen de data ook worden gedownload? Zijn er aanroepdiensten beschikbaar voor frequente hergebruikers?

Bruikbaarheid betreft belemmeringen veroorzaakt door bijvoorbeeld gebrekkige datakwaliteit, incomplete en niet-geactualiseerde data, ontbrekende of onvolledige metadata, gebrek aan standaardisatie, en bestanden beschikbaar in ongeschikte formaten. Maar ook dat er geen faciliteit is voor hergebruikers om feedback te leveren of verzoeken voor datasets in te dienen.

Een andere belemmering die uit het literatuuronderzoek naar voren komt, heeft ook met de hergebruiker zelf te maken. Niet alle hergebruikers beschikken over voldoende kennis en technische vaardigheden, zowel om de mogelijkheden te zien van de toegevoegde waarde van open data om bepaalde vraagstukken aan te pakken, maar ook ontbreekt het nog aan data-vaardigheden om de open data te kunnen gebruiken.³¹ Nog een belemmering is dat er een mismatch lijkt te zijn tussen de data die beschikbaar zijn, en de data waar hergebruikers behoefte aan hebben.³²

Er worden al decennialang onderzoeken naar de belemmeringen van hergebruik van overheidsdata uitgevoerd. Hoewel het beschikbaar stellen van overheidsdata als open data een aantal van de genoemde belemmeringen deels hebben ondervangen, worden in recente onderzoeken³³ nog steeds

³¹ Zie van Loenen, B. & F. Welle Donker (2014). De stand in opendataland. Delft, OTB - Research for the Built Environment: 116. <http://repository.tudelft.nl/view/ir/uuid%3A7b425db6-2f9a-4be6-b5f0-3492055268d8/> (geraadpleegd 30 juni 2019); van Loenen, B., F. Welle Donker & R. Braggaar (2016). De stand in opendataland 2016. Delft, Kenniscentrum Open Data: 96. <http://repository.tudelft.nl/islandora/object/uuid:b76cb0d8-e9b5-4125-8d7c-b452de1da0d0?collection=research> (geraadpleegd 30 juni 2019) en ; Berends et al. (2017a).

³² Zie ook IRM [Independent Reporting Mechanism] (2015). Aligning supply and demand for better governance. Washington, Independent Reporting Mechanism (IRM): 40. <http://www.opengovpartnership.org/sites/default/files/IRMReport-OpenData.pdf> (geraadpleegd 30 juni 2019)

³³ Zie ook Berends, J., W. Carrara, H. Vollers, T. Fechner & M. Kleemann (2017b). Analytical Report 5: Barriers in working with Open Data: 37. https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf,

belemmeringen genoemd die vooral betrekking hebben op bekendheid en bruikbaarheid. Uit ons literatuuronderzoek blijkt wel dat de financiële belemmeringen met het beschikbaar stellen van open data nu lijken af te nemen.³⁴

Door datasets via (centrale) dataportalen te ontsluiten kunnen een aantal van de technische belemmeringen worden ondervangen. In onze Social Media Open data Sentimentanalyse vormen de begrippen “gebruik”, “ervaring”, “portaal” en vindbaarheid” het uitgangspunt van de analyse.

5.2 Social media open data sentimentanalyse

Social media data kunnen als indicator worden gebruikt om aan te geven dat er iets aan de hand is of kan zijn. Daarnaast kan social media data context geven aan data.³⁵ Door de achterliggende data te analyseren, kan men een beeld krijgen van de beleving van open data hergebruikers.

5.2.1 Inleiding social media sentimentanalyses

Wij hebben onderzocht wat de gebruikersbeleving van open data is en welke eventuele barrières er worden ondervonden bij het verkrijgen en gebruiken van open data in Nederland, door middel van een social media analyse en Natural Language Processing (NLP). Social media platformen bieden een eenvoudige weg voor de gebruiker om meningen, klachten of juist complimenten te verkondigen. Alle berichten zijn digitaal en er is bij de berichten vaak veel metadata gekoppeld, zoals een locatie, tijdstempel en soms zelfs persoonlijke informatie. Dit in tegenstelling tot andere teksten op het internet, waarvan vaak buiten de opmaak (HTML) geen metadata van aanwezig is. Via NLP kunnen social media berichten automatisch worden geanalyseerd. Een ander voordeel van social media inhoud is dat er veel materiaal (de corpus) beschikbaar is voor tekstueel onderzoek, maar de kwaliteit is daarentegen vaak van lage kwaliteit.

Twitter is in dit geval uitgekozen als social media platform om het onderzoek uit te voeren omdat het openbare deel van Twitter beschikbaar zijn voor onderzoek. Door middel van een sentimentanalyse is eerst gemeten of een social media bericht (bekend als een Tweet op het Twitter platform) behorende bij bepaalde sleutelwoorden positief, neutraal of negatief is. Variabelen hierbij zijn bijvoorbeeld de context, toon en emotie. Door het sentiment van Tweets met sleutelwoorden als “open data” en “open geodata” te meten is een inschatting gemaakt van hoe op social media wordt gedacht over open data in Nederland.

Behalve een huidige stand van zaken kunnen de cijfers ook worden vergeleken met toekomstige cijfers om een beeld te geven van de ontwikkelingen op open data gebied in Nederland.

Daarnaast zijn de Tweets geclassificeerd om te kijken waar gebruikers tegenaan lopen tijdens het verwerken van open data. Van een aantal vooraf opgestelde oorzaken worden de aantallen berichten geteld die als zodanig geclassificeerd worden. Hierbij is gebruik gemaakt van het Machine Learning (ML) concept om het classificeren van grote aantallen teksten mogelijk te maken.

(geraadpleegd 30 juni 2019). Hun onderzoek was gebaseerd op een analyse van de European Data Portal Open Data Maturity in Europe rapporten van 2015 en 2016, European Data Portal (2017) Recommendations for Open Data Portals; en European Data Portal (2017), Re-using Open Data.

³⁴ Zie bijv. Berends et al. (2017b).

³⁵ Zie o.m. Van Loenen, B., F. Welle Donker & D. Kilic (2018). Overheid en “sociale” locatiedata: Een verkennend onderzoek naar de stand van social media (locatie)data gebruik Nederlandse overheid. Geonovum, Amersfoort: 38. <http://resolver.tudelft.nl/uuid:9187527d-6f48-4ccb-a6fa-c63930407999>

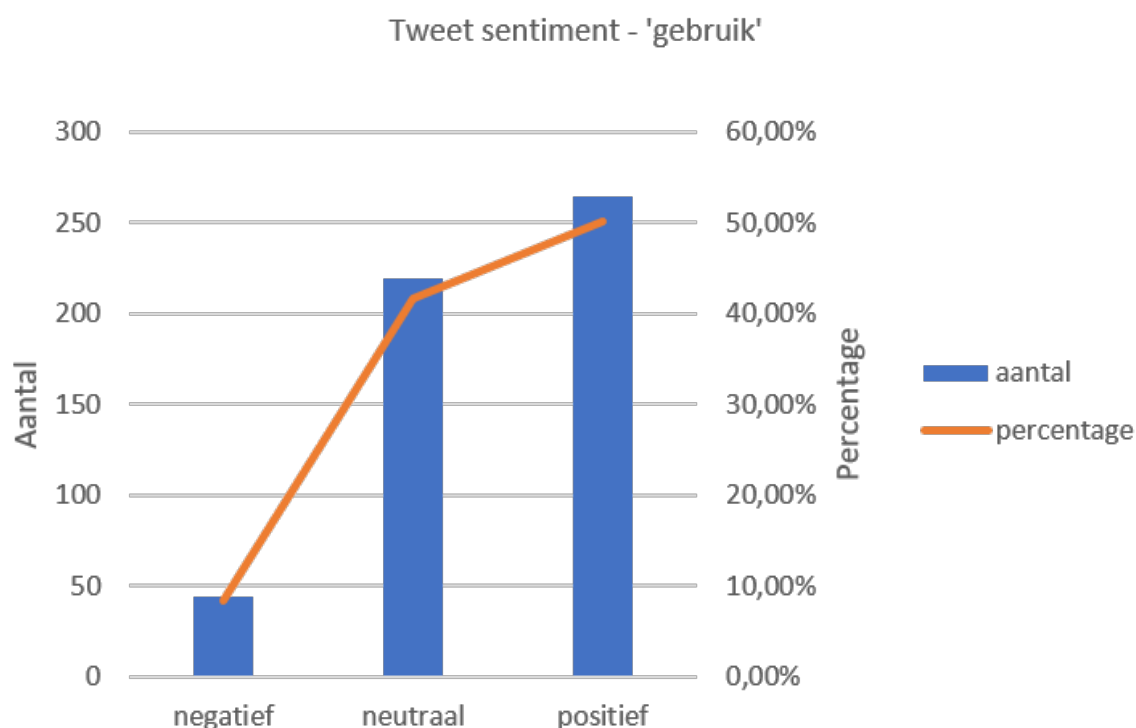
In Bijlage 4 staat een beschrijving van de gebruikte methodologie, inclusief een beschrijving welke tweets als negatief, positief of neutraal worden geclassificeerd.

5.2.2 Resultaten Twitter sentimentanalyse

Hieronder beschrijven we de resultaten per zoekopdracht. De zoekopdrachten gaan terug tot het jaar 2010 en tot 31 december 2018.

1 gebruik (527 tweets)

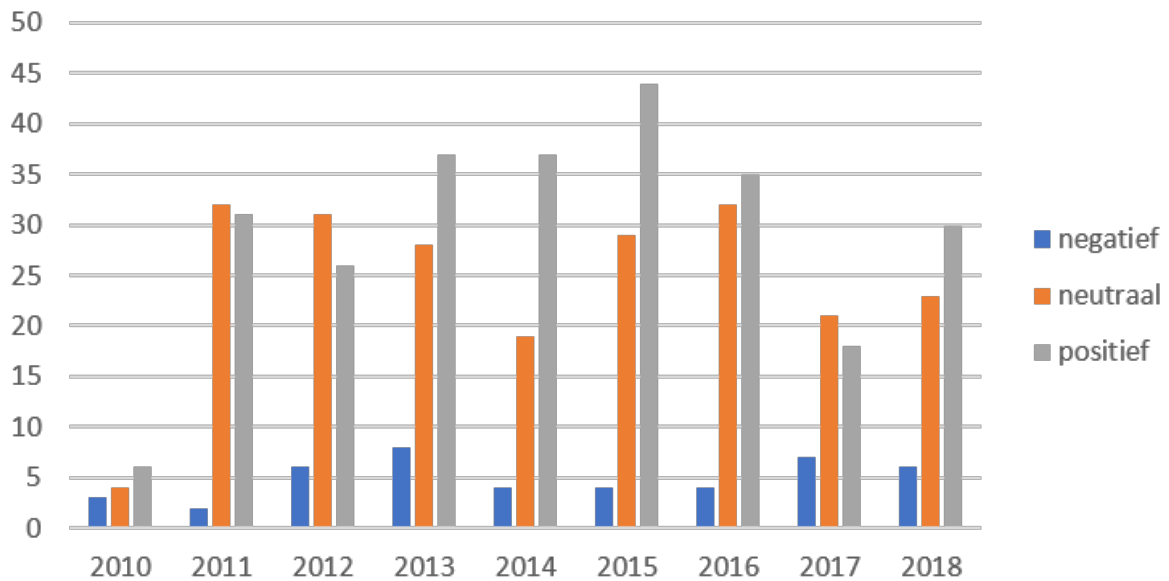
Zoekopdrachten waarin het gebruik werd vernoemd in combinatie met open data leverden 527 Tweets op. Het merendeel van de Tweets rond dit onderwerp werd gewaardeerd als positief. Dit kan duiden op positieve reacties naar aanleiding van het gebruik van open data of het aansporen van het gebruik van open data met hierin positieve termen.



Figuur 25: Tweet sentiment - 'gebruik' aantallen en als percentage 2010 t/m 2018 (527 tweets)

Relatief gezien zijn er over de jaren heen iets meer negatieve Tweets verschenen en wat minder neutraal gemarkeerde Tweets.

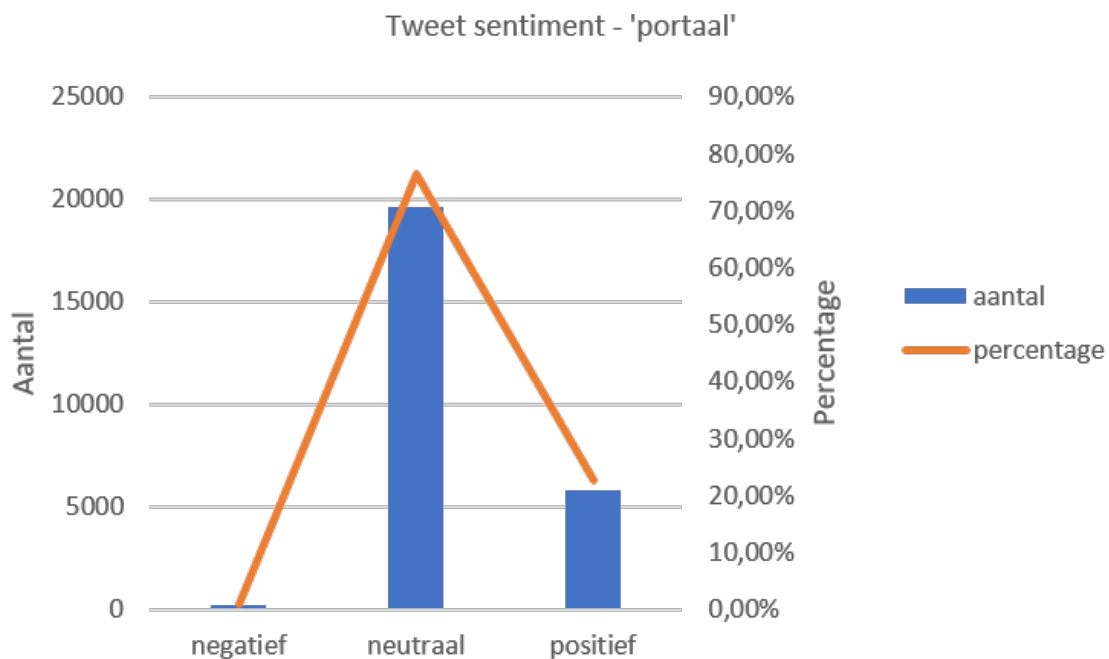
Tweet sentiment - 'gebruik'



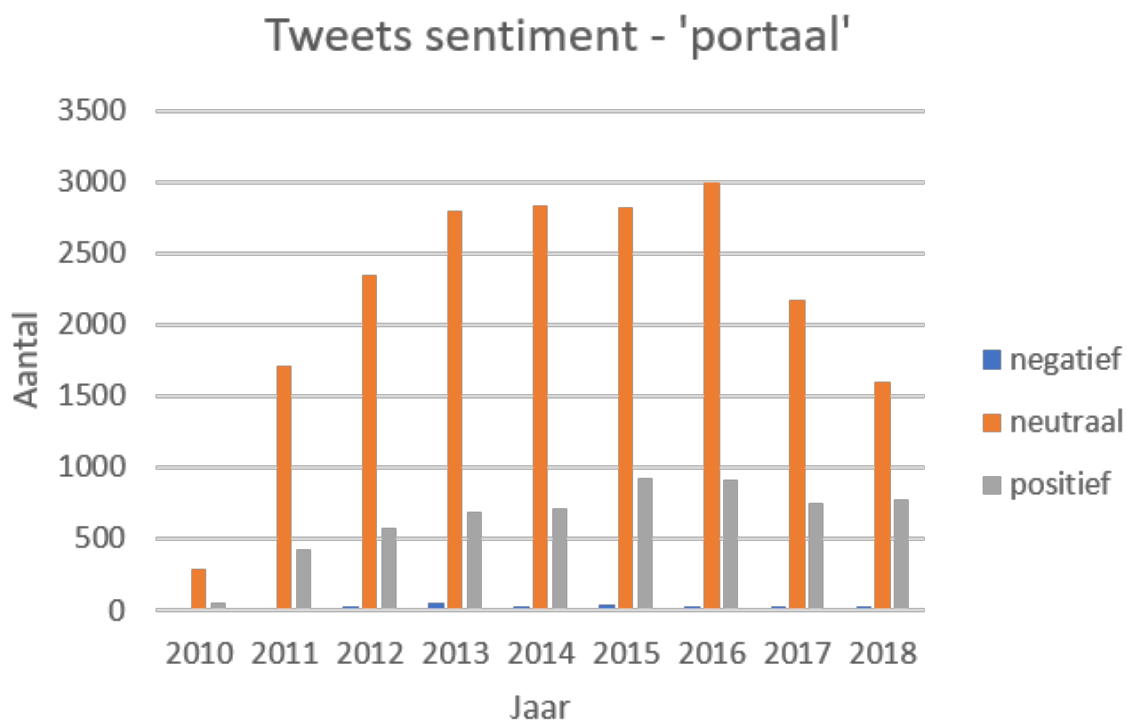
Figuur 26: Tweet sentiment - 'gebruik' 2010 t/m 2018 (527 tweets)

2. portaal (25.628 tweets)

Het merendeel van de verzamelde Tweets is afkomstig van zoekopdracht met de termen portaal en open data. Hiermee zijn 25.628 Tweets verzameld. Slechts een klein deel van de Tweets is hierbij negatief en het merendeel is niet gepolariseerd. Ook ongeveer een kwart van de Tweets is positief. Over de jaren heen is er een vrij consistent beeld van de sentimentswaarde. Het aantal Tweets is de laatste jaren afgenomen rond dit onderwerp. Dit zou kunnen komen doordat open data portals een bepaalde mate van volwassenheid hebben bereikt, en meer onderdeel van dagelijks leven geworden.



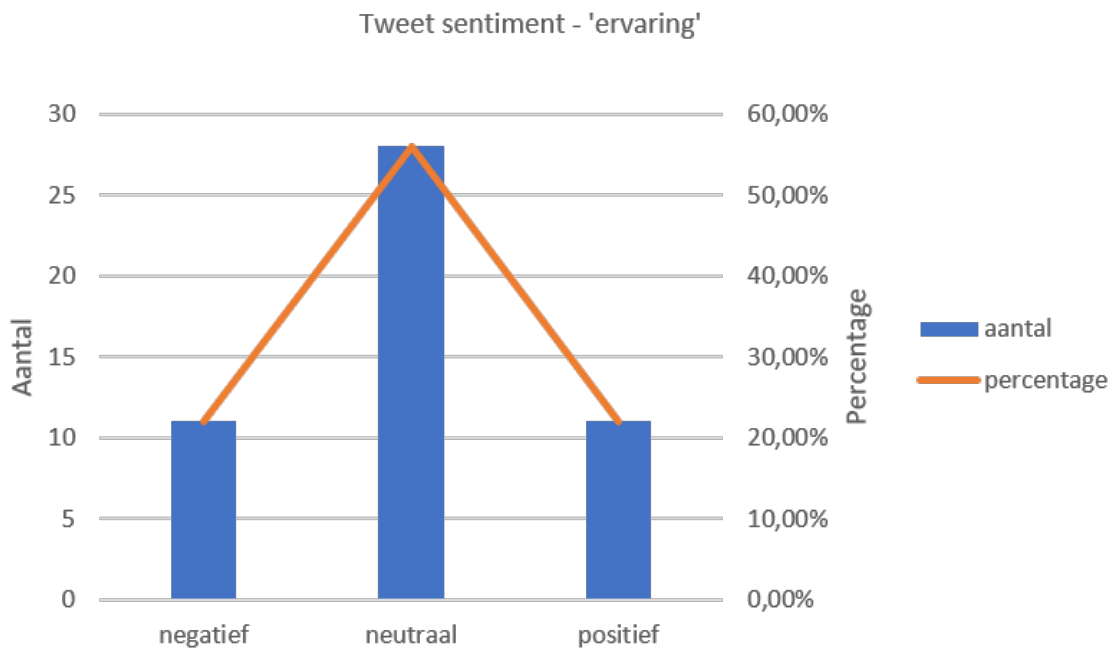
Figuur 27: Tweet sentiment - 'portaal' aantallen en als percentage 2010 t/m 2018 (25.628 tweets)



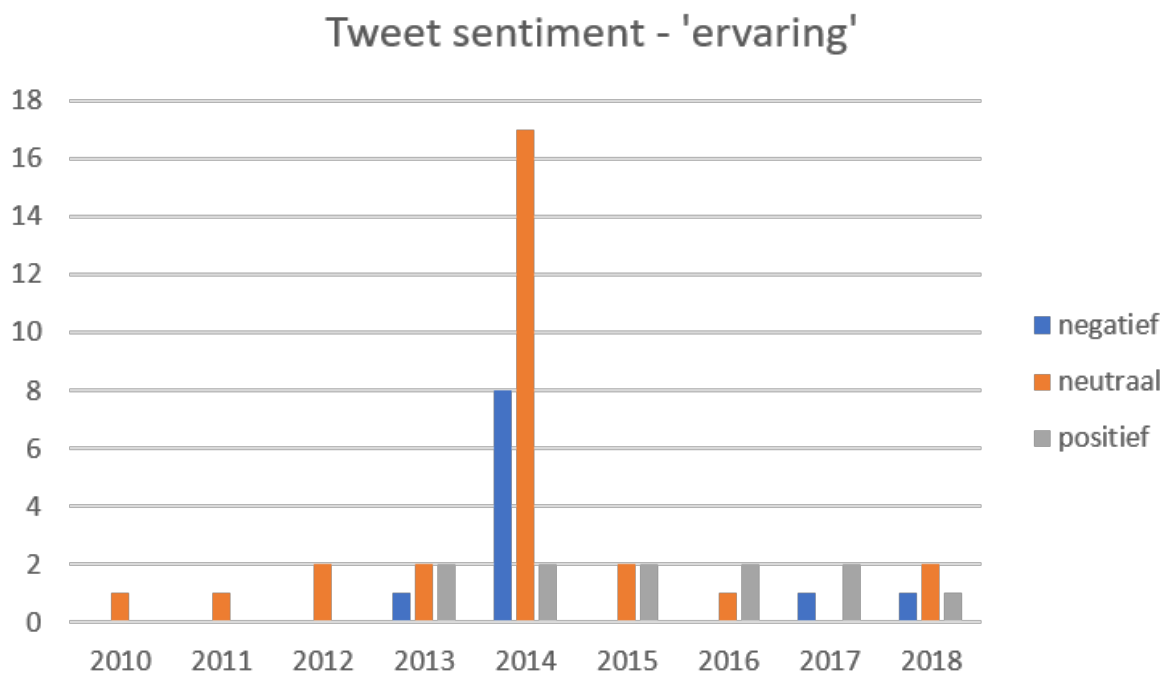
Figuur 28: Tweet sentiment - 'portaal' 2010 t/m 2018 (25.628 tweets)

3. ervaring (50 tweets)

Zoekopdrachten rondom het onderwerp ervaring en open data leverden slechts enkele Tweets op. Hierbij was de sentimentswaarde verdeeld in positieve en negatieve Tweets. Rondom 2014 waren de meeste Tweets verstuurd rondom dit onderwerp.



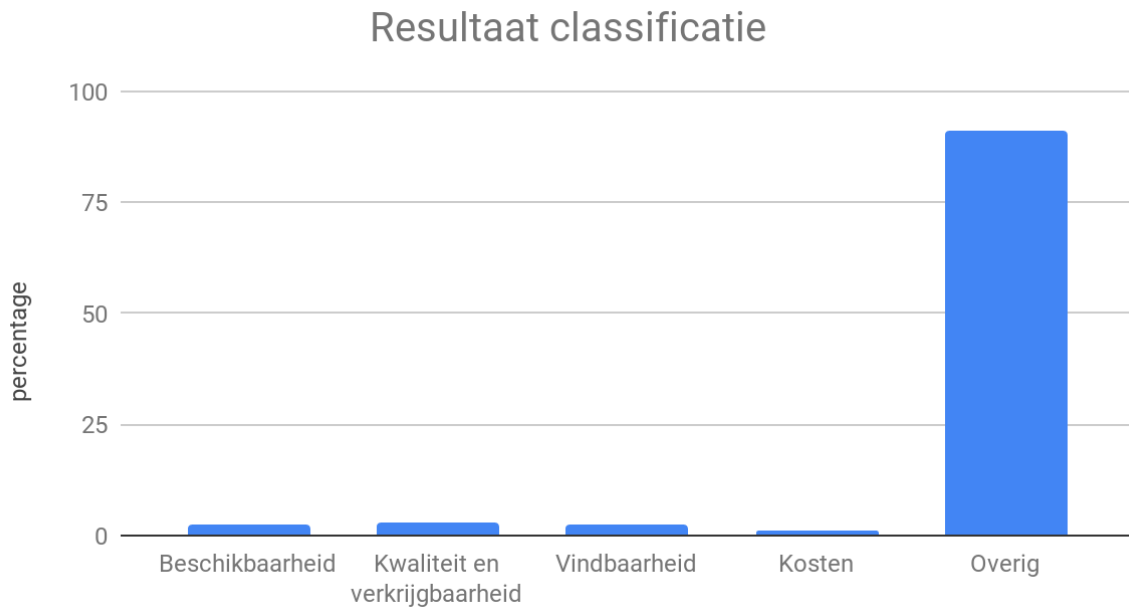
Figuur 29: Tweet sentiment - 'ervaring' aantallen en als percentage 2010 t/m 2018 (50 tweets)



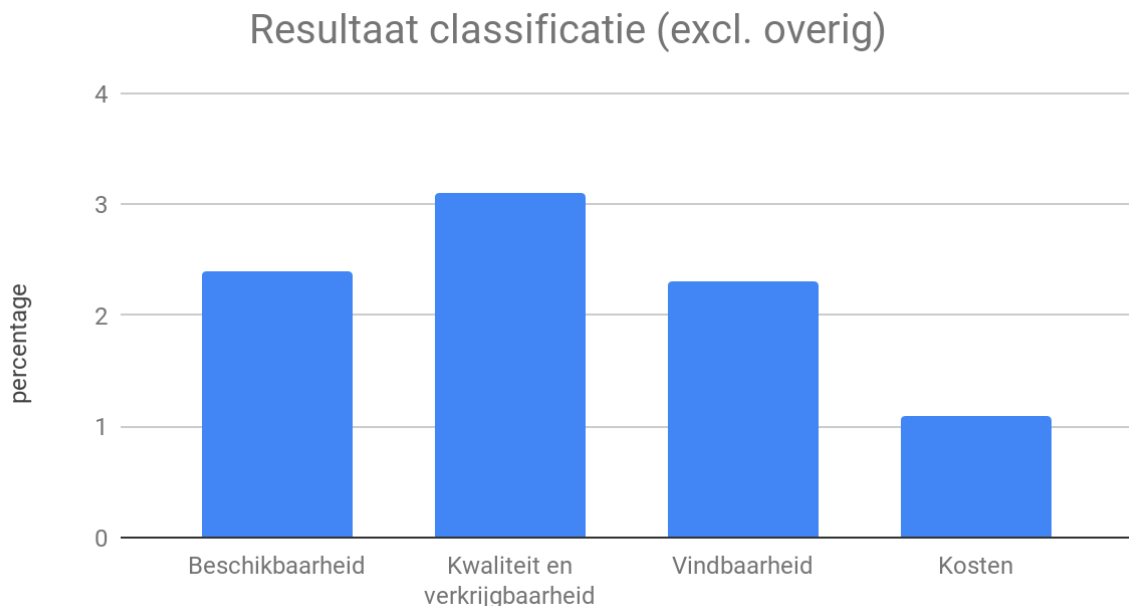
Figuur 30: Tweet sentiment - 'ervaring' 2010 t/m 2018 (50 tweets)

5.2.3 Resultaten classificatie

Figuren 31 en 32 tonen de resultaten van de classificatie van de Tweets, waar in Figuur 31 de ruwe classificatie wordt getoond inclusief de grootste categorie 'Overig', en in Figuur 32 wordt de classificatie zonder de categorie 'Overig' getoond.



Figuur 31: Resultaat classificatie per categorie



Figuur 32: Resultaat classificatie per categorie exclusief 'overig'

Hoewel Figuren 31 en 32 een beeld scheppen, kan men geen conclusies verbinden aan deze resultaten aangezien het slechts een bijzonder klein aantal tweets betreft.

5.2.4 Conclusies social media analyse

Twitter biedt een grote bron van data waarmee social media analyses kunnen worden uitgevoerd. Door middel van een sentimentanalyse is een inschatting gemaakt van het open data sentiment. Uit de resultaten bleek dat het merendeel van de Tweets over open data een neutrale sentimentele waarde hebben. Slechts een klein deel van de Tweets is geclassificeerd als een negatief bericht. Een mogelijke verklaring is dat Twittergebruikers voornamelijk berichten delen met een specifieke vraag of bedoeld zijn ter kennisgeving, bijvoorbeeld een nieuwtje. Verdere analyse op de inhoud van de

Tweets wees uit dat slechts een beperkt percentage van de Tweets inhoud betrof die ingedeeld kon worden in een van de hergebruiksbarrière categorieën. Het meeste commentaar dat gerelateerd kon worden aan eventuele barrières betrof de categorieën beschikbaarheid en datakwaliteit en verkrijgbaarheid. Tweets rondom vindbaarheid zijn slechts enkele keren gevonden en nog minder Tweets betroffen berichten over de (te) hoge kosten van open data.

Er zijn echter beperkingen aan dit onderzoek. Ten eerste kunnen social media data een scheef beeld geven over de werkelijke situatie. Verder kan er sprake zijn van een populatie met veel ruis waardoor het sentiment van een beperkte groep aangezien kan worden voor het sentiment van de samenleving. Een voorbeeld van ruis is bijvoorbeeld het gebruik van cynisme/ sarcasme en het gebruik van emoticones.³⁶ Verder hebben we maar één bron van social media onderzocht, en dan alleen het publieke deel van Twitter. Andere data waren niet beschikbaar voor dit onderzoek. Een andere beperking is dat de populariteit van Twitter niet onder alle open data hergebruikers even groot is. De geïnterviewde hergebruikers gaven aan Twitter nog maar weinig (meer) te gebruiken om het melden van problemen of om positieve of negatieve tweets te versturen. Zij gaven aan Twitter voornamelijk nog te gebruiken voor het melden van nieuwe toepassingen of aankondigingen van activiteiten.

5.3 Waar lopen hergebruikers tegenaan bij het gebruiken van open data?

Om Vraag 3 verder te beantwoorden (tegen welke belemmeringen lopen hergebruikers aan), hebben we een kwalitatieve analyse uitgevoerd, bestaande uit een online vragenlijst, een quick-scan peiling tijdens een workshop met wetenschappelijke medewerkers, een ronde van diepte-interviews, en een online validatie gehouden onder de geïnterviewde hergebruikers.

Hieronder zullen wij de uitkomsten van deze kwalitatieve analyse gezamenlijk beschrijven.

5.3.1 Deelvraag 3a. Hoe ervaren zij de kwaliteit van de aangeboden data?

Als men naar datakwaliteit van de open data in Nederland in algemene zin vraagt, wordt er vaak positief gereageerd. Twee-derde van de respondenten van de vragenlijst ervaren de algemene kwaliteit van open data als afdoende, maar met ruimte voor verbeteringen. Slechts een klein deel gaf aan dat de datasets niet goed tot slecht waren. Zowel in de vragenlijst als tijdens de interviews werd aangegeven aan dat er in de afgelopen jaren al veel stappen zijn gemaakt om datakwaliteit te verbeteren. Slechts een klein deel gaf aan dat de datasets niet goed tot slecht waren. De not-for-profit organisaties en semioverheden waren het meest kritisch op de kwaliteit van open data, terwijl juist het bedrijfsleven de data afdoende tot goed beoordeelden. Bij het doorvragen van specifieke datakwaliteitsaspecten blijkt zowel uit de enquête als uit de vraagg gesprekken dat in de beleving van hergebruikers bepaalde kwaliteitsaspecten van open data nog te wensen overlaat. Dit wordt deels veroorzaakt omdat elke (her)gebruikers zijn/ haar eigen wensen en behoeften heeft, afhankelijk van hun ervaring, kennis en de door hun gebruikte software. Verder kwamen er een aantal aandachtspunten naar boven met betrekking tot belemmeringen die direct met de datakwaliteit verband houden. Wij zullen deze aandachtspunten hieronder per stuk toelichten. We merken bij dit punt wel op dat een aantal bedrijven hun businessmodel volledig hebben afgestemd op de beschikbare open data; zonder open data hadden deze bedrijven niet bestaan. Verder kunnen wij opmerken dat belemmeringen die hergebruikers ervaren, sterk afhankelijk is van de mate van open data volwassenheid van de hergebruiker. Minder ervaren hergebruikers lijken meer belemmeringen te ervaren dan meer ervaren hergebruikers.

³⁶ Zie o.m. van Loenen et al. (2018)

5.3.2 Deelvraag 3b: Welke belemmeringen ervaren open data hergebruikers?

Zowel in de vragenlijst als tijdens de workshop en de interviews gaven hergebruikers aan dat ze een scala aan belemmeringen ondervinden. De meest genoemde belemmeringen zijn: de vindbaarheid van datasets, gebrekkige actualiteit, de structuur en semantiek van de data, gebrek aan standaarden, te geaggregeerde data, onbruikbare bestandsformaten, en een incomplete dekkinggraad. Hierbij was het opvallend dat in de enquête de overheden (zowel nationaal als lokaal) geen belemmeringen ondervonden met het vinden van datasets. Het lijkt dat vertegenwoordigers van de overheid open data beter weten te vinden dan ondernemers en burgers. Maar ook hergebruikers binnen de overheid ervaren problemen met de kwaliteit van open data. Hieronder worden de voornaamste belemmeringen beschreven.

5.3.2.1 Belemmeringen met de datasets

Bekendheid

Vrijwel alle geïnterviewde hergebruikers noemden bekendheid, d.w.z. kenbaarheid en vindbaarheid als een van de grootste belemmeringen:

- Kenbaarheid: veel datasets hebben geen duidelijke namen; als je niet weet hoe een bestand heet, kan je ook niet vinden
- Vindbaarheid: versnipperde vindplaatsen: veel data (van lokale overheden) worden niet via een centraal dataplatform aangeboden, en zijn daardoor datasets moeilijk te vinden (vindbaarheid). Er is geen verplichting voor lokale overheden om datasets bij data.overheid.nl aan te melden, waardoor beschikbare datasets van lokale overheden niet zijn doorgelinkt naar data.overheid.nl.

Actualiteit

De actualiteit (update frequentie) van de datasets laat te wensen over: sommige datasets worden eenmalig aangeboden en daarna niet meer geüpdatet als open data versie, vooral data die eenmalig voor een project of een hackathon beschikbaar zijn gesteld. Bij andere datasets is het onduidelijk wat de update frequentie is, en/of wanneer de volgende update verwacht kan worden. Daarnaast wordt in de metadata niet altijd vermeld wanneer de data daadwerkelijk zijn ingewonnen.

Sommige geïnterviewde hergebruikers gaven aan dat data pas relatief laat beschikbaar worden gesteld via data.overheid.nl hoewel die data via andere kanalen al beschikbaar zijn. Voor data die alleen waarde hebben als actuele data is dat een gemiste kans. Dit geldt niet alleen voor real-time data, maar ook voor datasets die per maand of per jaar door de overheid worden bijgewerkt. Ook kan er een tijd zitten tussen de actualisatiedatum van een dataset en het beschikbaar stellen van die data als een open data-versie. Verder gaven een aantal hergebruikers aan dat actualisatiedata niet synchroon loopt met peildata voor de industrie, bijv. data die nodig zijn voor subsidieaanvragen worden pas een week voor de deadline van de aanvraag als open data beschikbaar gesteld.

Structuur en semantiek

Een veelgehoorde klacht (door meerdere hergebruikers genoemd) is dat de datasets vaak een duidelijke structuur en/of beschrijving van de structuur ontbreken. Als voorbeeld werd door een aantal geïnterviewde hergebruikers gegeven dat:

- er informatie ontbreekt over welke attributen er in de data aanwezig zijn
- datamodellen worden niet beschreven, waardoor het moeilijk te achterhalen is waarom de datamodellen tot bepaalde resultaten komen
- onduidelijk structuur van de labels van de kolommen in een CSV-bestand, of wat die labels betekenen (vaak een voor hen nietszeggende afkorting of specifieke term).

- een gebrek aan eenduidige semantiek:
 - o de labels in de bestanden komen niet overeen met de benaming die door andere organisaties wordt gebruikt, bijv. niet dezelfde namen die door CBS of door andere gemeenten worden gebruikt.
 - o soms zelfs foutieve / verkeerde benamingen

Bovendien gebeurt het dat de structuur in updates wordt aangepast (kolommen gesplitst / extra kolommen toegevoegd, of juist weggehaald of hernoemd) zonder dat dat in de bijbehorende documentatie wordt beschreven. Voor hergebruikers vergt dat veel tijd om de data elke keer weer op te schonen en/of te herstructureren, vooral als er extra kolommen zijn toegevoegd. Dat maakt verder vergelijkingen met voorgaande versies erg moeilijk. In ons eigen kwantitatieve onderzoek liepen wij tegen vergelijkbare problemen aan van onduidelijke structuur die met updates veranderde waardoor het opschonen van de aangeleverde data meer tijd in beslag nam dan in voorgaande onderzoeken.

Gebrek aan standaardisatie

Gebrek aan standaardisatie van lokale overheidsdata/ gemeentelijke data met betrekking tot bijvoorbeeld gebruikte dataformaten en structuur. Op nationaal niveau is standaardisatie redelijk goed maar niet op lokaal niveau, zeker als die standaardisatie per jaar verandert. Verder lijkt er weinig coördinatie te zijn tussen gemeenten en tussen provincies en/of waterschappen m.b.t. standaardisatie. Voor dezelfde objecten worden door verschillende organisaties bijvoorbeeld verschillende kleuren gebruikt. Verder zijn overheidsorganisaties nu bezig eigen API's te ontwikkelen in plaats van gebruik te maken van bestaande platforms/ API's (al dan niet door commerciële partijen ontwikkeld). Verder lijkt er een gebrek aan coördinatie te zijn op het gebied van welke (gelijksoortige) datasets door lokale overheden beschikbaar worden gesteld. Nu wordt die afweging door elke lokale overheid gemaakt en niet alle lokale overheden stellen dezelfde typen open data beschikbaar, of beschikbaar in vergelijkbare formaten. Hierdoor kunnen er geen complete land dekkende / regio dekkende datasets worden samengesteld omdat er gaten in de dekkingsgraad zitten.

Te geaggregeerde data/ te laag detailniveau

Hiermee samenhangend werd als belemmering genoemd dat er een gebrek aan is aan standaard regels op gebied van anonimiseren en/of aggregeren van data. Elke overheidsorganisatie maakt een eigen overweging wat wel/niet toelaatbaar is in het kader van de Algemene Verordening Gegevensbescherming (AVG) (bescherming van persoonsgegevens) of op het gebied van het beschermen van concurrentiegevoelige data van bedrijven. Die afweging op lokaal niveau heeft niet alleen betrekking op welke datasets wel of niet als open data beschikbaar worden gesteld, maar ook op welk niveau open datasets worden bewerkt om geschikt te maken voor publicatie als open data. Sommige data-aanbieders aggregeren de data op geografisch niveau, bijvoorbeeld data op wijkniveau. Andere aanbieders aggregeren data door kolommen met gegevens die herleidbaar zouden kunnen zijn naar natuurlijke personen, geheel te verwijderen. Vooral voor data die als download data beschikbaar zijn, en waar die afwegingen vooraf worden gemaakt, worden volgens een aantal hergebruikers dergelijke afwegingen heel erg aan de voorzichtige kant gemaakt. Door deze mate van (over)aggregatie gaat veel detailniveau verloren die een aantal hergebruikers wel zouden willen hebben.

Niet geschikte bestandsformaten

Niet geschikte bestandsformaten is ook vaak genoemd als barrière. Een voorbeeld is een document die als open data-versie alleen in pdf wordt aangeboden terwijl er ook een origineel document in een Word- of Excelversie beschikbaar zou moeten zijn. Dat zijn weliswaar geen open formaten, maar wel formaten waarin documenten verwerkt kunnen worden (machine-verwerkbaar). Hergebruikers moeten momenteel contact met de data-aanbieders opnemen om de dataset in het door hen gewenste formaat op te vragen omdat de pdf-versie niet machine-leesbaar is en de structuur niet is

te herkennen. Niet alle ambtenaren zijn zich ervan bewust van de beperkingen van pdf-formaat voor hergebruikers. Hetzelfde geldt ook voor open geo-formaten: een aantal hergebruikers hebben aangegeven liever de (oorspronkelijke) shapefile te willen gebruiken dan de aangeboden GML bestanden. Ook zijn niet alle open data bestandformaten makkelijk te begrijpen voor niet-specialisten, bijv. DATEX II-formaat voor verkeersdata³⁷ is een complex formaat.

Fouten in data en gaten in de data

Een aantal van de hergebruikers gaven aan dat er fouten in de datasets kunnen zitten zoals spelfouten, foutieve georeferencing of geheel ontbrekende georeferencing in de dataset. Een andere belemmering is dat locatie referentiesystemen achter kan lopen bij real-time data.

Een andere gehoorde belemmering is dat er niet alleen gaten in de dekkingsgraad van de data zitten, maar ook gaten in de datasets zelf. Voor niet-overheid hergebruikers is het moeilijk om een directe terugmelding van fouten te maken omdat er geen faciliteiten daarvoor beschikbaar zijn, of erg moeilijk te vinden (vaak op een andere pagina). En als er een terugmelding wordt gemaakt, bijvoorbeeld via e-mail, dan duurt het lang voordat er gereageerd wordt op de melding.

5.3.2.2 Belemmeringen met de datadiensten

Naast belemmeringen met de data werden ook belemmeringen aangegeven met de manier waarop data worden aangeboden.

Kwaliteit van de dienstverlening: Storingen in servers/ API's

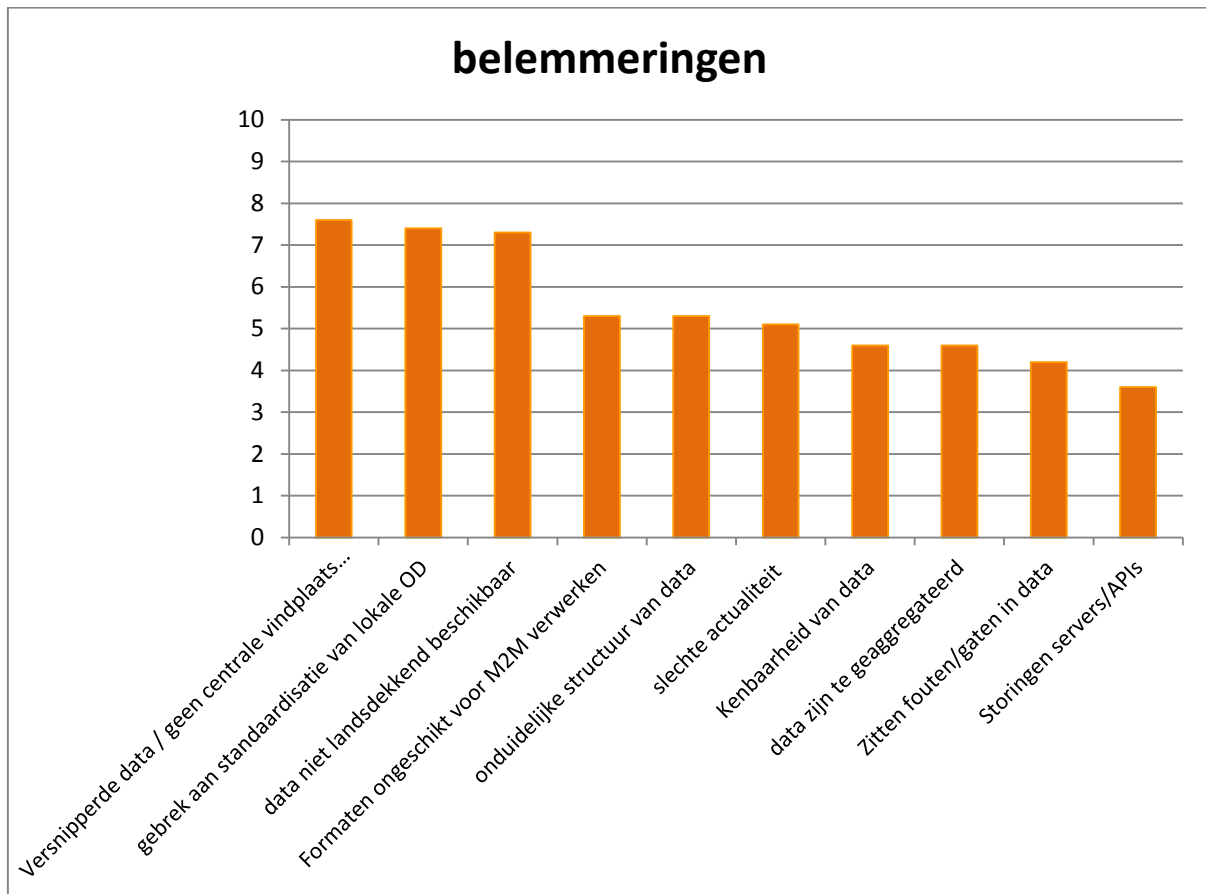
Een aantal hergebruikers dat vooral gebruik maken van (near) real-time data, vinden het een groot probleem wanneer de servers en/of API's er uit liggen. Voor hergebruikers van real-time data voor diensten zoals verkeersapps of ov-apps is het uitvallen van servers zeer nadelig. Voor data waarbij de actualiteit minder belangrijk is, kunnen hergebruikers terugvallen op data downloads met oudere data. De door de hergebruikers ontwikkelde diensten bieden dan geen actuele data aan maar dat is beter dan helemaal geen data aanbieden. Verder werd ook duidelijk gemaakt dat er vaak niet adequaat wordt gereageerd op meldingen van storingen, vooral wanneer een dergelijke storing op vrijdagmiddag plaats vindt. Verder zijn trage servers (vooral bij lokale overheden) een probleem.

Geen filteropties voor data

Voor sommige hergebruikers zijn de aangeboden download bestanden te groot om te kunnen verwerken met standaard software, bijv. te veel regels in een spreadsheet. Daardoor kunnen er data verloren gaan voor hergebruikers. Andere hergebruikers gaven aan dat zij slechts een deel van de attributen van de totale dataset nodig hebben, maar dat er geen filter mogelijkheden beschikbaar zijn.

Als laatste onderdeel van ons onderzoek hebben wij aan de geïnterviewde hergebruikers gevraagd om via een online evaluatie aan te geven welke van deze aspecten als grootste belemmeringen werden ervaren. In deze online evaluatie hebben wij gevraagd om de 10 meest genoemde belemmeringen een waarde toe te kennen van 1 tot 10, waarin 1 weinig prioriteit voor de hergebruiker had, en een 10 een hoge prioriteit had voor de hergebruiker. Uit deze prioritering kwam naar voren dat de versnipperde data / geen centrale vindplaats voor open data van lokale overheden, gebrek aan standaardisatie van open data van lokale overheden, en open data van lokale overheden niet beschikbaar als land dekkende data als de grootste belemmeringen werden ervaren, zie Figuur 33.

³⁷ DATEX II is de Europese standaard voor het uitwisselen van verkeersgegevens en reisinformatie in intelligente verkeerssystemen.



Figuur 33: prioritering van de belemmeringen voor open data hergebruikers (juli 2019)

5.3.3 Deelvraag 3c: Welke verbetermogelijkheden zien hergebruikers?

In zowel de vragenlijsten als tijdens de interviews gaven hergebruikers aan dat mogelijke verbeterpunten zijn om voor de manier waarop open data worden aangeboden. Deze aanbevelingen vloeien direct voort uit de belemmeringen die ondervonden worden. Zo werden meer actuele data, meer data op hoger detailniveau, een aantal keren genoemd, net als data sneller beschikbaar stellen. En duidelijker aangeven in de metadata wat de frequentie van data-updates is.

Er werden ook verbetermogelijkheden met betrekking tot de governance van open data gegeven door de hergebruikers. De meest genoemde waren:

- Er is meer regie en coördinatie nodig op de datastromen. De overheid zou meer gebruik kunnen maken van de kennis die er al is bij de hergebruiker via groepen en specialisten uit koepelorganisaties, bijvoorbeeld ICT Nederland of GeoBusiness Nederland.
- Datavalidatie onderdeel maken van open data governance. Validatiechecks zijn dan ingebakken in het open data proces.
- Data via verplichte standaarden publiceren, inclusief gestandaardiseerde en consistente codering.
- Meer integratie tussen dataplatforms bijvoorbeeld tussen data.overheid.nl en PDOK.
- Bindende afspraken met alle lagen van overheid en overheidsbedrijven over welke datasets ontsloten moeten worden.

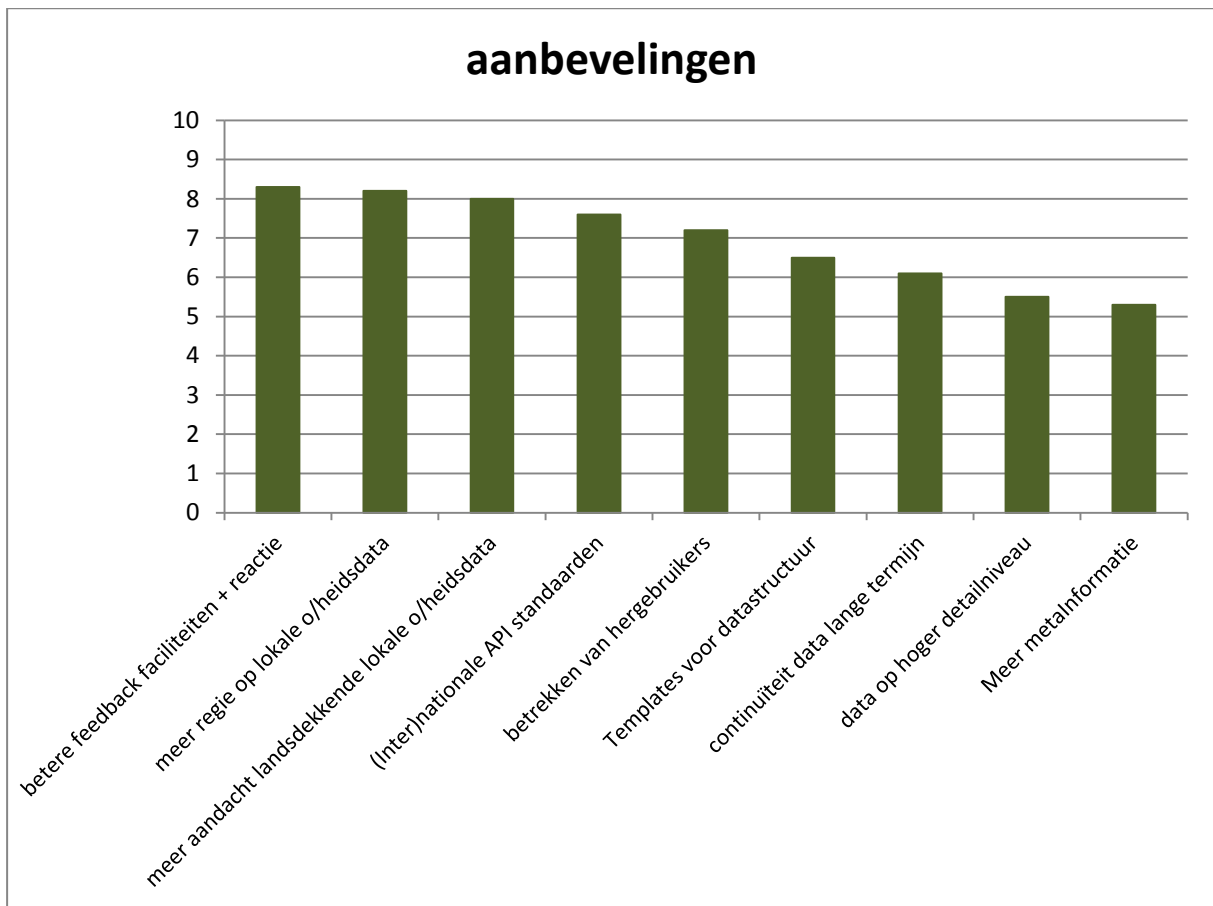
Verder werden er verbetermogelijkheden met betrekking tot wat de overheid zou kunnen doen om een betere afstemming tussen de overheid en de hergebruikers. Deze waren:

- Maak meer gebruik van feedback van gebruikers, en dat ook tijdig (sneller) terugkoppelen aan die gebruikers. Ook duidelijker maken aan andere hergebruikers dat er al meldingen zijn geweest van diezelfde fout/ ontbrekende data. En niet alleen een melding maken dat de feedback is geregistreerd, maar ook sneller de gemelde fouten oplossen, vooral als die betrekking heeft op servers die eruit liggen (een melding op vrijdagmiddag niet pas op maandagochtend afhandelen).
- Zorg dat er feedbackfaciliteiten op de webpagina's zijn waar de data beschikbaar zijn zodat een gebruiker niet de pagina hoeft te verlaten om een melding te maken.
- Zet als open data aanbieder wat vaker de pet van de burger/ bedrijfjes op. Ga eens langs bij een aantal van de bekende hergebruikers. Kleine ondernemers hebben geen tijd om bij Open Data Gebruikers bijeenkomsten langs te gaan (kost ten minste een halve dag inkomsten), dus: zoek als aanbieder actief de hergebruiker op.
- Betrek meer open source ontwikkelaars die tools/ toepassingen maken om (complexere) overheidsdata inzichtelijk te maken.

Verbetersuggesties van de hergebruikers met betrekking tot datatoepassingen waren:

- Maak een overzicht van de verrijkingen van open data die door anderen zijn gedaan.
- Maak gebruik maken van bestaande internationale standaarden voor API's i.p.v. eigen API-standaarden te ontwikkelen.
- Besteed meer aandacht aan de continuïteit en bruikbaarheid van data(toepassingen). Veel promotie van innovatie door de overheid is nu gebaseerd op functionaliteit (wat kunnen we doen met welke data?). Maar wat gebeurt er na de hackathon/ project / looptijd van de subsidie?
- Denk meer vanuit maatschappelijke vraagstukken en welke data daarbij bij kunnen dragen aan oplossingen. Bijvoorbeeld, hoe gaan we om met de vergrijzing van de bevolking in Nederland? Of, Hoe gaan we statushouders aan werk helpen waar nu een tekort aan is? Daarvoor zijn meer demografische data nodig over bijv. welke leeftijdscategorieën nog werkzaam zijn en in welke sectoren voor welke salarissen en in welke plaatsen zodat er betere afstemming gemaakt kan worden tussen werkzoekenden en vacatures. Zo maakt UWV sinds eind 2018 informatie over werkzoekenden en geanonimiseerde cv's beschikbaar als open data. Maar hergebruikers geven aan dat die benodigde data in een te geaggregeerd formaat worden aangeboden om echt bruikbaar te zijn, en worden (nog) niet geactualiseerd.
- Houd eigen monitoring-/meetsystemen (bijv. CO₂, waterkwaliteit) in de gaten en bedenk hoe eigen data (en die van derden) bij zouden kunnen dragen om dergelijke monitorprocessen te verbeteren.

Als laatste onderdeel van ons onderzoek hebben wij aan de geïnterviewde hergebruikers gevraagd via een online evaluatie aan te geven welke van deze aanbevelingen de hoogste prioriteit zou moeten krijgen, waarin 1 een zeer lage prioriteit had voor de hergebruiker, en een 10 de hoogste prioriteit in de optiek van de geïnterviewde. Uit deze evaluatie kwam naar voren dat betere feedback faciliteiten en een tijdige reactie op de geleverde feedback, meer regie op het beschikbaar stellen van open data van lokale overheden en die data land dekkend te maken, en gebruik van (inter)nationaal herkende standaarden voor API's als grootse prioriteiten naar boven kwam, zie Figuur 34.



Figuur 34: prioritering van aanbevelingen voor de overheid van hergebruikers (juli 2019)

Een interessante suggestie was om de overheid te verplichten zelf alleen data te gebruiken die via één centraal dataportaal worden aangeboden. Dan zouden overheidsgebruikers tegen dezelfde belemmeringen aanlopen als hergebruikers buiten de overheid. En daarmee samenhangend zou de private sector meer betrokken willen worden bij de ontwikkeling van een dergelijk centraal dataportaal. Overigens waren circa een derde van de respondenten al meer of minder betrokken bij het open data proces, hetzij als deelnemer in de Open Data Gebruikersgroep, als datahouder/ beleidsmaker of als ontwikkelaar in opdracht van de overheid.

Het is ook interessant om te zien dat er door hergebruikers een aantal aanbevelingen worden gemaakt die al zijn geïmplementeerd. Zo geeft de huidige versie van data.overheid.nl via 'Impact' een overzicht van een aantal toepassingen gebaseerd op open data. Verder is één van de ambities die in NL DIGITAAL worden uitgesproken juist om meer datagedreven aan de slag te gaan met maatschappelijke vraagstukken.

6 Vraag 4: Wat zijn behoeften van de hergebruikers?

6.1 Deelvraag 4a: Welke behoefte hebben hergebruikers m.b.t. het aanbod van open data?

Hergebruikers hebben vooral behoefte aan meer data van lokale overheden en betere vindbaarheid van die datasets. Er ontbreken nu veel datasets van gemeenten en van provincies die wel door andere lokale overheden als open data worden verstrekt. Daardoor zijn data niet land/ regio dekkend beschikbaar.

In de enquête gaven een aantal hergebruikers aan behoefte te hebben aan meer verschillende soorten data (“alle data die de overheid zelf ook gebruikt”) en aan geautomatiseerde processen om data sneller te laten ontsluiten als open data. Vooral hergebruikers voor maatschappelijke producten en/of diensten maken gebruik van alle mogelijke manieren om aan data te komen. Deze categorie maakt ook het meest gebruik van dataverzoeken in kader van de Wet openbaarheid van bestuur (Wob), zogenaamde Wob-verzoeken, vergeleken met andere categorieën hergebruikers. Maar ook hergebruikers voor commerciële producten/ diensten maken nog steeds gebruik van Wob-verzoeken. Dit zou een indicatie kunnen zijn dat niet alle data actief beschikbaar wordt gesteld als open data of dat de data niet gevonden kunnen worden.

Meer actuelere data

Verder hebben zij behoefte aan meer actuele data, onder andere is er behoefte aan actuelere statistische data, maar ook van snelheidsprofielen, en geografische informatie. Maar gebruikers hebben ook behoefte aan de oudere versies zodat er vergelijkingen / tijdreeksen gemaakt kunnen worden. Daarmee samenhangend willen een aantal hergebruikers dat de release datum van periodieke updates beter afgestemd worden met andere overheidsprocessen, bijv. financiële data in december actualiseren terwijl de indiendata voor subsidies (die dergelijke data nodig hebben voor de aanvragen) 1 januari is.

6.2 Deelvraag 4b: Welke behoefte hebben hergebruikers m.b.t. de manier waarop data worden aangeboden?

Uit de kwalitatieve analyse kwamen de volgende behoeften van open data hergebruikers naar voren met betrekking tot de manier waarop data worden aangeboden.

Meer maatwerkopties voor detailniveau/ benodigde attributen van datasets

Hergebruikers hebben daarnaast behoefte aan data met een hoger detailniveau. De hergebruikers realiseren zich dat persoonsgegevens beschermd moeten worden en dat daarvoor een zekere mate van aggregatie nodig is om datasets te anonimiseren. Maar de hergebruikers gaven aan niet zo zeer geïnteresseerd te zijn in persoonsgegevens op zich, maar wel in een hoger detailniveau. Er zou gekeken moeten worden of er geen alternatieve mogelijkheden zijn om meer gedetailleerdere data te kunnen krijgen, zonder persoonsgegevens. Desnoods zou een hergebruiker die een doelbinding voor een bepaalde dataset (die nu niet als open data beschikbaar zijn) aan kan tonen, bijvoorbeeld voor een maatschappelijk doel, die dataset zou kunnen krijgen met een gebruikersovereenkomst. De hergebruiker moet dan ervoor zorgen dat de afgeleide dataproducten/-diensten niet herleidbaar zijn naar natuurlijke personen. Andere gebruikers hebben juist behoefte aan minder gedetailleerde, en minder complexe datasets. Zij zouden graag van bepaalde datasets slechts enkele attributen willen hebben in plaats van de hele dataset.

Meer verschillende formaten en data diensten

Verder zouden data in meerdere formaten moeten worden aangeboden. Voor de een is een open source formaat meer geschikt, een ander wil liever een propriety formaat omdat de data dan makkelijker te verwerken zijn. Zo ondersteunen CMS'en (content management systemen) geen uitlevering in CSV-formaat. Ook ontbreekt georeferencing in de aangeboden open source formaten die wel in de oorspronkelijke shapefiles (Esri propriety GIS-software) zit.

Data zouden zowel via API's als via datadumps beschikbaar gesteld moeten worden. Maar dan moeten die API's wel volgens internationaal erkende protocollen/ standaarden zijn ontwikkeld. Er wordt steeds meer data via API's ontsloten maar (vooral lokale) overheden gebruiken daarvoor eigen API's die voor hergebruikers niet handig zijn of gemakkelijk zijn te begrijpen. Regelmatige afnemers gebruiken vaker API's maar voor eenmalige/ onregelmatige gebruikers kost het te veel tijd om uit te zoeken hoe de API's werken. De API van KNMI werden door verschillende geïnterviewde hergebruikers als een goed voorbeeld genoemd.

Daarmee samenhangend zouden hergebruikers per request via API's meer gedetailleerde/ minder geaggregeerde data kunnen krijgen. Privacy-checks moeten dan wel technisch worden ingericht om bijvoorbeeld scraping te vermijden.

Standaard datastructuren

Verder werd aangegeven dat er een behoefte is aan het gebruik van standaard templates voor gelijksoortige documenten en eenduidige semantiek om patronen in datastructuren te herkennen.

Metadata

Hergebruikers hebben behoefte aan uitgebreidere metadata en/of aanvullende documentatie, en die zouden in meerderde talen beschikbaar gesteld moeten worden. Een aantal van de geïnterviewde gebruikers gaven aan dat als zij gebruik maken van datasets uit andere landen, de kans groot is dat er hergebruikers uit het buitenland ook gebruik zullen maken van Nederlandse datasets. Dit is bevestigd in onze kwantitatieve analyse van de open data portalen beschreven in Hoofdstuk 3. Bovendien zijn niet alle hergebruikers in Nederland de Nederlandse taal machtig, zoals bleek uit de discussie tijdens de workshop van 8 maart met internationale studenten en wetenschappers. Zij gebruiken vooral data via CBS waar een Engelstalige versie beschikbaar is, of geodata via PDOK waarbij zij hun wetenschappelijk netwerk gebruiken om de relevante datasets te herkennen.

In de metadata en/of aanvullende documentatie zou moeten worden meegeven:

- Indicatie hoe de dataverstrekker zelf de kwaliteit van de dataset beschrijft. Dus niet data 'as is' beschikbaar stellen, maar meer detail geven waarvoor de dataset oorspronkelijk voor werd gebruikt en welke bewerkingen er zijn uitgevoerd met de dataset om als open data beschikbaar te stellen.
- Beschrijving hoe de data tot stand zijn gekomen
- Beschrijving wat de velden betekenen
- Als er kolommen zijn toegevoegd/ verwijderd, beschrijving van welke kolommen dan zijn gesplitst en/of weggehaald, en waarom
- Als de data betrekking hebben op data met codes die bij andere organisaties worden gebruikt (bijv. ISCO-beroepscode of CBS-codes voor wijken), dan die codes ook meegeven in de labels van de dataset zelf.

Betrek de gebruiker in het open data proces met feedback faciliteiten

Een suggestie was om een platform te creëren met een "ontiegelijk eenvoudige feedbackmogelijkheid": de overheid als platform dus. Creëer een feedback knop op de download site zodat er meteen feedback gegeven kan worden zonder de site te verlaten, een terugkoppelingknop dus. . Op moment dat je de data hebt gevonden: duidelijk dat er feedback mogelijk is, na gebruik van

de download is een gebruiker weg van het contactmoment. Het terugmeldsysteem verbeterdekaart.kadaster.nl werd als voorbeeld genoemd. De suggesties waren:

- Reageer als overheid binnen redelijke tijd (dus niet na 6 maanden)
- Informeer de gebruiker over wat er met zijn feedback is gedaan, en
- Bedank de gebruiker en/of geef de gebruiker credits hiervoor (plaats op een site, geef mogelijkheid van likes of reacties).
- Zorg dat er een automatische kwaliteitscontrole is: controleer de eigen bestanden (bijvoorbeeld OpenStreetMap) met eigen bestanden (i.c. basisregistratie Topografie).
- Laat overheden feedback geven op elkaars data ter verbetering van de kwaliteit: zoals bij basisregistraties

Voorbeeld feedback win-win-win

Een klant van een open data bedrijf constateert dat er iets niet beschikbaar is: het open data bedrijf maakt een melding bij de bron. Bron lost het binnen een maand op en meldt dit terug. Waardering van het open data bedrijf door klant wordt vergroot. Zie hoe dit gaat met bugs in software.

Win 1: klant heeft actuelere data

Win 2: open data bedrijf heeft beter imago en betere data

Win 3: open data aanbieder heeft actuelere data zonder dat ze zelf op zoek is gegaan naar een verbetering

6.3 Deelvraag 4c. Welke behoefte hebben hergebruikers m.b.t. de vindbaarheid van data?

De meeste respondenten en een aantal van de geïnterviewde hergebruikers gaven aan dat - hoewel er altijd ruimte voor verbetering is en een aantal ook graag meer betrokken zouden worden via klankbordgroepen, online platforms en open data community's, of als ontwikkelaar - de meeste redelijk tevreden zijn met de ingeslagen weg.

Veel hergebruikers gaven aan behoefte te hebben aan één centraal platform waar alle open data van lokale overheden vindbaar zijn. Er ontbreken nu heel veel datasets van gemeenten en van provincies. Dat centrale platform zou aangestuurd moeten worden door een (interbestuurlijk) organisatie die verplichtingen stellen voor gebruik van standaardisatie/ formaten geschikt voor machine-to-machine uitwisseling. Verder zouden er dan randvoorwaarden opgesteld worden m.b.t. welke lokale datasets verplicht als open data beschikbaar gesteld moeten worden en met welke update frequentie.

Een andere behoefte betreft vindbaarheid van datasets via algemene zoekmachines of via de zoekfunctie van open data portalen. Dus niet alleen op trefwoorden kunnen zoeken, maar ook op algemene termen, bijvoorbeeld 'data over wegen'.

Verder hebben hergebruikers behoefte aan duurzame beschikbaarheid van open data. Dat betekent dat (oudere versies van) datasets beschikbaar moeten blijven. Verder zou er meer gebruik moeten worden gemaakt van machine-leesbare persistente identifiers zoals DOI's, en er zou meer gebruik gemaakt moeten worden van URI's in plaats van URL's voor datasets om vindbaarheid te verbeteren, en machine-to-machine verwerking te bevorderen. Het verschuiven van bestanden levert dode links en nog meer uitzoekwerk op.

Een andere suggestie was om bij persberichten van de overheid waar nieuwe datasets/ onderzoeksresultaten bekend worden gemaakt, meteen ook een link naar de betreffende data te plaatsen. Dat scheelt weer uitzoekwerk.

Veel hergebruikers zouden niet meer zonder open data kunnen. Vooral bij de not-for-profit organisaties en bij de micro en kleine ondernemingen is hun businessmodel geheel of grotendeels afhankelijk van open data. Zonder open data zouden een aantal van deze ondernemingen niet bestaan en zouden vele toepassingen niet zijn ontwikkeld. Sommige hergebruikers zouden bereid zijn om te betalen voor data met een hogere kwaliteit/ meer attributen/ lager aggregatieniveau die met een Service Level Agreement worden aangeboden, mits de prijs niet te hoog is. De marges voor diensten gebaseerd op open data zijn erg laag, ook voor de ontwikkelde betaaldiensten, dus de prijs voor data als grondstof is erg belangrijk. Andere hergebruikers krijgen geen keuze omdat niet alle benodigde data beschikbaar zijn als open data, maar alleen als gesloten data tegen een vergoeding.

6.4 Overige behoeften: governance van open data

De governance van open data is volgens respondenten, met name op het lokale niveau, slecht: het aanbod is erg versnipperd en vrijblijvend. De respondenten wensen dat de governance voor open data van lagere overheden centraal wordt geregeld, met een interbestuurlijk orgaan met een duidelijk mandaat aan het roer.

Daarnaast is de behoefte geuit dat open data op een consistente en bestendige manier in een overheidsorganisatie wordt ingericht. Dat betekent dat er zekerheid wordt geboden dat datasets beschikbaar blijven als open data en dat open data in de organisatie ingebed worden. Hergebruikers zijn nu nog erg afhankelijk van welwillende ambtenaren met inhoudelijke kennis. Hergebruikers hebben de indruk dat opendatabeleid in een lagere overheid sterk verbonden is met een dergelijke ambtenaar: wanneer de ambtenaar niet meer bij open data betrokken is, verdwijnt open data binnen de organisatie van de radar.

7 Conclusies en aanbevelingen open data hergebruikers

7.1 Conclusies

Open data worden op grote schaal en door een diverse groep gebruikt voor allerlei doeleinden. Hoewel er verschillende categorieën hergebruikers zijn, is het niet zo dat dit type hergebruiker open data voor een specifiek doel gebruiken. Het bedrijfsleven beperkt zich niet tot hergebruik van open data voor alleen commerciële producten en diensten, en not-for-profit organisaties beperken zich niet tot alleen maatschappelijke diensten. Open data worden zowel gebruikt om producten en diensten voor derden te ontwikkelen als voor interne bedrijfsvoering bij alle categorieën hergebruikers. Uit de kwantitatieve analyse van logfile gegevens blijkt dat vooral overheden, de private sector, de hogere onderwijssector, en het bank- en verzekeringswezen de grootgebruikers zijn van open data. Opvallend is ook dat Nederlandse open data ook door buitenlandse organisaties worden gebruikt. Verder blijkt dat het open data portaal van CBS vaker op afzonderlijke dagen worden gebruikt vergeleken met PDOK en data.overheid.nl. PDOK-datasets worden het meest aangeropen, vooral ook omdat veel van de datasets gebruikt worden in andere diensten zoals het WOZ-waardeloket. Data.overheid.nl lijkt vooral gebruikt te worden als een catalogusdienst voor andere dataplatforms.

Uit de kwalitatieve analyse blijkt dat open data zowel via downloaddiensten als via API's worden hergebruikt. Hergebruikers maken gebruik van meerdere zoekstrategieën en van meerdere platforms om benodigde data te verkrijgen. Het feit dat er nog steeds gebruik wordt gemaakt van verzoeken voor datasets in kader van Wet openbaar bestuur, zogenaamde Wob-verzoeken, bij de data-verstrekkers zelf is een indicatie dat nog niet alle data waar behoefte aan zijn, al beschikbaar zijn als open data.

Uit dit onderzoek blijkt dat hergebruikers vooral belemmeringen ervaren met open data van lokale overheden. Het aanbod is versnipperd, er is geen eenduidige standaardisatie met betrekking tot de datastructuur en bestandsformaten, en niet alle lokale overheden maken gelijksoortige datasets beschikbaar als open data waardoor er geen landsdekking of regiodekking bereikt kan worden.

Uit dit onderzoek blijkt ook dat hergebruikers vooral behoefte hebben aan meer samenwerking op het gebied van open data. Hergebruikers willen graag betrokken worden bij de governance van het open data proces, maar geven ook aan dat – vooral ondernemers – geen tijd hebben om deel te nemen aan formele bijeenkomsten. Liever hebben hergebruikers dat de overheid naar hun komt om een beter beeld te krijgen van hoe open data worden ingezet en hoe de governance verbeterd zou kunnen worden. Op de korte termijn hebben hergebruikers de meeste behoefte aan betere feedback faciliteiten en een tijdige reactie op de geleverde feedback. Ook hebben hergebruikers een urgente behoefte aan meer regie op het beschikbaar stellen van open data van lokale overheden en die datasets land dekkend te maken, en gebruik van (inter)nationaal herkende standaarden voor API's.

Uit dit onderzoek blijkt ook dat open data nu een vaste waarde zijn voor veel hergebruikers. Er zijn nu bedrijven die zonder open data niet zouden kunnen bestaan met hun businessmodel geheel afhankelijk van open data. Ook ontdekken steeds meer hergebruikers uit andere sectoren de toegevoegde waarde van open data. Waren er een aantal jaren geleden nog vooral hergebruikers uit de geo-sector, nu ziet men breder hergebruik in andere sectoren. Uit dit onderzoek blijkt ook dat hergebruikers open data gebruiken om producten en diensten te ontwikkelen die een bijdrage kunnen leveren aan maatschappelijke vraagstukken.

7.2 Beperkingen van dit onderzoek

Dit onderzoek kent een aantal beperkingen wat betreft de gebruikte methodologieën. Ten eerste is er de algemene beperking van logfile gegevens van open data portals dat men alleen een beeld krijgt van hoe vaak een dataset wordt aangeropen. Maar dat zegt nog niet zo veel over daadwerkelijk hergebruik aangezien datasets ook aangeropen worden als een potentiële gebruiker de dataset alleen maar bekijkt zonder te gebruiken. Bovendien wordt bij aanroepen van data via een kaartdienst, bijvoorbeeld een Web Mapping Service (WMS), elke keer dat de kaart in of uit gezoemd wordt, dit als een nieuw verzoek geregistreerd. Ook kunnen datasets worden aangeropen als onderdeel van andere datasets, bijvoorbeeld een topografische dataset als onderliggende kaart voor andere datasets. Zo is de Basisregistratie Adressen en Gebouwen (BAG) is één van de meest aangeropen datasets omdat BAG-data gebruikt (moeten) worden voor andere diensten zoals WOZ-waardeloket. En bij data aanroepen via mobiele diensten is het IP-nummer van de gebruiker zelf niet te achterhalen, alleen het IP-nummer van de ISP.

Door de logfile gegevens van januari en februari te onderzoeken, kunnen er gebruikers geregistreerd zijn die maar eens per jaar open data gebruiken, bijvoorbeeld burgers die de WOZ-waarde van hun woning willen controleren nadat zij de jaarlijkse aanslag binnen krijgen. Andere hergebruikers kunnen juist gemist worden.

Een laatste beperking van het kwantitatieve deel van dit onderzoek is dat wij drie nationale open data platforms hebben geanalyseerd, terwijl uit het kwalitatieve deel van dit onderzoek blijkt dat veel hergebruikers hun data veelal direct bij de bronhouders opvragen, of via andere open data platforms voor bijvoorbeeld lokale open data of NDW voor verkeersgegevens.

Een beperking van kwalitatieve deel van dit onderzoek betreft het tijdsbestek waarin de online enquête was uitgezet van november 2018 tot en met januari 2019. Door de vakantieperiode viel de response van de vragenlijst lager uit dan verwacht. Hoewel online vragenlijsten een goede manier is om zo veel mogelijk gebruikers in een korte tijd te benaderen, moet men bij de antwoorden ook rekening houden met het feit dat er een mogelijke bias bij de antwoorden omdat respondenten de vragenlijsten kunnen gebruiken om problemen aan te kaarten, terwijl de tevreden gebruikers minder snel geneigd zullen zijn om deel te nemen. Verder duurde de kwantitatieve analyse van de logfile gegevens langer dan gepland waardoor een aantal hergebruikers niet geselecteerd konden worden in de longlist van diepte-interviews.

Ondanks deze beperkingen kunnen er wel een aantal aanbevelingen worden gemaakt.

7.3 Aanbevelingen

Uit het kwantitatieve deel van dit onderzoek komen een aantal aanbevelingen naar boven, zowel op het gebied van het aanbod van open data via dataplatforms als op het gebied van de dataset zelf. Wij sluiten dit hoofdstuk af met algemene aanbevelingen.

7.3.1 Aanbevelingen op het gebied van data platforms

Uit dit onderzoek kwam naar voren dat hergebruikers van open data over het algemeen tevreden zijn met het aanbod en de kwaliteit van open data, zeker op nationaal niveau. Open data worden herbruikt voor het creëren van commerciële en maatschappelijk producten en diensten. Hoewel wij voor dit onderzoek niet de maatschappelijke kosten en baten van open data in beschouwing hebben genomen, kunnen wij uit dit onderzoek concluderen dat er waarschijnlijk een positieve balans is tussen de baten en de kosten van open data. Desondanks toont dit onderzoek dat er nog steeds behoefte is aan:

- Meer samenwerking tussen de nationale overheid en lokale overheden, en tussen overheden en de hergebruikers. Maak meer gebruik maken van klankbordgroepen, online platforms en open data community's. Betrek hierbij ook organisaties zoals Logius en Civity (als exploitant van dataplatform.nl voor gemeentelijke open data)
- Betrek hergebruikers als ontwikkelaar bij de open data platforms om de dataplatforms gebruikersvriendelijker te maken.
- Verbeter de vindbaarheid van open data, dus open datasets via algemene zoekmachines op een intuïtieve manier kunnen vinden.
- Maak alle open data – ook van lokale overheden en van overheidsbedrijven met een publieke taak – via één centraal platform kenbaar en vindbaar.
- Meer samenwerking tussen de verschillende data platforms, zowel op nationaal gebied als op lokaal gebied.
- Stel een register op van ontbrekende datasets. Dit zijn zowel datasets waar verzoeken voor zijn ingediend als datasets die nog in onderzoek zijn voor open data. Maak in dit register duidelijk wie de bronhouder van de dataset is, en wat de status van de dataset is. Zorg dat het register bijgehouden wordt.
- Verplicht de overheid om via dezelfde (centrale) dataportaal data te gebruiken als niet-overheidsgebruikers. Dan lopen overheden tegen vergelijkbare belemmeringen aan, en zullen die belemmeringen sneller worden weggenomen.

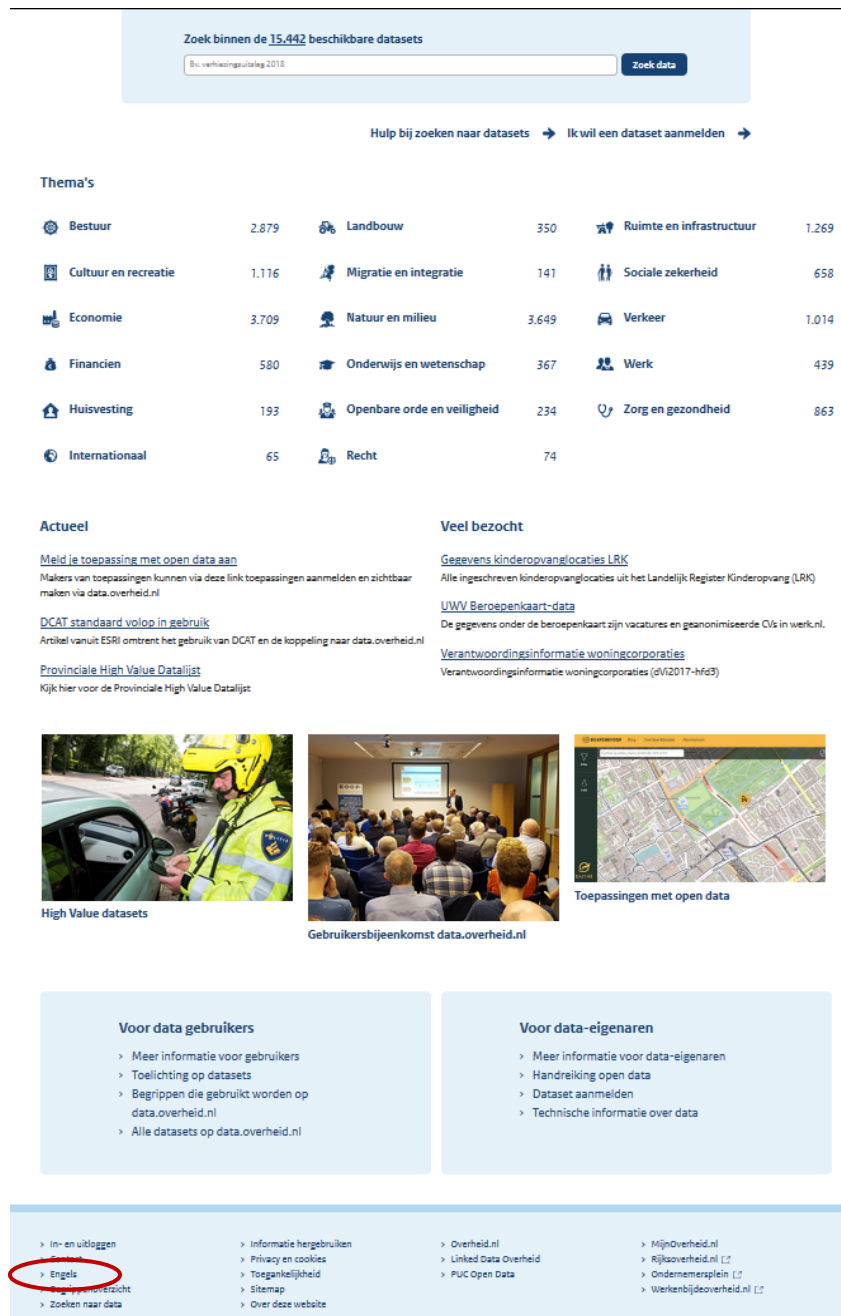
7.3.2 Aanbevelingen op het gebied van open data

Naast aanbevelingen op de manier waarop open beschikbaar worden gesteld via open data platforms, zijn er ook nog stappen te maken op het gebied van de datasets zelf.

- Stel niet alleen data beschikbaar die door de overheid is geproduceerd, maar ook de data die door de overheid zelf wordt gebruikt en afkomstig is van andere bronnen. Maak dit duidelijk tijdens het aanbestedingsproces voor het inwinnen van dergelijke data dat de afgeleide data als open data beschikbaar gesteld kunnen worden. Betrek die partijen in het open data proces.
- Verbeter de actualiteit van de datasets: zorg ervoor dat de tijd tussen nieuwe versies die bij de overheid beschikbaar komen en wanneer de open data versie beschikbaar en kenbaar wordt gemaakt, zo kort mogelijk is.
- Vind een balans tussen bescherming van persoonsgegevens en de benodigde bewerkingsslagen om data beschikbaar te maken als open data. Zorg er voor dat het anonimiseren/ aggregeren van datasets op een gestandaardiseerde en consistente manier wordt uitgevoerd voor gelijksoortige data.
- Ontwikkel geautomatiseerde processen als onderdeel van de data governance strategie om data sneller te ontsluiten als open data.
- Zorg ervoor dat metadata en aanvullende documentatie compleet zijn, en in meerdere talen en in een machine-verwerkbaar formaat beschikbaar zijn.
- Zorg ervoor dat open data zoveel mogelijk in meerdere machine-verwerkbare bestandsformaten (dus geen pdf) beschikbaar zijn. Als de originele versie in een niet-open formaat is gecreëerd, bijvoorbeeld als shapefile of word document, dan die versies ook beschikbaar stellen naast een open bestandsformaat versie.
- Data via verplichte standaarden publiceren, inclusief gestandaardiseerde en consistente codering voor de structuur.
- Maak bindende afspraken over welke datasets ontsloten moeten worden.

Verder blijkt uit ons onderzoek dat de open data portalen in Nederland ook door buitenlandse bezoekers worden gevonden. Om de open data portalen gebruikersvriendelijker voor niet-

Nederlandstalige hergebruikers te maken, zouden de portalen ten minste in een Engelstalige versie beschikbaar moeten zijn. Voor CBS open data en bij data.overheid.nl zijn er wel opties beschikbaar om naar Engels over te schakelen. Bij CBS open data staat die optie rechtsboven in het scherm, maar op data.overheid.nl is de optie om naar Engels over te schakelen niet makkelijk te vinden namelijk helemaal onderin het scherm, zie Figuur 35. Bij PDOK is er geen mogelijkheid om naar een Engelstalige versie over te schakelen.



Figuur 35: screenshot openingspagina data.overheid.nl met optie voor Engelstalige versie onderaan pagina (geraadpleegd 10 juli 2019)

7.3.3 Van data-aanbod naar data-gedreven

Voor de overheid om datagedreven te werken, zijn er nog een aantal stappen nodig. Een van de eerste stappen is de governance van (open) data. Datamanagement, zeker op nationaal niveau en met de vele betrokken dataverstrekken en stakeholders, is geen sinecure. Het zou echter de governance van open data ten goede komen als de overheid alle overheidsdata – ook de niet-

openbare data – via een centraal datapunt zouden gebruiken, zoals nu ook bij gemeente Amsterdam en Eindhoven het geval is.

Het is belangrijk dat het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties de regie oppakt, niet alleen op nationaal niveau maar vooral ook op lokaal niveau. Het beschikbaar stellen van open data is geen doel meer op zichzelf maar een onderdeel van een datastrategie om de transitie naar datagedreven organisaties te verwezenlijken. Om datagedreven te kunnen werken moeten data kunnen stromen. En dat kan alleen als er een datastrategie belegd is zowel op macroniveau (nationaal) als op mesoniveau (regionaal) en microniveau (organisatie), waarbij het voor eenieder duidelijk is wie data-eigenaren zijn en waar de verantwoordelijkheden liggen. Datakwaliteit wordt ook steeds belangrijker naarmate meer geautomatiseerde besluitvorming gebaseerd is op algoritmes. En dat vergt dat overheden transparant zijn over de bron en de inhoud van de data. Maar van beleid naar uitvoering gaan, vergt een cultuurverandering binnen organisaties en een lange adem. Daarom is het essentieel dat er een duidelijke lange termijn visie aanwezig is en dat de geplande fases ook werkelijkheid worden.

Een andere aanbeveling is om niet vanuit aanbod te denken, maar vanuit maatschappelijke vraagstukken. Voor welke uitdagingen zijn welke data nodig om tot een aanpak te komen? Ook dat behelst het betrekken van het bedrijfsleven, maatschappelijke organisaties, maar ook betrokken burgers. Dat zou kunnen door specifieke platforms op te zetten om zo een community op te bouwen rondom een bepaald maatschappelijk vraagstuk. Stakeholders en andere geïnteresseerde partijen zouden zich (tijdelijk) kunnen aansluiten bij een dergelijke community.

Het belangrijk dat er geïnvesteerd wordt in de skills die nodig zijn om de groeiende berg aan data op een effectieve manier te beheren en te kunnen gebruiken. Die skills moeten zowel binnen de overheid worden opgebouwd als bij burgers en open data hergebruikers. Met meer burgers en hergebruikers met data skills, zullen data effectiever ingezet kunnen worden om maatschappelijke vraagstukken aan te pakken en het vertrouwen in een transparante overheid te vergroten.

Ten slotte is het belangrijk dat dit onderzoek regelmatig herhaald wordt om de voortgang van open data hergebruik te monitoren. In 2014 en 2016 zijn als onderdeel van andere onderzoeken al eerder kwantitatieve analyses op basis van logfile gegevens uitgevoerd voor PDOK. Het betrekken van andere open data platforms in de analyse heeft zijn waarde bewezen voor dit onderzoek. In vervolgonderzoeken zouden ook andere, sectorspecifieke, open data platforms zoals NDW moeten worden.

Dankwoord

Wij willen graag alle mensen bedanken die hebben meegewerkt aan het tot stand komen van dit rapport. Allereerst willen wij Erwin van Mierlo, Dennis Ramondt en Leen Roosendaal van CBS, Rob Lamers en Alexander Maljaars van PDOK, en Hayo Schreijers en Jelle Verburg, en andere medewerkers van KOOP bedanken voor het verstrekken van de benodigde logfile gegevens en het faciliteren van ons onderzoek. Ook willen we de leden van de begeleidingscommissie bedanken: Marcel Hopmans, Ron Nederpelt, Jaron Haas en Paul Suikerbuik. En uiteraard willen wij alle mensen bedanken die bereid waren om geïnterviewd te worden en/of onze vragenlijst hebben ingevuld. Met hun input hebben ze waardevolle bijdrage geleverd aan de totstandkoming van dit rapport.

Lijst van afkortingen

API	Application Programme Interface
AVG	Algemene Verordening Gegevensbescherming
BAG	Basisregistratie Adressen en Gebouwen
BoW	Bag of Words
BRT	Basisregistratie Topografie
BZK	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
CBS	Centraal Bureau voor de Statistiek
CMS	Content Management System
IDF	Inverse Document Frequency
IP-adres	Internet Protocol adres
ISP	Internet Service Provider
KOOP	Kennis- en Exploitatiecentrum Officiële Overheidspublicaties
M2M	Machine-to-machine
ML	Machine Learning
NLP	Natural Language Processing
PDOK	Publieke Dienstverlening op de Kaart
TM	Term Frequency
WMS	Web Mapping Service
Wob	Wet openbaarheid van bestuur
WOZ	wet Waardering Onroerende Zaken

Literatuurlijst

- Berends, J., W. Carrara, H. Vollers, T. Fechner & M. Kleemann (2017b). Analytical Report 5: Barriers in working with Open Data: 37. https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n5_-_barriers_in_open_data.pdf, geraadpleegd 30 juni 2017.
- Berends, J., W. Carrara, W. Engberts & H. Vollers (2017a). Re-using Open Data. A study on companies transforming open data into economic & societal value. Brussels, Capgemini Consulting: 2016. https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf, geraadpleegd 6 maart 2019.
- Braggaar, R. (2018). Wi-Fi network-based indoor localisation: The case of the TU Delft campus. Fac. Architecture & the Built Environment. Delft, Delft University of Technology. Thesis MSc. Geomatics for the Built Environment: 92. <https://repository.tudelft.nl/islandora/object/uuid:bb3009cb-605e-4b80-9cf6-85f7dbc9ace8?collection=education>.
- Cipriano, P., C. Easton, E. Roglia & G. Vancauwenberghe (2013). A European Community of SMEs built on Environmental Digital Content and Languages. Final Report: 160. http://www.smespire.eu/wp-content/uploads/downloads/2014/03/D1.3_FinalReport_1.0.pdf, geraadpleegd 3 juli 2019.
- ConsultingWhere Ltd & ACIL Tasman (2012). Assessing the value of Ordnance Survey OpenData™ to the economy of Great Britain. Full study interim report: 107.
- De Smedt, T. & W. Daelemans (2012). Pattern for Python. *Journal of Machine Learning Research*. <http://doi.org/10.3928/1081597X-20130426-01>
- Deloitte LLP (2012). Open Growth: Stimulating demand for open data in the UK. A briefing note from Deloitte Analytics. London, Deloitte Touche Tohmatsu Limited: 12. <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf>, geraadpleegd 30 juni 2019.
- den Haan, D. (2018). Intermediation as a dissolver of barriers. How intermediaries can help overcome barriers in open data use. Utrecht, Utrecht University. Thesis MSc. Geo Information and Management Applications: 59. <https://dspace.library.uu.nl/handle/1874/368825>, geraadpleegd 30 juni 2019.
- Grothe, M. (2016). INSPIRE verankert zich in milieubeleid. Geo Info. 2016-2: p.44-47. https://www.geonovum.nl/uploads/documents/201602GeoInfo_INSPIRE_0.pdf, geraadpleegd 30 november 2018.
- Implementatiewet EG-richtlijn infrastructuur ruimtelijke informatie (Besluit Inspire) (2009). Netherlands, Stb. 2009, 310.
- IRM [Independent Reporting Mechanism] (2015). Aligning supply and demand for better governance. Washington, Independent Reporting Mechanism (IRM): 40. <http://www.opengovpartnership.org/sites/default/files/IRMReport-OpenData.pdf> (geraadpleegd 30 juni 2019)
- Janssen, M. & A. Zuiderwijk (2014). Infomediary Business Models for Connecting Open Data Providers and Users. Social Science Computer Review **32**(5): 563-576 doi: 10.1177/0894439314525902.
- Knops, R.W. (2019). Aanbieding NL DIGITAAL: Data Agenda Overheid. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Den Haag: 3. kst-26643-597. <https://zoek.officielebekendmakingen.nl/kst-26643-597.html>, geraadpleegd 30 Juni 2019.
- Labots, W. (2016). Gebruiker in kaart. Analyse van loggegevens van de PDOK-webservices om groepen gebruikers van open data te identificeren. Delft, Delft University of Technology: 33. Thesis BSc Technology Policy & Management.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). NL DIGITAAL. Data Agenda Overheid. Den Haag: 31. kst-26643-597. <https://zoek.officielebekendmakingen.nl/blg-876545.pdf>, geraadpleegd 30 juni 2019.

- Richtlijn 2007/2/EG van het Europees Parlement en de Raad van 14 maart 2007 tot oprichting van een infrastructuur voor ruimtelijke informatie in de Gemeenschap (INSPIRE) (2007). Publicatieblad van de Europese Gemeenschap, L108/50-63, Europese Unie. p.1-14. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:NL:PDF>.
- RIPE NCC. (n.d.). 4.2.4 Description of the INETNUM Object. Retrieved February 4, 2019, from <https://www.ripe.net/manage-ips-and-asns/db/support/documentation/ripe-database-documentation/rpsl-object-types/4-2-descriptions-of-primary-objects/4-2-4-description-of-the-inetnum-object>
- Vancauwenberghe, G., F. Welle Donker, & B. van Loenen (2017). Business Models for Geographic Information. In J. B. Kruse, J. Crompvoets, & F. Pearlman (Eds.), *GEOValue: The Socioeconomic Value of Geospatial Information*. (pp. 85–115). CRC PRes.
- van Loenen, B., F. Welle Donker & D. Kilic (2018). Overheid en “sociale” locatiedata: Een verkennend onderzoek naar de stand van social media (locatie)data gebruik Nederlandse overheid. Geonovum, Amersfoort: 38. <http://resolver.tudelft.nl/uuid:9187527d-6f48-4ccb-a6fa-c63930407999>, geraadpleegd 30 juni 2019.
- van Loenen, B., J. Ubacht, W. Labots & A. Zuiderwijk (2017). Log File Analytics for Gaining Insight into Actual Use of Open Data. 17th European Conference on Digital Government (ECDG 2017). J. Vieira Borges and J. C. Dias Rouco. Lisbon, Academic Conferences and Publishing International Ltd: 238-246. <http://kcopendata.eu/wp-content/uploads/2017/06/2017-bvl-ECDG17-paper-only.pdf>, geraadpleegd 28 juni 2019.
- van Loenen, B., F. Welle Donker & R. Braggaar (2016). De stand in opendataland 2016. Delft, Kenniscentrum Open Data: 96. <http://repository.tudelft.nl/islandora/object/uuid:b76cb0d8-e9b5-4125-8d7c-b452de1da0d0?collection=research>, geraadpleegd 30 juni 2019.
- van Loenen, B. & F. Welle Donker (2014). De stand in opendataland. Delft, OTB - Research for the Built Environment: 116. <http://repository.tudelft.nl/view/ir/uuid%3A7b425db6-2f9a-4be6-b5f0-3492055268d8/>, geraadpleegd 30 juni 2019

Bijlage 1 Overzicht van gebruikte datasets door grootgebruikers per open data portaal

	81588NED	82245NED	83487NED	83220NED	37296NED	83148NED	81589NED	80590NED	84166NED	70072NED
Kamer van Koophandel Nederland	734	22	22	22	4	444	249	3	137	112
baten-lastendienst Logius	88	91	92	89	65	25	46	79	56	21
ING Group	198	127	124	123	26	26	25	40	37	4
Hogeschool Van Arnhem en Nijmegen	277	48	45	45	141	26	93	33	22	21
Fontys Hogescholen	255	63	59	59	151	41	42	39	19	24
Erasmus Universiteit Rotterdam	91	39	37	35	65	28	45	30	22	32
Hogeschool Van Amsterdam	160	14	12	11	154	28	60	93	31	18
Belastingdienst/Centrum voor Infrastruc...	265	193	192	191	12	60	40	16	25	6
Rijksuniversiteit Groningen	67	55	55	51	78	16	25	23	9	24
Universiteit Utrecht	40	26	33	20	92	19	16	19	8	49
Wageningen Universiteit	52	48	47	46	42	15	25	6	9	14
Ministerie van Justitie	105	160	156	152	34	25	10	16	12	12
Ministerie van Economische zaken	67	41	41	42	4	49	40	17	21	14
Universiteit van Tilburg	39	31	32	30	49	4	22	44	18	18
Uva.nl	78	24	23	25	47	12	20	9	7	14
DNB.nl	3	4	4	4	6	1	8	14	7	1
Ministerie van Verkeer en Waterstaat/Rijk...	7	70	70	69	3	5	10	11	5	10
Gemeente Heerlen	7	59	59	58	17	7	3	1	2	12
Zadkine	69	8	8	8	11	30	14	9	10	5
Gemeente Delft	2	6	6	6	4	-	9	-	2	8
Het Kadaster en de Openbare Registers	6	12	12	12	3	3	2	2	2	10
Within Reach Group B.V.	5	36	36	36	1	2	1	1	1	-

Figuur 36: gebruikte datasets via CBS open data portaal door grootgebruikers

	brtachtergrondkaart	bag	kadastralekaartv3	luchtfoto/rgb	"luchtfoto/rgb"	opentopoachtergr...	brtachtergrondkaa...	brtachtergrondkaa...	opentopo	bagterugmeldingen
Het Kadaster en de Openbare Registers	3,5 mln.	1,1 mln.	2,3 mln.	1,1 mln.	1,2 mln.	241,1K	429,9K	198,3K	215,6K	119,2K
Within Reach Group B.V.	2,7 mln.	2,1 mln.	1 mln.	995,8K	1,1 mln.	569,4K	537,4K	111,1K	681,5K	308,3K
Ministerie van Verkeer en Waterstaat/Rijk...	1,7 mln.	198,5K	269,3K	192,9K	340,9K	1,1 mln.	257,4K	343,4K	37,2K	26,1K
NAT Publishing Amsterdam-Grontmij Bu...	1,1 mln.	127,1K	218,6K	1,6 mln.	499,7K	449,7K	238,7K	45,9K	399,6K	30,6K
baten-lastendienst Logius	1,3 mln.	304K	314,2K	165,4K	320,6K	8,4K	121,2K	145,3K	4,6K	36,1K
Gemeente Heerlen	1,3 mln.	296,6K	293,9K	72,3K	98K	24K	39,3K	20,5K	9,7K	59,1K
Ministerie van Economische zaken	1,2 mln.	424,8K	63,2K	20,5K	36,6K	11,5K	197K	18,6K	28	93,6K
ING Group	1,1 mln.	787,3K	68,7K	16,2K	28K	461	59,5K	24K	-	35,2K
Hogeschool Van Amsterdam	1 mln.	52,1K	28,3K	21,1K	83,5K	257,6K	116,5K	53,4K	5,7K	2,5K
Wageningen Universiteit	510,2K	134K	144,8K	109,9K	212K	19,7K	59,5K	204,6K	13K	1,7K
Kamer van Koophandel Nederland	326,1K	1,1 mln.	6,9K	486	3,1K	366	9,8K	3,8K	24	222,3K
Gemeente Delft	400,1K	128K	17,3K	6,7K	13,2K	38,2K	10,7K	14,1K	1,5K	24,1K
Universiteit Utrecht	304,6K	85,5K	30,7K	128,3K	143,6K	166,5K	65K	12,5K	3,2K	11,1K
Belastingdienst/Centrum voor Infrastruc...	208,3K	351,7K	94,6K	58,6K	-	526	24,8K	8,5K	24	56,2K
Ministerie van Justitie	320,9K	140K	70,1K	15,1K	31,6K	3,3K	68K	67,8K	423	7,3K
Rijksuniversiteit Groningen	147,5K	111,9K	60,7K	94,6K	73,2K	8,9K	31,9K	12,1K	79,1K	14,2K
Uva.nl	115,6K	29,6K	18K	10,6K	4,1K	2K	26,2K	90,5K	3,9K	1,7K
Hogeschool Van Arnhem en Nijmegen	112,8K	11,1K	10,4K	5,1K	6,4K	719	10,7K	2,7K	656	929
Universiteit van Tilburg	51,6K	19,4K	20,6K	573	8,8K	127	14,3K	55,8K	-	1,3K
Fontys Hogescholen	52,5K	24,9K	21,8K	5,7K	6,6K	189	12,9K	38,8K	343	2,1K
Erasmus Universiteit Rotterdam	45,5K	26,6K	14K	2,8K	10,1K	9K	11,2K	7,6K	-	1,4K
Zadkine	55,3K	13,6K	6,9K	2,5K	7,6K	364	6,1K	2,2K	-	2,4K

Figuur 37: gebruikte datasets via PDOK open data portaal door grootgebruikers

	Immigratie: Vertrek ...	Immigratie: Toelati...	Immigratie: Nation...	Immigratie: Opvan...	Gegevens kindero...	Verantwoordingsi...	Nationale bewegw...	Immigratie: Toelati...	Aanhuurtransactie...	Trajectcontrole sy...
baten-lastendienst Logius	-	7	-	-	53	68	1	-	56	-
Ministerie van Justitie	172	136	116	92	-	-	-	60	-	52
DNB.nl	-	-	-	-	-	3	-	-	-	-
Ministerie van Verkeer en Waterstaat/Rijk...	-	-	-	-	-	-	62	-	-	-
Rijksuniversiteit Groningen	-	2	-	-	-	-	-	-	-	-
Ministerie van Economische zaken	-	-	4	-	-	-	-	-	-	-
Universiteit van Tilburg	-	-	-	-	-	2	-	-	-	-
Uva.nl	-	-	-	-	-	-	-	-	-	-
Fontys Hogescholen	-	-	-	-	-	-	-	-	-	-
Wageningen Universiteit	-	-	-	-	-	-	-	-	-	-
Zadkine	-	-	-	-	3	-	-	-	-	-
Within Reach Group B.V.	-	-	-	-	8	-	-	-	-	-
Universiteit Utrecht	-	-	-	-	-	-	-	-	-	-
Erasmus Universiteit Rotterdam	-	-	-	-	-	-	-	-	-	-
Hogeschool Van Arnhem en Nijmegen	-	-	-	-	-	-	-	-	-	-
Belastingdienst/Centrum voor Infrastruc...	-	-	-	-	9	-	-	-	-	-
Hogeschool Van Amsterdam	-	-	-	-	-	-	-	-	-	-
Gemeente Delft	-	-	-	-	-	-	-	-	-	-
Kamer van Koophandel Nederland	-	-	-	-	-	-	-	-	-	-
Het Kadaster en de Openbare Registers	-	-	-	-	-	-	-	-	-	-
ING Group	2	-	-	-	-	-	-	-	-	-
Gemeente Heerlen	-	-	-	-	13	-	-	-	-	-

Figuur 38: gebruikte datasets via data.overheid.nl door grootgebruikers

Bijlage 2 Lijst van geïnterviewden voor onderzoek

contactpersoon	Naam organisatie	functie	datum	Interviewer
Tom Kunzler	Open State Foundation	adjunct directeur	4-6-2019	Frederika Welle Donker
Martijn de Meulder	Waterkaart Nederland	ontwikkelaar en freelance journalist	6-6-2019	Frederika Welle Donker
Thijs Brentjes	MobilityLabel	ontwikkelaar en mobility analyst	13-6-2019	Bastiaan van Loenen
Marjan Schnetz	Ockham Groep BV	datagoeroe	14-6-2019	Frederika Welle Donker
Stephan Okhuijsen	Datagraver	datagraver / adviseur	14-6-2019	Frederika Welle Donker
Alex Olieman	Waar Overheid	ontwikkelaar	14-6-2019	Frederika Welle Donker
Duco Mansvelder	SIVI	consultant	21-6-2019	Frederika Welle Donker

Bijlage 3 Overzicht van IP register velden

Attribute Name	Presence	Repeat	Indexed	
inetnum:	mandatory	single	primary/lookup	key
netname:	mandatory	single	lookup key	
descr:	optional	multiple		
country:	mandatory	multiple		
geoloc:	optional	single		
language:	optional	multiple		
org:	optional*	single	inverse key	
sponsoring-org:	optional	single		
admin-c:	mandatory	multiple	inverse key	
tech-c:	mandatory	multiple	inverse key	
abuse-c:	optional	single	inverse key	
status:	mandatory	single		
remarks:	optional	multiple		
notify:	optional	multiple	inverse key	
mnt-by:	mandatory	multiple	inverse key	
mnt-lower:	optional	multiple	inverse key	
mnt-routes:	optional	multiple	inverse key	
mnt-domains:	optional	multiple	inverse key	
mnt-irt:	optional	multiple	inverse key	
created:	generated	single		
last-modified:	generated	single		
source:	mandatory	single		

Bijlage 4 Methodologie Sentiment Analyse

B.4.1 Verzamelen van data

Om het sentiment en de beleving van open data van de gebruikers te bepalen is allereerst data verzameld vanaf het Twitter platform. Voor het verzamelen van de data is een Twitter developers account aangemaakt.³⁸ Vervolgens is binnen het developers account een applicatie aangemaakt die toegang verschaft tot de benodigde API sleutels en tokens.

Via een Python script is vervolgens data verzameld met behulp van diverse query's. Deze query's zijn steeds opgebouwd uit een aantal zoekwoorden. De steekwoorden of tags waarmee Twitter berichten via onderwerp vindbaar maakt op het platform worden hashtags genoemd.³⁹ Hashtags kenmerken zich doordat deze geen spaties of andere speciale leestekens bevatten. De berichten hebben idealiter twee hashtags die op iedere plek binnen het bericht kunnen voorkomen. Bij het uitvoeren van de zoekopdracht wordt niet alleen naar de hashtags gekeken, maar ook naar de inhoud van het bericht. Zo zijn ook berichten opgehaald waar bijvoorbeeld helemaal geen hashtags in voorkwamen of hashtags die niet overeenkwamen met de zoekopdracht, maar waarbij wel de inhoud overeenkwam.

Er is gezocht met de volgende zoekwoorden en hashtags combinaties:

- (#)opendata AND gebruik
- (#)opendata AND ervaring
- (#)opendata AND portaal
- (#)opendata AND vindbaarheid

Door het gebruik van Nederlandse woorden in de zoekopdracht is de taal gelimiteerd tot alleen Nederlandse berichten. Het limiteren van de geografische locatie van de berichten zorgde ervoor dat er te weinig resultaten konden worden verzameld aangezien niet alle Tweets zijn voorzien van een geolocatie. Daarnaast zijn de geolocaties niet altijd eenduidig waardoor een bericht onbedoeld buiten de zoekopdracht kan vallen.

De officiële Twitter 'search API' voor het programmatisch ophalen van Tweets is gelimiteerd zodat alleen berichten van de afgelopen week kunnen worden opgehaald. Daarnaast is er ook een quotum wat het maximale aantal opgehaalde berichten beperkt. Deze beperkingen gelden ook voor alle libraries en programma's die indirect van deze API gebruik maken. Voor het verzamelen van de Tweets is daarom gebruik gemaakt van de GetOldTweets library.⁴⁰ Hiermee zijn tussen de periode 2010 tot 2018 alle Tweets behorende bij de zoektermen opgehaald. Zo worden de Twitter archieven via het web (net als een gewone gebruiker) gecrawld in plaats van via de API en kent daardoor geen beperkingen.

³⁸ <https://developer.twitter.com/en/apps>

³⁹ <https://help.twitter.com/en/using-twitter/how-to-use-hashtags>

⁴⁰ <https://github.com/Jefferson-Henrique/GetOldTweets-python>

B.4.2 Data voorbereiding

Om de Tweets te kunnen verwerken is voorbereiding noodzakelijk. Er wordt namelijk alleen gewerkt met tekst en niet met numerieke waarden. Een zin kan hierbij worden gezien als een opeenvolging van losse woorden die afzonderlijk en in verband met elkaar kunnen worden gezien. Er is aangenomen dat de waarde van woorden wordt beïnvloed door interpunctie en hoofdlettergebruik. Het opschonen van de data is daarnaast belangrijk omdat naast de gebruikelijke tekens, zoals letters en cijfers ook tekens kunnen voorkomen die niet deel uitmaken van de Nederlandse taal. Voorbeelden hiervan zijn emoticons en andere tekens die vaak gebruikt worden om een mening of bericht te voorzien van extra emotie.

1. Hoofdletters naar kleine letters
2. Verwijderen leestekens
3. Verwijderen vreemde tekens (emoticons e.d.)
4. Verwijderen van stopwoorden

B.4.3 Sentimentanalyse

Na voorbereiding van de data zijn de Tweets vervolgens geanalyseerd op sentiment. Dit houdt in dat elke Tweet voorzien is van een label positief, neutraal of negatief. In dit geval is elke Tweet een 'feature' van maximaal 280 karakters en de 'target' is een polariteitswaarde tussen de -1 (negatief) en 1 (positief). Het gebied van -0,10 tot 0,10 heeft een waarde van neutraal.

Belangrijke factor voor het classificeren zijn de bijvoeglijke naamwoorden die gebruikt worden in een Tweet. Deze adjectieven worden vaak gebruikt om een ervaring of mening te verkondigen zoals "slechte data" of "moeilijk vindbare data". Elk woord is in dit proces gevectoriseerd zodat een numerieke waarde overblijft. Het vectoriseren is gedaan met behulp van de Bag of Words (BoW) methode waarbij met behulp van een vocabulaire van bekende woorden woordfrequenties van de inkomende tekst worden verzameld. Frequenties zijn op basis van Term Frequency (TF) en hierdoor gewogen voor de lengte van de Tweet. Daarnaast is door middel van Inverse Document Frequency (IDF) de frequentie van woorden die heel vaak voorkomen minder zwaar gewogen. Door middel van N-grams houdt het model rekening met de context van woorden. Een 3-gram kan bijvoorbeeld zijn 'goede open data'. Zonder N-grams zou 'slechte kwaliteit open data' toch nog positief kunnen worden beoordeeld omdat ieder woord afzonderlijk een bepaalde score wordt toegekend. Door 'slechte' en 'kwaliteit' te verbinden kan de hele Tweet worden bestempeld als negatief.

Voorbeelden van geanalyseerde Tweets:

- Voorbeeld van positief gelabelde Tweet:

"Zeker doen, mooie ervaring ! En ik ben beschikbaar voor goed gesprek ;-) over #communitymngt #hackathons of #opendata [https:// twitter.com/simoonfransen/status/734655401170440192 ...](https://twitter.com/simoonfransen/status/734655401170440192)"

Scores:

```
[(['zeker'], 0.3, 0.5, None), (['mooie', '!'], 0.875, 1.0, None), (['beschikbaar'], 0.4, 0.4, None), (['goed'], 0.55, 0.9, None), ([';-)'], 0.25, 1.0, 'mood')]
```

Het bijwoord zeker en de bijvoeglijke naamwoorden 'mooie', 'goed' en 'beschikbaar' zorgen hier voor een positieve score.

- Voorbeeld van neutraal gelabelde Tweet:

"#ZZP #freelance #zpz Drupal developer met ckan opendata ervaring : Voor een migratie project zijn wij op zoek n... [http:// bit.ly/cVweeg](http://bit.ly/cVweeg)"

Scores:

[]

In deze Tweet zijn geen woorden gebruikt die met een bepaalde sentimentswaarde kunnen worden gewaardeerd. De Tweet is daarom gemarkeerd als neutraal.

- Voorbeeld van negatief gelabelde Tweet:

"Sodeju. Uit eigen ervaring weet ik dat het ontvangen van een simpele dataset ongeveer 100 werkdagen kost #opendata <https://twitter.com/OpendataNL/status/903184701120237568> ..."

Scores:

[(['eigen'], 0.0, 0.3333333333333333, None), (['simpele'], -0.25, 0.95, None)]

In dit geval wordt alleen het woord *simpele* bestempeld als een negatieve sentimentswaarde. De Tweet zelf is hierdoor licht negatief beoordeeld.

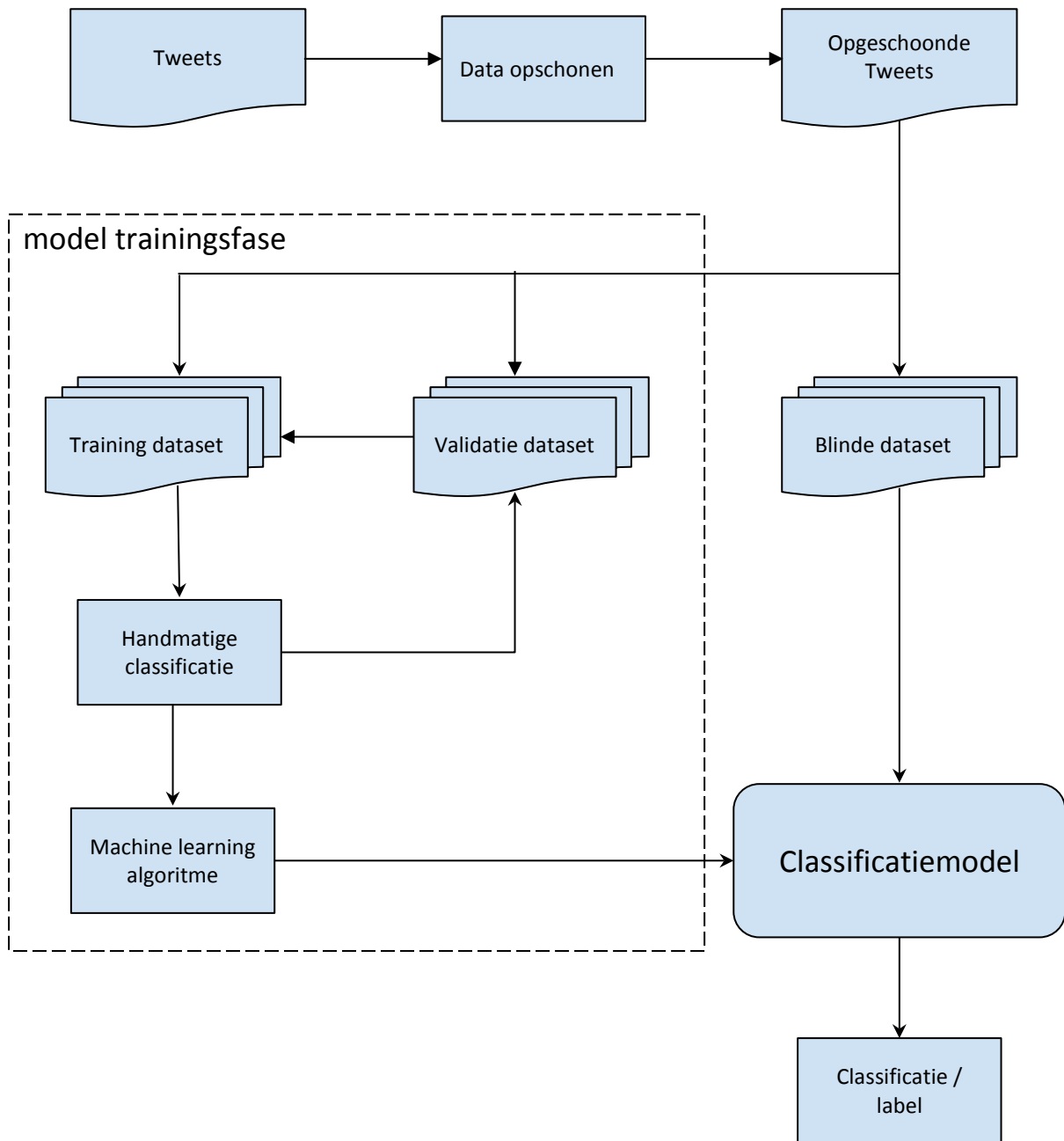
B.4.4 Classificatie Tweets

Om erachter te komen wat de meest voorkomende barrières zijn die gebruikers ondervinden zijn de Tweets geïdentificeerd in 5 categorieën:

1. Problemen met de beschikbaarheid of data zelf data, bijvoorbeeld slechte kwaliteit, fouten.
2. Problemen met de verkrijgbaarheid van data, bijvoorbeeld portaal offline.
3. Problemen met de metadata, bijvoorbeeld slechte vindbaarheid.
4. Kosten te hoog, bijvoorbeeld door hoge administratie- of verwerkingskosten.
5. Overige berichten. In deze categorie vallen alle Tweets die niet gerelateerd kunnen worden aan potentiële gebruiksbarrières.

Deze categorieën zijn vooraf bepaald op basis van belemmeringen geïdentificeerd in Hoofdstuk 4. Wij hebben van die genoemde belemmeringen aangenomen dat dit anno 2018 nog steeds potentiële barrières kunnen zijn voor hergebruikers van open data.

Het classificatiemodel is gebruikt om voor alle Tweets een label te genereren en ze te kwalificeren in een van de bovenstaande categorieën. Machine learning is hierbij toegepast om met behulp van vooraf gelabelde Tweets, de training dataset, automatisch onbekende Tweets (de blinde dataset) te labelen aan de hand van statistische gelijkheid met de trainingsdata in een van de categorieën (Figuur 39). De gelijkheid tussen de Tweets met een van de categorieën is bepaald door eerst de relevante woorden per Tweet te bepalen.



Figuur 39: classificatieproces Tweets

De Tweets zijn vervolgens opgedeeld in een training dataset, een validatie dataset en een blinde dataset. De Tweets uit de training dataset zijn handmatig ingedeeld per categorie. Hiermee is vervolgens met een machine learning algoritme het classificatiemodel getraind. Gedurende de trainingsfase is de kwaliteit van het model getoetst door vooraf gevalideerde Tweets door het model te laten classificeren. De verhouding tussen het aantal correct geclassificeerde Tweets en het totaal aantal Tweets geeft de nauwkeurigheid van het classificatiemodel aan.

Van elke Tweet is een woordentelling bijgehouden. Met de gewogen waarde (TF-IDF waarde) is de relevantie per woord vervolgens bepaald. Met de relevante woorden uit deze blinde dataset is daarop vergeleken met welke relevante woorden van de training dataset categorieën de beste

overeenkomst is. De standaard classificatie die is toegekend aan een Tweet is overige categorie, welke over het algemeen meest voorkomend is (de baseline).

B.4.5 Trainen van classificatiemodel

Voorbeelden van Tweets per categorie:

1. Problemen met de beschikbaarheid van open data en slechte kwaliteit, foutieve data etc.

"@GeoBusinessNL vat samen: initiatieven als @PDOK_online zijn niet de bottleneck in gebruik #opendata , maar de instanties die niet meedoen"

2. Problemen met de verkrijgbaarheid van data, bijvoorbeeld portaal offline.

"zo moeilijk is het niet. Er zit momenteel een beperking op gebruik #opendata NS. En @floorter wil daar terecht vanaf. Kan dat?"

3. Problemen met de metadata, bijvoorbeeld slechte vindbaarheid.

"@ArchivePortal #opendata aanbieden op website met onveilig certificaat zal het gebruik niet stimuleren pic.twitter.com/8nuxCcxmld"

4. Kosten te hoog, bijvoorbeeld door hoge administratie- of verwerkingskosten.

"open data, maar waar betaald de gebruiker dan voor?"

5. Overige berichten. In deze categorie vallen alle Tweets die niet gerelateerd kunnen worden aan potentiële gebruiksbarrières.

"Uniek! Overheid in verlegenheid door gebruik #opendata , de ultieme confrontatie met transparantie. Doorbraak? https://twitter.com/h20_bug/status/812327501171191812 ..."

Het classificatiemodel is vervolgens gebruikt om iedere Tweet in te delen in één van de vijf categorieën. Bijlage 3 toont het script wat is gebruikt voor het classificeren van circa 25.000 Tweets.

Als Tweets willekeurig zouden worden geclassificeerd dan is de nauwkeurigheid 20% (1 van de 5 categorieën).

Bijlage 5 Python script Tweets ophalen en sentiment analyse

```
# -*- coding: utf-8 -*-
import sys
# check python version for correct version of module
# got3 is still in beta so just use python27 for now
if sys.version_info[0] < 3:
    import got
else:
    import got3 as got
from pattern.db import Datasheet
from pattern.nl import sentiment as sentiment_nl, positive
from pattern.web import plaintext

def main():
    search_term = raw_input('Search term: ').strip()

    csv_file = Datasheet()
    csv_file.append(['date', 'tweet_id', 'username', 'tweet_text', 'retweets', 'permalink',
'favorites', 'mentions', 'geolocation', 'hashtags', 'sentiment', 'subjectivity',
'sentiment_label'])

    def saveTweet(t):
        tweet_text = plaintext(t.text)
        sentiment = sentiment_nl(tweet_text)
        polarity = sentiment[0]
        subjectivity = sentiment[1]
        if polarity <= -.1:
            polarity_label = 'negatief'
        elif -.1 <= polarity <= .1:
            polarity_label = 'neutraal'
        else:
            polarity_label = 'positief'
        csv_file.append([t.date, t.id, t.username, tweet_text, t.retweets, t.permalink,
t.favorites, t.mentions, t.geo, t.hashtags, polarity, subjectivity, polarity_label])

    #Get tweets by query search (date boundaries: "yyyy-mm-dd")
    tweetCriteria = got.manager.TweetCriteria().setQuerySearch('opendata AND
{}'.format(search_term)).setSince("2010-1-1").setUntil("2018-12-31").setMaxTweets(100000)

    tweets = got.manager.TweetManager.getTweets(tweetCriteria)

    print len(tweets)

    for tweet in tweets:
        saveTweet(tweet)

    csv_file.save('twitter_results_{}.csv'.format(search_term))

if __name__ == '__main__':
    main()
```

Bijlage 6 Python script sentiment analyse assessments per woord

```
# -*- coding: utf-8 -*-
from pattern.nl import sentiment as sentiment_nl

tweet = """Aan de slag met #opendata ? Benut de ervaring van deelnemers #pnhslim
via de Open Data toolbox op http://www.noord-holland.nl/web/Projecten/NoordHolland-Slimmer/Toolbox-Open-Data.htm ..."""

print sentiment_nl(tweet).assessments
```

Bijlage 7 Rapportage online enquête

Vragenlijst Open Data gebruikers

Inhoudsopgave

1	Algemeen	71
	Algemeen	71
2	Vragenlijst	72
	Tot welke categorie (her)gebruiker rekent u uzelf het meest?	72
	korte toelichting categorie (her)gebruiker, indien u dit wenst te geven	73
	Voor welk doeleinde gebruikt u open data het meest? (meerdere antwoorden mogelijk)	74
	Via welk platform / data portaal vindt u de benodigde open data? (meerdere antwoorden mogelijk)	75
	Welk open data platform heeft uw voorkeur en waarom? (optioneel)	76
	Via welke type diensten(en) gebruikt u open data?	77
	Welke verbeterpunten ziet u voor de overheid m.b.t. vindbaarheid van open data?	78
	Welke verbeterpunten ziet u voor de overheid m.b.t. manier waarop open data wordt aangeboden?	79
	Hoe ervaart u de kwaliteit van open data in het algemeen? Hierbij kunt u denken aan compleetheid van de data, metadata, bruikbare formaten, etc.	80
	Welke belemmeringen ervaart u bij het gebruik van open data?	81
	Welke verbeterpunten ziet u voor de overheid m.b.t. kwaliteit van open data?	82
	Hoe bent u nu bij het open data proces betrokken?	83
	Hoe zou u graag (meer) bij het open data proces betrokken willen worden?	84
	Heeft u nog opmerkingen m.b.t. deze vragenlijst? Laat hier uw e-mailadres achter als u op de hoogte gehouden wilt worden van de uitkomsten van dit onderzoek	85

1. Algemeen

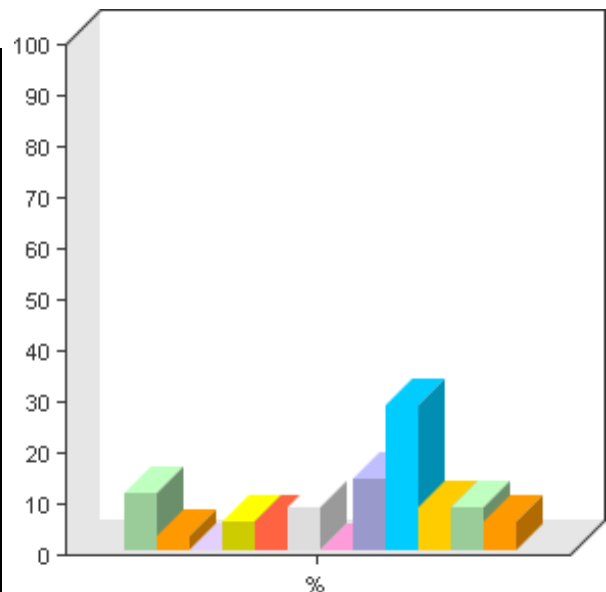
Gegevens	
Vragenlijst	Vragenlijst Open Data gebruikers
Inhoud	16 vragen
Algemene statistieken	
Totaal aantal benaderd	567
Aantal begonnen	35 (6%)
Aantal afgerond	35 (6% totaal, 100 % begonnen)

2. Vragenlijst

Algemeen	
Vraag / variabele	Tot welke categorie (her)gebruiker rekent u uzelf het meest?
Type	Enkele-keuze
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Tot welke categorie (her)gebruiker rekent u uzelf het meest?

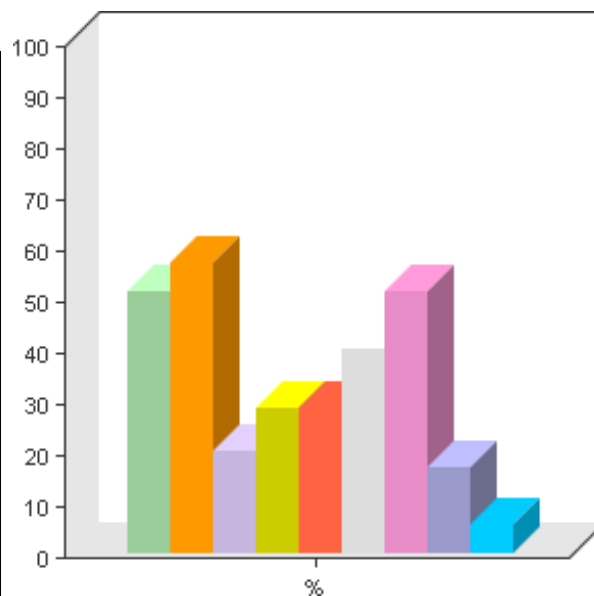
Antwoord mogelijkheid	Aantal	Percentage
overheid, nationaal	4	11%
overheid, provincie/waterschap	1	3%
overheid, gemeente	0	0%
semi-overheid / private organisatie met publieke taak	2	6%
wetenschap / onderwijs	2	6%
bedrijfsleven, groot (>250 medewerkers)	3	9%
bedrijfsleven, middelgroot (50-250 medewerkers)	0	0%
bedrijfsleven, klein (10-49 medewerkers)	5	14%
bedrijfsleven, micro (10	29%
not-for-profit organisatie	3	9%
Burger	3	9%
anders, namelijk:	2	6%



Algemeen	
Vraag / variabele	Voor welk doeleinde gebruikt u open data het meest? (meerdere antwoorden mogelijk)
Beantwoord door	35 (100%)

Voor welk doeleinde gebruikt u open data het meest? (meerdere antwoorden mogelijk)

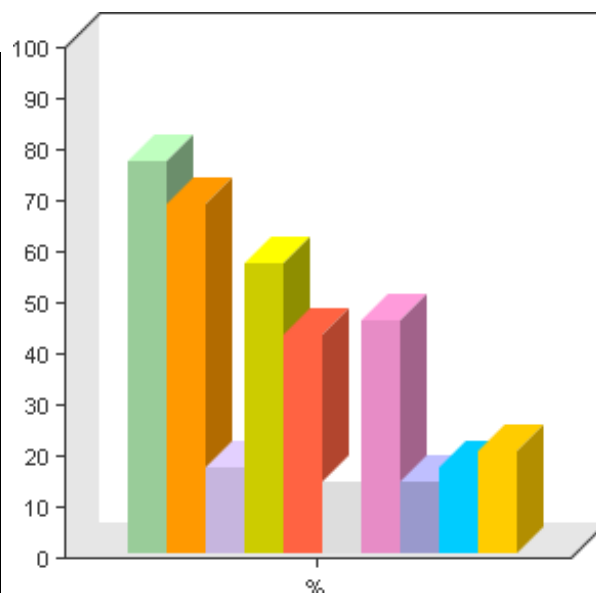
Antwoord mogelijkheid	Aantal	Percentage
voor commerciële producten / diensten (for profit)	18	51%
voor maatschappelijke producten / diensten (not-for-profit)	20	57%
als lokkertje voor andere producten / diensten / dienstverlening	7	20%
voor interne bedrijfsvoering (interne efficiëntie / effectiviteit)	10	29%
voor het uitvoeren van beleid / maken van beslissingen / publieke dienstverlening	10	29%
voor onderzoek / onderwijs	14	40%
voor experimenteren / R&D	18	51%
voor privédoeleinden	6	17%
anders, namelijk:	2	6%



Algemeen	
Vraag / variabele	Via welk platform / data portaal vindt u de benodigde open data? (meerdere antwoorden mogelijk)
Beantwoord door	35 (100%)

Via welk platform / data portaal vindt u de benodigde open data? (meerdere antwoorden mogelijk)

Antwoord mogelijkheid	Aantal	Percentage
via data.overheid.nl	27	77%
via PDOK.nl (Publieke Dienstverlening op de Kaart)	24	69%
via dataplatform.nl	6	17%
via CBS open data portaal	20	57%
via Nationaal Georegister	15	43%
via Provinciaal Georegister	5	14%
via dataportaal op website van data-aanbieder	16	46%
via contactformulier op website van data-aanbieder	5	14%
via een officieel verzoek onder de Wet openbaarheid van bestuur (Wob-verzoek)	6	17%
anders, namelijk:	7	20%



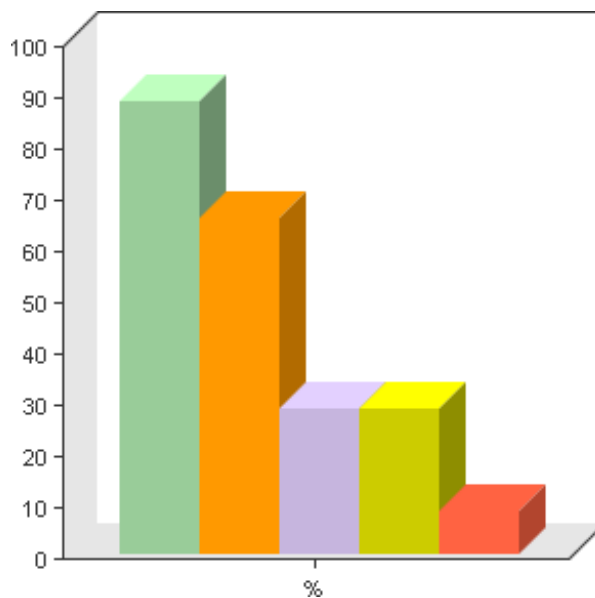
Algemeen	
Vraag / variabele	Welk open data platform heeft uw voorkeur en waarom? (optioneel)
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Welk open data platform heeft uw voorkeur en waarom? (optioneel)

(geen statistieken beschikbaar)

Algemeen	
Vraag / variabele	Via welke type diensten(en) gebruikt u open data? (meerkeuze)
Beantwoord door	35 (100%)

Antwoordmogelijkheid	Aantal	Percentage	
download dienst	31	89%	
API	23	66%	
catalogus / zoekdienst	10	29%	
visualisatie	10	29%	
anders, namelijk:	3	9%	



Algemeen	
Vraag / variabele	Welke verbeterpunten ziet u voor de overheid m.b.t. vindbaarheid van open data?
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Welke verbeterpunten ziet u voor de overheid m.b.t. vindbaarheid van open data?

(geen statistieken beschikbaar)

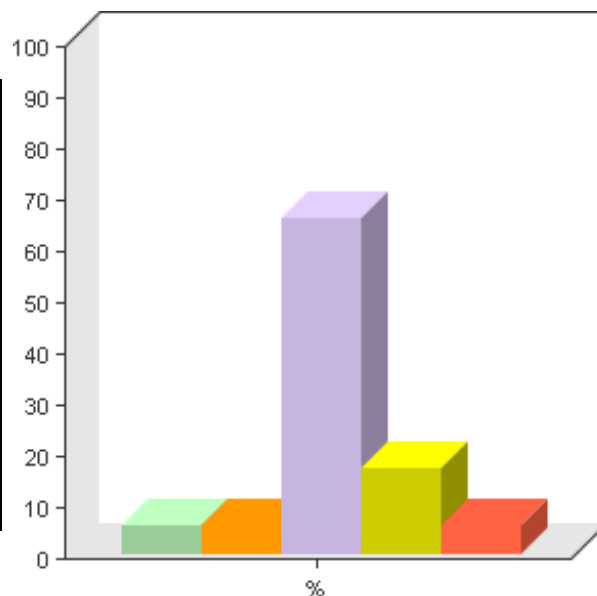
Algemeen	
Vraag / variabele	Welke verbeterpunten ziet u voor de overheid m.b.t. manier waarop open data wordt aangeboden?
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Welke verbeterpunten ziet u voor de overheid m.b.t. manier waarop open data wordt aangeboden?

(geen statistieken beschikbaar)

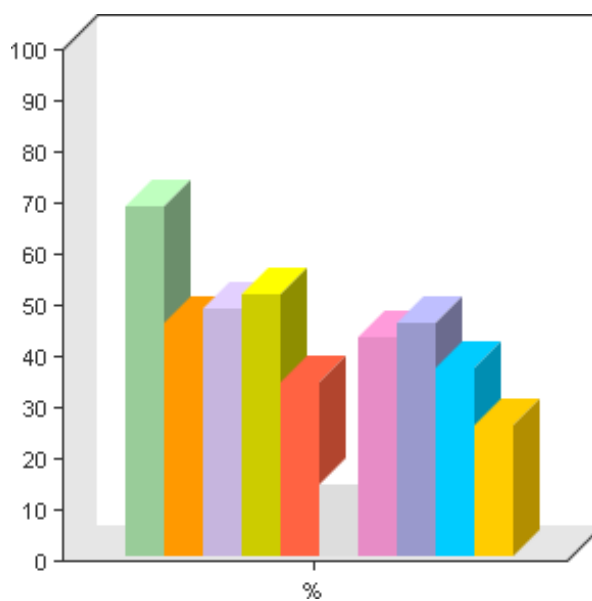
Algemeen	
Vraag / variabele	Hoe ervaart u de kwaliteit van open data in het algemeen? Hierbij kunt u denken Hierbij kunt u denken aan compleetheid van de data, metadata, bruikbare formaten, etc.
Beantwoord door	35 (100%)

Antwoordmogelijkheid	Aantal	Percentage
waardeloos, moet veel investeren in data opschonen / converteren / additionele data toevoegen	2	6%
niet goed, maar beter dan niets of betalen voor de data	2	6%
afdoende, maar er is genoeg ruimte voor verbeteringen	23	66%
goed, zeker voor mijn beoogd doeleinde	6	17%
anders, namelijk:	2	6%



Algemeen	
Vraag / variabele	Welke belemmeringen ervaart u bij het gebruik van open data?
Beantwoord door	35 (100%)

Antwoordmogelijkheid	Aantal	Percentage
data is niet actueel	24	69%
data is incompleet qua dekkingsgraad	16	46%
data te aggregaerd	17	49%
metadata is incompleet / afwezig	18	51%
aangeboden dataformaat is niet bruikbaar voor mij	12	34%
databestand is te groot	5	14%
data is onvolledig	15	43%
kan data niet makkelijk vinden / weet niet wie data-houder is	16	46%
kan niet alle benodigde data via één platform vinden	13	37%
anders, namelijk:	9	26%



Algemeen	
Vraag / variabele	Welke verbeterpunten ziet u voor de overheid m.b.t. kwaliteit van open data?
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

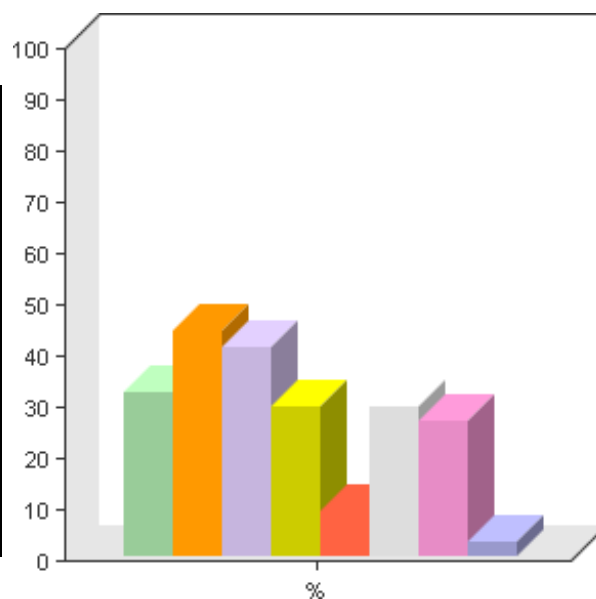
Welke verbeterpunten ziet u voor de overheid m.b.t. kwaliteit van open data?

(geen statistieken beschikbaar)

Algemeen	
Vraag / variabele	Hoe bent u nu bij het open data proces betrokken?
Beantwoord door	34 (97%)

Hoe bent u nu bij het open data proces betrokken?

Antwoordmogelijkheid	Aantal	Percentage
via een User Group	11	32%
via Social Media	15	44%
via nieuwsbrieven / e-mail / informele bijeenkomsten	14	41%
als data-houder	10	29%
als beleidsmaker	3	9%
ad-hoc, alleen als ik iets nodig heb	10	29%
anders, namelijk:	9	26%
helemaal niet	1	3%



Algemeen	
Vraag / variabele	Hoe zou u graag (meer) bij het open data proces betrokken willen worden?
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Hoe zou u graag (meer) bij het open data proces betrokken willen worden?

(geen statistieken beschikbaar)

Algemeen	
Vraag / variabele	Heeft u nog opmerkingen m.b.t. deze vragenlijst? Laat hier uw e-mailadres achter als u op de hoogte gehouden wilt worden van de uitkomsten van dit onderzoek
Type	Open
In vragenlijst	Vragenlijst Open Data gebruikers
Taalinstelling	Nederlands
Filter	Geen filter
Beantwoord door	35 (100%)

Heeft u nog opmerkingen m.b.t. deze vragenlijst? Laat hier uw e-mailadres achter als u op de hoogte gehouden wilt worden van de uitkomsten van dit onderzoek

(geen statistieken beschikbaar)

Kenniscentrum open data

Faculteit Bouwkunde, TU Delft
Julianalaan 134, 2628 BL Delft
Postbus 5030, 2600 GA Delft

Telefoon: +31 (0)15 278 30 05
E-mail: opendata-bk@tudelft.nl
www.kcopendata.eu