

OPEN DATA BEST PRACTICES

Bart Hanssens (BOSA)

OPEN DATA TASK FORCE opendata@belgium.be

14/10/2020

Contents

Inleiding	2
Wat is open data ?	2
Europese open data richtlijn en omzetting naar Belgisch recht	2
De federale open data strategie.....	3
De open data taskforce	3
Het federale open data portal data.gov.be.....	3
Praktische richtlijnen.....	4
Aanbestedingen en overdracht van eigendomsrechten	4
Open data en beheer van gegevens in het algemeen.....	4
Kwaliteit van gegevens.....	5
Interpretatie van gegevens	5
Gebruiksvoorwaarden en licenties.....	6
Berekenen en aanrekenen van vergoedingen	7
Publicatie en vermelding op data.gov.be.....	7
Bestandsformaten.....	8
Openstellen van gegevens via API's / web services	8
Contactgegevens	9
Vragen van burgers en bedrijven beantwoorden	9
Zelf open data (her)gebruiken.....	10
Bijkomende informatie.....	11
Linken	11

Inleiding

Wat is open data ?

Open data zijn gegevens die:

- geen persoonsgebonden of andere gevoelige informatie meer bevatten
- makkelijk herbruikbaar zijn in automatische processen
- gratis hergebruikt mogen worden (ook voor commerciële doeleinden)

Meestal worden deze gegevens gepubliceerd door overheidsdiensten, maar deze kunnen ook gepubliceerd worden door bedrijven, academische instellingen of particulieren.

Open data is eigenlijk niet nieuw, overheden publiceren al jaren (bijvoorbeeld) statistieken die door journalisten, onderzoekers of bedrijven worden hergebruikt. Maar open data kreeg vanaf +/- 2010 een stevige duw in de rug door enkele initiatieven van onder andere de Verenigde Staten en het Verenigd Koninkrijk.

Europese open data richtlijn en omzetting naar Belgisch recht

Europa had al langer een richtlijn over de toegankelijkheid van data van overheidsdiensten (de Public Sector Information), om de transparantie van overheden te bevorderen. Deze richtlijn werd herwerkt om ook het hergebruik van deze gegevens te stimuleren, hooguit tegen een kleine vergoeding en bij voorkeur zelfs helemaal gratis.

De redenering is dat overheidsdiensten werken met belastinggeld, zodat het verzamelen van gegevens al betaald wordt door burgers en bedrijven. Voor zover deze gegevens geen persoonsgebonden of ander gevoelige data bevatten, is het dan ook logisch dat de belastingbetalers erover kunnen beschikken, en zelf toepassingen kunnen maken.

Bedrijven die gegevens hergebruiken om succesvolle toepassingen te maken, creëren namelijk economische meerwaarde en betalen in principe dan ook weer belastingen op de extra winsten. Andere partijen, zoals academici of geëngageerde burgers, kunnen dan weer analyses of toepassingen maken met educatieve of sociale meerwaarde.

Federaal werd de tweede versie van de Europese PSI-Richtlijn omgezet naar de “Wet van 4 mei 2016 inzake het hergebruik van overheidsinformatie” en het “Koninklijk Besluit van 2 juni 2019 inzake het hergebruik van overheidsinformatie”. De omzetting van de derde versie is in voorbereiding.

Voor meer informatie, zie:

- <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>
- <http://www.ejustice.just.fgov.be/eli/wet/2016/05/04/2016009236/justel>
- <http://www.ejustice.just.fgov.be/eli/besluit/2019/06/02/2019013333/justel>

De federale open data strategie

Daarnaast heeft de federale overheid in 2015 een open data strategie opgesteld, waarin zowel vraaggerichte en proactieve initiatieven aangemoedigd worden. De drempel voor hergebruik moet ook zo laag mogelijk worden gehouden.

De open data taskforce

Om de federale overheidsdiensten te ondersteunen, heeft het KB een open data taskforce in het leven geroepen, die bestaat uit leden van de FOD BOSA DG Digitale Transformatie en de Dienst Administratieve Vereenvoudiging

Zij helpen onder andere de federale overheidsdiensten met juridisch en technisch advies, en werken samen met open data experts van de Gewesten en de Europese Commissie. Daarnaast staat de taskforce in contact met de Belgische open data community “OpenKnowledge Belgium”, waarin academici en bedrijven vertegenwoordigd zijn.

Daarnaast beantwoorden ze eenvoudige vragen van burgers en bedrijven, en sturen ze de meer gedetailleerde vragen door naar de overheidsdienst(en) die het best geplaatst zijn om te antwoorden. Dit kunnen zowel federale als regionale of lokale overheidsdiensten zijn.

De open data taskforce is te bereiken via het contactformulier op het federale open data portaal en via opendata@belgium.be.

Het federale open data portal data.gov.be

Het federale open data portaal bevat een overzicht van verwijzingen naar Belgische open data sets, ongeacht wie deze data publiceert (federale staat, provincies, gewesten, gemeentes, intercommunales, bedrijven ...) of waar de bestanden of web services beheerd worden.

De nadruk ligt hierbij op “verwijzingen” of linken: de open data bestanden worden dus niet gekopieerd naar 1 centraal platform. Dit zou ook praktisch niet haalbaar zijn, want de data is afkomstig van verschillende overheidsniveaus, en in een aantal gevallen moeten deze bestanden (ook) al verplicht ter beschikking worden gesteld via andere portalen en platformen.

Wel wordt de structuur van de metadata (titel, beschrijving, linken, thema’s) zoveel mogelijk gelijkgeschakeld volgens de Europese DCAT-AP norm. Deze metadata worden namelijk ook gedeeld met het Europese open data portaal, dat de open data sets van alle EU-lidstaten in kaart brengt.

Burgers en bedrijven, maar ook overheidsdiensten, kunnen op het portaal zoeken naar datasets, en kunnen vragen stellen via het contactformulier. Daarnaast bevat het federale portaal ook bijkomende documentatie, presentaties en verslagen van werkgroepen.

- <https://data.gov.be/nl/info-faq>

Praktische richtlijnen

Aanbestedingen en overdracht van eigendomsrechten

Overheidsdiensten kunnen natuurlijk geen gegevens als open data publiceren als ze geen eigenaar zijn van deze data. Aangekochte gegevens, of gegevens van andere overheidsdiensten, kunnen dus niet zomaar worden vrijgegeven.

Dit geldt ook voor gegevens die worden geproduceerd of verzameld in het kader van een overheidsopdracht: de overheidsdienst wordt na afloop niet automatisch eigenaar van de gegevens, ook niet van de “ruwe” data die vaak nodig is om tot het eindproduct of de dienst te komen. Zonder expliciete bepaling, worden deze gegevens geacht eigendom te blijven van de aannemer.

Aanbeveling: zorg ervoor dat in overheidsopdrachten een clausule staat die expliciet bepaalt dat de rechten op de producten en diensten - en alle “ruwe” data die verzameld of geproduceerd werden om de producten en diensten te realiseren - overgedragen worden aan de aanbestedende overheidsdienst.

Open data en beheer van gegevens in het algemeen

Zoals reeds aangehaald is open data niet echt nieuw, en zijn veel overheidsdiensten al vertrouwd met het publiceren en beheren van gegevens via bepaalde (internationale) normen.

Voor andere diensten kan “open data” de aanleiding zijn tot vragen en discussies die eigenlijk niets te maken hebben met “open”, maar wel met het beheer van data in het algemeen. “Wie is de eigenaar van de data? Wie kan meer vertellen over de inhoud en de manier waarop deze gegevens werden verzameld?” Dit zijn vragen die men kan – en zelfs moet – stellen bij het (her)gebruiken van data in het algemeen, zelfs als het gaat om hergebruik binnen de eigen overheidsdienst.

Aanbeveling: wees kritisch maar constructief, het openstellen van data kan een aanzet zijn om beter om te gaan met gegevens in het algemeen.

Kwaliteit van gegevens

Soms wordt “onvoldoende kwaliteit” aangehaald als een reden om gegevens niet als open data ter beschikking te stellen.

Het is echter moeilijk te verantwoorden waarom data niet goed genoeg zou zijn om te publiceren, maar wel goed genoeg is om intern te blijven gebruiken... Kwaliteit kan ook verschillen per toepassing: een onvolledig stratenplan kan voldoende zijn voor een toeristische gids, maar misschien niet voor het leveren van pakjes. Het is dan ook de verantwoordelijkheid van de hergebruiker om de kwaliteit in te schatten en te oordelen of deze voldoende is voor zijn/haar toepassing.

Aanbeveling: publiceer de data en vermeld (indien mogelijk) kort de gekende beperkingen. Stel daarnaast een plan op om de kwaliteit indien nodig te verbeteren, en sta daarbij open voor eventuele vragen en suggesties van externe partijen.

Interpretatie van gegevens

Ook “externen kunnen deze data niet correct interpreteren” wordt wel eens aangehaald als reden om gegevens niet als open data te publiceren.

De overheid heeft inderdaad heel wat expertise in huis, maar het is dan ook nodig om deze expertise helder te documenteren, zodat deze op z'n minst doorgegeven kan worden aan nieuwe medewerkers, zonder dat de continuïteit van de dienstverlening in het gedrang komt. En als deze documentatie dan toch beschikbaar is, dan kan deze net zo goed gedeeld worden met het grote publiek... Maar ook buiten de overheid zijn er heel wat experts die gegevens misschien vanuit een andere hoek kunnen bekijken, wat tot nieuwe inzichten of toepassingen kan leiden.

Aanbeveling: publiceer de data en verduidelijk (indien nodig) hoe gegevens verzameld werden en hoe ze geïnterpreteerd moeten worden.

Gebruiksvoorwaarden en licenties

Het Koninklijk Besluit voorziet in een aantal mogelijkheden wat betreft de gebruiksvoorwaarden via een “getrapt” model.

In regel wordt er geen enkele voorwaarde opgelegd: kosteloos hergebruik is dus toegestaan, ook voor commerciële doeleinden, en bronvermelding is daarbij niet verplicht. Dit is te vergelijken met de internationaal gekende “Creative Commons Zero (CC0)” verklaring.

Hiervan kan afgeweken worden, als er een goede reden wordt vermeld op de website van de overheidsdienst, én als deze motivering vooraf naar de open data taskforce werd verstuurd voor advies (anders is de licentie namelijk niet geldig).

Een eerste afwijking is de mogelijkheid om bronvermelding te verplichten, wat neerkomt op de internationale “Creative Commons Attribution (CC-BY)” licentie. Onder deze licentie is hergebruik dus nog steeds gratis, ook voor commerciële toepassingen.

Als dit niet volstaat, dan mag een overheidsdienst per uitzondering kiezen om “marginale kosten” aan te rekenen. Het KB bevat een opsomming van welke kosten er aangerekend mogen worden, maar bevat geen lijst van exacte bedragen. Als er dus geen andere wetten of besluiten bestaan die een vast tarief opleggen voor een bepaalde set van gegevens, dan moet de overheidsdienst zelf een berekening maken om het exacte bedrag te bepalen, en deze berekening publiceren. Deze optie wordt meestal afgeraden: in veel gevallen kost dit meer aan administratie (berekenen, voorstel sturen, factureren...) dan wat de verkoop van data zou kunnen opbrengen.

Als ook dit niet zou volstaan, dan is er nog de mogelijkheid om per uitzondering een andere licentie op te stellen. Ook deze manier van werken wordt meestal afgeraden: in het verleden werden er soms heel uitgebreide licenties opgesteld, die ofwel niet veel verschilden van de internationaal gekende Creative Commons licenties, ofwel in praktijk gewoon niet gecontroleerd of afgedwongen werden. Het correct opstellen van een eigen licentie kost een overheidsdienst tijd en moeite, en is ook niet klantvriendelijk voor hergebruikers, die vaak gegevens van verschillende bronnen willen combineren.

Aanbeveling: gebruik zoveel mogelijk de “Creative Commons Zero (CC0) verklaring of (na motivatie en mits positief advies van de taskforce) de “Creative Commons Attribution (CC-BY)”. Vermijd eigen licenties.

Berekenen en aanrekenen van vergoedingen

Het is niet verplicht - maar wel toegelaten - om een “marginale vergoeding” aan te rekenen voor het aanleveren van gegevens, en in een paar specifieke gevallen is de kostprijs in een Koninklijk of Ministerieel Besluit vastgelegd.

In de andere gevallen moet er een gedetailleerde berekening voorgelegd kunnen worden, zodat het berekenen en factureren uiteindelijk meer kan kosten dan wat het mogelijk opbrengt. Daarbij komt nog dat de meeste aanvragen afkomstig zijn van burgers en bedrijven die belastingen betalen, en dus al een financiële bijdrage leveren aan de werking van de overheid. Gegevens moeten dan ook zo veel mogelijk gratis gepubliceerd worden.

Deze redenering geldt ook in omgekeerde richting: het is niet omdat een bedrijf geld kan verdienen met diensten of combinaties van gegevens die gebaseerd zijn op open data, dat de overheid daarom kosten mag aanrekenen.

Tenzij in zeer specifieke gevallen, is het afsluiten van exclusiviteitscontracten verboden.

Aanbeveling: reken geen kosten aan voor het ter beschikking stellen van gegevens. Als dit toch absoluut noodzakelijk is, publiceer dan de gedetailleerde berekening om de kostprijs te rechtvaardigen.

Publicatie en vermelding op data.gov.be

Het federale open data portaal data.gov.be verzamelt linken en andere metadata van open data bestanden. De bestanden zelf worden niet gekopieerd, deze blijven gewoon op de website of het portaal van de overheidsdiensten staan.

Het is uiteraard niet de bedoeling om deze linken, titels... die al beschreven werden, nog eens manueel in te voeren op data.gov.be: deze informatie moet zoveel mogelijk automatisch overgenomen kunnen worden.

Bij voorkeur gebeurt dit via het aanmaken en uitwisselen van een “DCAT-AP” metadata-bestand, maar afhankelijk van het systeem waarin de gegevens beheerd worden of de waarmee de website gebouwd werd, kan dit ook via een API, newsfeed of – in het slechtste geval – door het verwerken (“scrapen”) van een overzichtspagina waarop alle open data sets vermeld staan. Contacteer de open data task force, zodat zij samen met uw technische medewerkers kunnen bekijken om deze koppeling te realiseren.

Aanbeveling: contacteer de open data task force om af te spreken op welke manier de linken naar datasets (automatisch) gepubliceerd en bijgewerkt kunnen worden op het federale open data portaal.

Bestandsformaten

Er bestaat geen “generiek open data formaat”: net zoals er geen generiek formaat bestaat voor alle toepassingen.

Statistieken worden bijvoorbeeld vaak in XLSX of CSV gepubliceerd, voor geografische informatie wordt vaker ShapeFile gebruikt, en GTFS is populair voor dienstregelingen van het openbaar vervoer.

Het is belangrijk dat de gegevens beschikbaar zijn in een formaat dat makkelijk automatisch te verwerken is: zo is het voor machines heel moeilijk om op een gestructureerde manier gegevens te halen uit een PDF- of DOCX-bestand met beelden en tabellen, of uit kaarten met alleen maar kleurcodes.

Het is natuurlijk nog steeds toegestaan om deze rapporten, kaarten of infographics te publiceren, maar zorg en dan ook voor dat de “ruwe” data beschikbaar is in een formaat zoals CSV, JSON, KML, ShapeFile, sqlite, TTL, XLSX of XML.

Aanbeveling: publiceer de “ruwe” gegevens waarop rapporten en kaarten gebaseerd zijn ook in gekende formaten die makkelijk automatisch te verwerken zijn.

Openstellen van gegevens via API's / web services

Naast downloadbare bestanden, vragen gebruikers soms ook naar de beschikbaarheid van API's of web services, zodat de gegevens makkelijker geïntegreerd kunnen worden in apps of eigen systemen.

Dit is vooral van belang bij gegevens die erg vaak wijzigen, of grote verzamelingen data waar de gebruiker maar een kleine deel van wenst te gebruiken, zoals bijvoorbeeld treinvertragingen of geografische gegevens. Het is niet haalbaar – en ook niet nodig – om een API te voorzien voor alle mogelijke datasets, er zal dus altijd een afweging gemaakt moeten worden tussen de kost om de API te ontwikkelen en onderhouden, en de meerwaarde of besparingen die gebruikers kunnen realiseren door het gebruik ervan.

Algemeen bekeken, is het aanbieden van een API alvast nuttig als er op een website al een zoekscherm aangeboden wordt, waarmee gebruikers manueel opzoeken kunnen doen. In bepaalde gevallen kan men de API en het zoekscherm via aparte kanalen laten verlopen, zodat het zoekscherm blijft functioneren als de API (te) zwaar belast zou worden.

Het blijft in heel wat gevallen ook aan te raden om – naast de API – regelmatig een volledig bestand als download aan te bieden, zodat de gebruiker kan kiezen. Een goed voorbeeld zijn de statistieken van infocenter.belgium.be, die zowel als download als via API beschikbaar zijn.

Aanbeveling: evalueer bij het bouwen van nieuwe systemen de mogelijkheid om een API te voorzien, zodat gebruikers deze kunnen integreren in eigen systemen. Een vuistregel: als een online databank een zoekscherm heeft om gegevens manueel op te zoeken, dan is het nuttig om ook een API te voorzien om opzoeken te kunnen automatiseren.

Contactgegevens

Net zoals voor klassieke brochures of andere informatie op overheidswebsites, moet het voor elke dataset duidelijk zijn via welk kanaal gebruikers vragen kunnen stellen.

Bij voorkeur wordt er dan ook een emailadres gepubliceerd op de plaats waar de datasets ter beschikking worden gesteld. Gebruik daarvoor een algemeen adres of een mailgroep in plaats van 1 specifieke persoon, zodat eventuele vragen ook tijdens vakantie of afwezigheid opgevolgd kunnen worden. Dit mag hetzelfde adres zijn als voor andere vragen

Een alternatief is om een link te voorzien naar een algemeen of specifiek contactformulier.

Aanbeveling: zorg voor een emailadres waar burgers, bedrijven en andere overheidsdiensten vragen over datasets naartoe kunnen sturen.

Vragen van burgers en bedrijven beantwoorden

Burgers en bedrijven (of andere overheidsdiensten) kunnen dus vragen stellen over reeds gepubliceerde gegevens, of kunnen op zoek zijn naar gegevens waarvan ze vermoeden dat uw dienst deze bezit.

Het is best mogelijk dat uw dienst niet over deze gegevens beschikt: laat dit de aanvrager dan ook zo snel mogelijk weten. Als de vraag eigenlijk niet voor uw dienst bedoeld is, probeer de vraag dan zelf door te sturen naar de juiste dienst (dit kan ook een dienst van een regionale of lokale overheid zijn), of stuur de vraag door naar de federale open data task force (opendata@belgium.be), en informeer de aanvrager naar wie de vraag is doorgestuurd.

Informeer de aanvrager ook als de vraag niet meteen duidelijk is, of als het beantwoorden ervan lang kan duren. Volgens het KB moet de vraag immers binnen 10 werkdagen (30 bij een meer complexe vraag) beantwoord zijn, maar een korte ontvangstbevestiging komt professioneel over.

Het antwoord kan uiteraard ook negatief zijn: als de gegevens geen eigendom zijn van de overheid, of persoonlijke gegevens bevat, of uit veiligheidsoverwegingen niet vrijgegeven kan worden, dan kan de aanvrager de gegevens niet verkrijgen. In dat geval moet de aanvrager wel op de hoogte gebracht worden waarom de gegevens niet worden vrijgegeven.

De overheid is niet verplicht om “onredelijk” veel inspanningen te doen: het is niet de bedoeling om heel specifieke verwerkingen te doen om tot een gepersonaliseerde dataset te komen.

Aanbeveling: stuur bedrijven en burgers een korte ontvangstbevestiging. Behandel de vraag (indien mogelijk) binnen de 10 werkdagen, en vermeld de reden als het antwoord negatief is. Als de vraag niet voor uw dienst bedoeld is, stuur de vraag dan door naar de juiste dienst of naar opendata@belgium.be.

Zelf open data (her)gebruiken

Overheidsdiensten publiceren niet alleen open data, ze kunnen op hun beurt ook open data van andere diensten hergebruiken.

Een mooi voorbeeld hiervan zijn de “BeST” (Belgian Streets) adresbestanden: in samenwerking met de 3 Gewesten, publiceert de federale overheid een reeks bestanden met alle straatnamen en geldige huisnummers als open data, en deze bestanden worden op hun beurt gebruikt om web services aan te bieden aan federale overheidsdiensten.

Aanbeveling: het hergebruik van open data door overheidsdiensten wordt sterk aangemoedigd. Hoewel het niet verplicht is, is het aan te raden om de overheidsdienst te contacteren die deze data publiceert. Op die manier krijgen zij ook een beter overzicht van het gebruik en eventuele specifieke noden.

Bijkomende informatie

Linken

België

- Open Data Task Force: opendata@belgium.be
- Federaal open data portaal: <https://data.gov.be>
- Verslagen en presentaties werkgroep portalen: <https://data.gov.be/nl/dcat-werkgroep>
- Wet inzake het hergebruik van overheidsinformatie:
<http://www.ejustice.just.fgov.be/eli/wet/2016/05/04/2016009236/justel>
- Koninklijk Besluit inzake het hergebruik van overheidsinformatie:
<http://www.ejustice.just.fgov.be/eli/besluit/2019/06/02/2019013333/justel>

Europa:

- Europees data portaal: <https://www.europeandataportal.eu/nl>
- Europese open data richtlijn: <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>