



# COOCK API-ECONOMIE Best Practice Guide "Open data"



## Inhoud

<b>1</b>	<b>Wat is Open Data?</b> .....	3
1.1	Open versus FAIR? .....	3
1.2	Voordelen van Open Data.....	3
1.3	Checklist "starten met gebruik van Open Data" .....	4
1.4	Private Open Data? .....	6
<b>2</b>	<b>Wat is Linked Data?</b> .....	7
<b>3</b>	<b>Wat met privacy in Open Data?</b> .....	8
3.1	Data anonimiseren en pseudonimiseren .....	8
3.2	Als anonieme Open Data niet voldoende is... Code of Conduct en DjustConnect.	9
<b>4</b>	<b>E-learning en nuttige links</b> .....	9
<b>5</b>	<b>Bronnen</b> .....	10

# 1 Wat is Open Data?

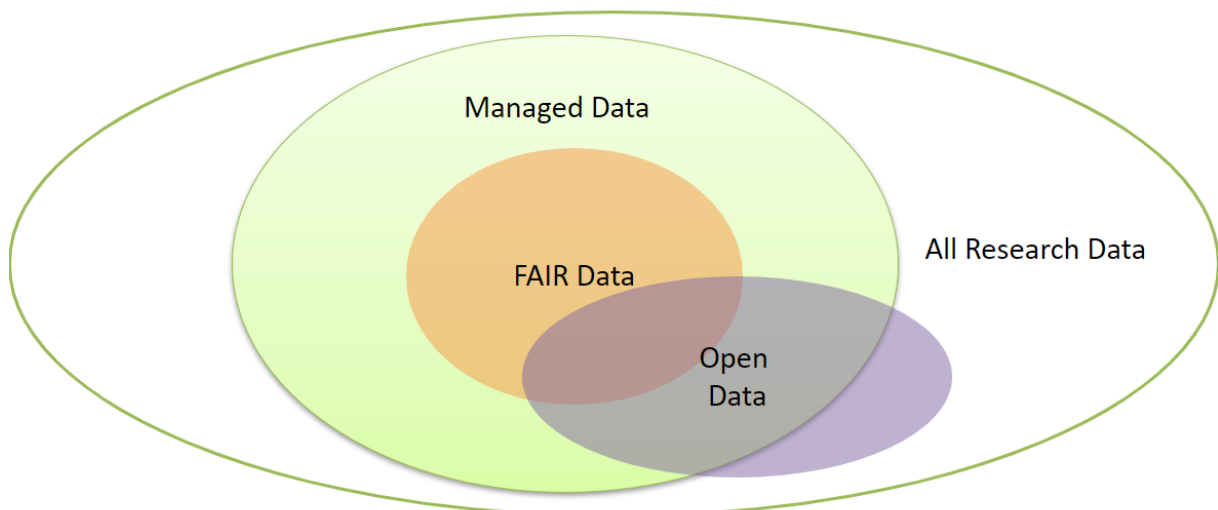
Open data zijn gegevens die worden verzameld en waar geen of minimale gebruiksbependingen op rusten. Ze zijn elektronisch beschikbaar en maken bij voorkeur gebruik van open standaarden.

Het gebruik van open data verhoogt niet alleen de transparantie, maar het maakt ook de weg vrij voor innovatie en de ontwikkeling van nieuwe producten of diensten.

Het is belangrijk op te merken dat niet alle informatie van de publieke sector Open Data is.

## 1.1 Open versus FAIR?

Naast open data wordt de term FAIR data vaak gebruikt. FAIR data (**F**indability, **A**ccessibility, **I**nteroperability, and **R**euse) wil zeggen dat de data vindbaar, toegankelijk, interoperabel en herbruikbaar is. In **Fout! Verwijzingsbron niet gevonden.** zien we een visualisatie van verschillende data. In deze grafiek kan gezien worden dat Open data en FAIR data niet dezelfde begrippen zijn. Waardevolle Open data kan bijvoorbeeld niet teruggevonden worden, en FAIR data hoeft niet noodzakelijk Open te zijn.



Figuur 1 visualisatie van Open Data

## 1.2 Voordelen van Open Data

De voordelen van Open Data zijn divers en variëren enorm. De kern van Open data bevat onder meer een verbeterde efficiëntie van het openbaar bestuur, economische groei in de particuliere sector en het verbeteren van de maatschappelijke welvaart (zie Figuur 2)

- Open data zorgt voor grotere transparantie over de werking van de overheid.
- Open data zorgt ook voor meer efficiëntie zowel binnen en tussen overheden als bij bedrijven en organisaties extern aan de overheid.

- Open data leidt tot innovatie en het ontstaan van nieuwe en vernieuwende producten en diensten.
- De maatschappelijke welvaart kan worden verbeterd, omdat de samenleving baat heeft bij informatie die transparanter en toegankelijker is.
- De economie kan voordeel halen uit een betere toegang tot informatie, inhoud en kennis en zo bijdragen aan de ontwikkeling van innovatieve diensten en het creëren van nieuwe bedrijfsmodellen



Figuur 2 Voordelen van Open Data

### 1.3 Checklist "starten met gebruik van Open Data"

Er zijn verschillende belangrijke stappen die doorlopen moeten worden voordat data kan gebruikt worden:

1. Het doel

**Definieer jouw doel:** Er is veel Open Data beschikbaar o.a. demografische gegevens, geografische gegevens, marktinformatie, ... in diverse domeinen onderwijs, toerisme, handel, erfgoed, wonen, mobiliteit, zorg en gezondheid, maar ook specifiek landbouw, klimaat, milieu, water. Zie onderstaande links om deze types Open Data te vinden:

- [Vlaamse Landbouw en Visserij in cijfers](#)
- [Marktinformatie](#) Vlaamse Landbouw en Visserij
- [Geopunt](#) met geografische informatie
- [Vlaams Open Data Portaal](#)
- [Provincies in cijfers](#)

Vaak bevatten deze sites dezelfde datasets en is het even zoeken. Digitaal Vlaanderen bouwt daarom aan een centrale [Datavindplaats](#) die toegang geeft tot het meest uitgebreide aanbod aan informatie.

Er zijn ook portalen waar Europese Open Datasets te vinden zijn. [Data.europ.eu](http://Data.europ.eu) is het officiële portal voor Europese data.

De Open Data kan voor meerdere doelstellingen gebruikt kan worden. Op het eerste zicht lijken vooral de laatste categorieën interessant voor agrovoedingsbedrijven, maar vaak ligt een grotere toegevoegde waarde in de combinatie van meerdere datasets. Zo kan Open Data je een beter inzicht verschaffen op gebied van marketing, supply chain management, strategische planning en business intelligence. Je kan de verkregen informatie ook toevoegen aan je bestaand digitaal product of service om toegevoegde waarde te bieden aan jouw klanten.

**Check de metadata en de data labels:** Als je weet voor welk doel je data nodig hebt, dan is het belangrijk om te kijken of de data aansluit bij je behoeften en dus bij je doeleinde, door te kijken naar de metadata en datalabels.

Over welke gegevens gaat het? Soms heb je meer baat bij geaggregeerde gegevens vb. het aantal huishoudens met kinderen in een bepaalde regio of de ruimtelijke spreiding van kippenexploitatie. Maar soms heb je nood aan individuele gegevens over vb. de hoofdteelt op een perceel en ben je niets met het huidige areaal aan aardappelen in Vlaanderen. Deze informatie noemt men metadata.

Metadata zegt dan weer iets over de data zelf. Het zijn gegevens die de karakteristieken van de data(set) beschrijven. Metadata is dus data over de data. Deze metadata is cruciaal om de dataset die je wil gebruiken correct te interpreteren zonder hulp van de persoon, sensor of technologie die de data heeft gecreëerd.

Belangrijke vragen hierbij zijn:

- Hoe zijn de gegevens gekomen? Gaat het over een al dan niet betrouwbare streekproef vb. gegevens gekomen vanuit het [landbouwmonitoringsnetwerk](#) van het Departement Landbouw en Visserij of een enquête uitgevoerd bij x aantal consumenten.
- Hoe zijn de gegevens verwerkt?
- Zijn het ruwe gegevens of samenvattingen?
- Hoe zal de vorm jouw analyse/product/toepassing beïnvloeden?
- Welke syntactische (taal) en semantische (betekenis) transformaties zal je moeten maken?
- Is dit compatibel met andere datasets die je hebt?

Naast de metadata, kan aan de gegevens nog een tweede soort informatie toegevoegd zijn, nl. data labels. Door data te labelen worden de ruwe gegevens (ongelabelde data) uitgebreid met een tag, label of klasse die betekenis of informatie toekent aan de gegevens. Op een foto genomen met een thermale camera van [een achterpoot/klauw](#) van een melkkoe wordt een bepaalde zone gelabeld als Mortellaro letsel. Betekenisvolle en interpreteerbare data wordt dus toegevoegd aan een thermaal beeld met temperaturen. Data labels zijn niet enkel bijzonder nuttig, maar verhogen ook de toegevoegde waarde van de dataset.

## 2. Open vergunning (licentie)

**Controleer de openheid:** Kijk naar de licentie-informatie die over de dataset wordt verstrekt. Zorg ervoor dat er een licentie beschikbaar is die je toestaat de gegevens te gebruiken op de manier die je van plan bent vb. is commercieel hergebruik toegestaan? Veel voorkomende licenties zijn de [Creative Commons licenties](#). Aan de hand van deze licentie wordt bepaald hoe data mag hergebruikt worden.

**Controleer de vereisten inzake naamsvermelding:** Het is mogelijk dat in de licentie staat dat mensen die de data (her)gebruiken, degene die de data publiceert moeten crediteren, wat betekent dat je de eigenaar moet crediteren op de voorgestelde manier, wanneer je jouw product of dienst beschikbaar stelt. Dit wordt attributie genoemd.

**Controleer de Share-Alike vereisten:** Dit wil zeggen dat als je de data verrekt, integreert of verwerkt met andere data, je de verplichting hebt om ook de geaggregeerde data of de resultaten vrij te geven als Open Data en je eigen gegevens onder een soortgelijke licentie te publiceren nadat je andere gegevens hebt toegevoegd aan de oorspronkelijke bron. Zorg er dus voor dat de licentie in overeenstemming is met je beoogde doel voor het (her)gebruik van de data.

### 3. Bestandsformaat

Nadat je hebt besloten dat een specifieke dataset precies is wat je zoekt, kan je waarschijnlijk kiezen om de datasets te downloaden waarbij kan gekozen worden tussen verschillende opties. Afhankelijk van je computervaardigheden, kan je het bestandstype kiezen dat het meest geschikt is. Het meest gebruikelijke bestandsformaat voor gegevens in tabelvorm is '.csv'. Hierbij kan je andere informatie aan het bestand toevoegen of berekeningen met de data maken. Datasets die kunnen worden aangepast, worden bij voorkeur gepubliceerd in een open bestandsformaat (bijvoorbeeld .csv in plaats van .xls). De meeste datasets zijn beschikbaar in een open bestandsformaat, maar houd er rekening mee dat sommige formaten (bv. '.pdf') niet kunnen worden aangepast, geopend worden of moeilijker verwerkbaar zijn.

### 4. Kwaliteit van de data

Op de pagina waar je de dataset wilt downloaden, moet je informatie vinden over de laatste datum waarop het bestand is gewijzigd. Indien je gegevens voor een bepaalde periode nodig hebt, moet je nagaan of de informatie over de periode wordt verstrekt of dat deze onlangs is bijgewerkt (en er bijgevolg andere versies in omloop zijn). Je moet nagaan of de informatie die je in het bestand verwachtte aan te treffen, ook daadwerkelijk is opgenomen en of je de verschillende labels begrijpt.

Belangrijke vragen hierbij zijn:

- Hoe actueel zijn de gegevens?
- Hoe regelmatig worden ze bijgewerkt?
- Begrijp je alle velden en hun context?
- Voor hoe lang worden ze gepubliceerd?
- Wat is de verbintenis van de uitgever?
- Wat weet je over de nauwkeurigheid van de data?
- Hoe wordt omgegaan met ontbrekende data?

Sommige websites geven je informatie over de kwaliteit van de (meta)data. Zo beoordeeld [Data.europa.eu](http://Data.europa.eu) de metadata op:

- Vindbaarheid vb. trefwoorden beschikbaar of zoeken in de tijd mogelijk?
- Toegankelijkheid vb. download URL actief?
- Interoperabiliteit vb.
- Herbruikbaarheid vb. formaat, type, al dan niet machineleesbaar
- Contextgebonden parameters vb. is de datum van beschikbaarstelling gekend? Datum bewerking gekend? Is informatie over bestandsgrootte en rechten beschikbaar?

## 1.4 Private Open Data?

Als er gesproken wordt over Open Data, wordt meestal gedacht aan gegevens of informatie afkomstig van overheidsinstanties, net omwille van hun verplichting in kader van openbaarheid van bestuur. Toch is het ook mogelijk voor private bedrijven om hun gegevens te delen als Open Data.

Een gekend voorbeeld van een privaat Open Data portaal komt van het peer-to-peer ridesharing, voedsel levering en transport netwerk bedrijf Uber. Het Uber Movement Open Data portaal biedt geanonimiseerde gegevens van meer dan twee miljard vervoersbewegingen voor niet-commercieel hergebruik. Met name stedenbouwkundigen

hebben baat bij de beschikbaarheid van dit soort gegevens. Zij kunnen hun plannen en beleid verfijnen en aanpassen met deze private Open Data.

Ook heel wat bedrijven in de energiesector stellen gegevens over energie ter beschikking die nuttig kunnen zijn voor verschillende partijen in het energielandschap. Op deze manier kunnen ze tonen van waar je energie komt, hoeveel en wat verbruikt wordt, hoeveel groene energie... wat naamsbekendheid en betrouwbaarheid (transparantie) opwekt.

Potentieel kunnen publieke Open Data en private Open Data worden gecombineerd en leiden tot meer diepgaande en verrijkte kennis.

## 2 Wat is Linked Data?

Om het begrip Linked Data goed te verstaan, is het belangrijk om het belang ervan uit te leggen. Iedereen kent het Internet. Het internet is een verzameling van web pagina's die traditioneel (ongestructureerde) informatie, meestal tekst, bevatten. Gebruikers van het internet klikken door om meer informatie te verkrijgen over een bepaald onderwerp. We linken dus tijdens het lezen en verwerken in ons hoofd de informatie die we vinden op 1 plaats aan extra informatie die we vinden op een andere plaats. Dit laatste is nu exact wat computers niet kunnen. Een zoekmachine kan een pagina lezen en kijken welke andere webpagina's door middel van links vanop de originele pagina gemaakt zijn. Maar deze zoekmachine is niet in staat om te verstaan hoe de stukken informatie op verschillende pagina's onderling gerelateerd zijn, terwijl net die relaties betekenis geven aan de informatie.

Om computers de relaties te laten begrijpen, moeten wij als mensen een gegevensstructuur toevoegen aan de gegevens. Concreet komt het erop neer dat mensen stukjes informatie deconstrueren en tags toevoegen die wel door computers kunnen begrepen en gevonden worden. Nog belangrijker is het om nadien de relaties tussen de tags vast te leggen. Zo bekom je een grote grafiek van geconnecteerde tags. Hierdoor weet een computer welke stukjes tekst waar precies over gaan en kan de zoekmachine via de grafiek de relaties kennen. Linked Data is een methode of best practice om deze koppelingen tussen gegevens mogelijk te maken.

Linked Data levert dus een reeks design principes voor het delen van machinelesbare, gekoppelde data beschikbaar op het web. Vier ontwerpprincipes voor linked (Open) data werden beschreven (Berners-Lee):

1. Ken aan elk gegeven een unieke naam toe. Hiervoor wordt gebruik gemaakt van URI's (Uniform Resource Identifiers). Door aan woorden unieke namen toe te kennen, kan een computer of machine weten dat een woord in tekst 1 hetzelfde is dan een ander woord in andere tekst.
2. Gebruik HTTP URI's zodat gebruikers deze unieke namen kunnen opzoeken op het internet.
3. Om ervoor te zorgen dat elke URI bruikbare informatie geeft, wordt gebruik gemaakt van standaarden. In de eerste plaats de RDF (Resource Description Framework) standaard. Dit is een formaat dat gebruikt wordt voor het publiceren van gegevens op het internet gebaseerd op een grafische voorstelling van deze gegevens. Om de in RDF formaat gepubliceerde gegevens op te vragen, wordt dan weer gebruik gemaakt van gestandaardiseerde querytaal, nl. SPARQL.
4. Net zoals je links zou toevoegen aan een webpagina, kan je tenslotte aan elke URI links naar andere URI's toevoegen zodat meer gegevens geconnecteerd worden. Door nieuwe informatie met bestaande bronnen te verbinden, maximaliseren we het hergebruik en de koppeling tussen

bestaande gegevens en creëren we een rijk onderling verbonden netwerk van machinaal verwerkbaar informatie.

Wanneer bij Linked Data gebruik wordt gemaakt van Open data wordt gesproken van linked Open data (LOD). Vandaag is dat in de meeste datasets het geval. Open Data wordt echt als de voorloper van linked data gezien. Linked Open data heeft als doel het internet te verrijken door open datasets nog beter te ontsluiten via de linked-datamethode. Dit ontsluiten wordt voorgesteld als een wolk van gekoppelde datasets, ook wel een kennisgrafiek (knowledge graph) genoemd.

### 3 Wat met privacy in Open Data?

#### 3.1 Data anonimiseren en pseudonimiseren

Voor overheidsorganisaties is er een verplichting om overheidsinformatie openbaar te maken als er een beroep op de Wet Openbaarheid van Bestuur (Wob) wordt gedaan. Vele overheidsinstanties proberen dan ook zo veel mogelijk van hun data te ontsluiten. Dit gebeurt uiteraard zonder daarbij de bescherming van de persoonsgegevens uit het oog te verliezen.

Met privacygevoelige data of persoonsgegevens hebben bedrijven en instellingen de beschikking over gegevens die uiterst kwetsbaar voor misbruik zijn. Dit is ook de reden dat er strenge wetgeving bestaat rondom het waarborgen van de privacy. Deze wetgeving is onder andere vastgelegd in de Algemene Verordening Gegevensbescherming (AVG). Dit betekent dat je zeer zorgvuldig met privacygevoelige informatie moet omgaan. Dit betreft gegevens, zoals onder meer de naam en adresgegevens, het Rijksregisternummer en bank- en creditcardgegevens. De overheid, maar ook bedrijven of andere instellingen zijn verplicht om deze data te anonimiseren of pseudonimiseren, als je bepaalde gegevens in de openbaarheid brengt.

Bij pseudonimisering kunnen anonieme gegevens met de juiste sleutel inzichtelijk gemaakt worden, waardoor herleiding naar natuurlijke personen weer mogelijk is. Anonimisering draait daarentegen om versleuteling die niet omkeerbaar is: als persoonsgegevens eenmaal zijn ontdaan van identificerende data, is het niet meer mogelijk om deze later weer met personen in verband te brengen.

Het verschil tussen geanonimiseerde en gepseudonimiseerde gegevens is belangrijk, omdat de AVG niet van toepassing is op volledig geanonimiseerde data. Voor gepseudonimiseerde data gelden de regels uit de AVG nog wel.

**Anonimiseren** is het proces waarbij je ervoor zorgt dat de gegevens van een persoon verwijderd worden zodat deze niet meer gebruikt kunnen worden om de betreffende persoon te identificeren. Hierbij kan nooit meer een link gelegd worden naar de originele gegevens in het onderzoek.

Vaak toetsen we een anoniem resultaat aan de hand van drie criteria:

- Herleidbaarheid, oftewel de mogelijkheid om een persoon direct te identificeren
- Koppelbaarheid, de mogelijkheid om andere gegevens in verband te brengen met een persoon wat tot een identificatie kan leiden
- Deduceerbaarheid, de mogelijkheid om informatie af te leiden wat gebruikt kan worden om een persoon te identificeren.

Op het moment dat een toetsing van de drie bovenstaande criteria dus niet kan leiden tot een identificatie van een persoon dan kun je concluderen dat de data voldoende anoniem



is. Geanonimiseerde gegevens vallen niet meer onder de AVG-wetgeving en kunnen vervolgens geëxporteerd worden voor verdere verwerking buiten de beveiligde omgeving. Het is de verantwoordelijkheid van de persoon of instantie die de gegevens ontsluit om te verzekeren geen privacygevoelige data te exporteren buiten de beveiligde silo.

Om gegevens te **pseudonimiseren** moeten alle verwijzingen naar personen in databestanden worden vervangen door een gegenereerde code. Een codelijst bevat dan de koppeling tussen de code en een persoon. Deze codelijst moet veilig worden bewaard. Dit kan binnen een beveiligde omgeving. De beveiligde omgeving bevat ook mogelijkheden om zulke codelijsten te genereren. Gepseudonimiseerde gegevens vallen zoals eerder vermeld nog steeds onder de AVG-wetgeving.

### 3.2 Als anonieme Open Data niet voldoende is... Code of Conduct en DjustConnect

In sommige gevallen heb je als bedrijf niet voldoende informatie ter beschikking door het gebruik van een Open Data bron. Als het jouw doel is om individuele Open Data van vb. een landbouwer te koppelen aan de gegevens die jij reeds hebt van deze landbouwer, kom je met Open Data niet ver genoeg.

Zoals hierboven beschreven valt de niet anonieme data onder de AVG-wetgeving, beter bekend als GRPR-wetgeving. Het delen en gebruiken van deze gegevens is mogelijk, maar enkel onder wettelijk bepaalde specifieke condities. Een van de opties is het verkrijgen van een uitdrukkelijk toestemming van de betrokken persoon.

Volgens de wetgeving zijn niet-persoonsgebonden gegevens, gegevens niet geen betrekking hebben op een geïdentificeerde of identificeerbare natuurlijke persoon of op persoonsgegevens die zodanig anoniem zijn gemaakt dat de betrokkene niet of niet meer identificeerbaar is". In de praktijk is het niet zo eenvoudig om het onderscheid te maken tussen persoonsgegevens en niet-persoonsgegevens. Daarenboven ervaren de mensen een bepaald gegeven toch als persoonlijk of hun persoonlijk eigendom of op zijn minst gevoelige informatie terwijl dit gegeven misschien strikt genomen niet onder de GDPR wetgeving valt. Om het data delen van niet-persoonsgegevens in landbouw te regelen, werd in 2018 een [Code of Conduct](#) voor het delen van data in landbouw opgesteld. Deze vrijwillige code erkent de landbouwers als eigenaar van de ruwe data en beschrijft hoe een landbouwer correct moet geïnformeerd worden over het data delen. Kijk [hier](#) voor meer informatie.

Om het delen van niet-anonieme dataset in de agrovoedingssector te faciliteren, is er in Vlaanderen het datadeelplatform [DjustConnect](#). DjustConnect verzekert veilig en transparant delen van gegevens na het verkrijgen van de juiste toestemmingen van de data-eigenaar, die in vele gevallen de landbouwers is. 100% Code of Conduct proof!

## 4 E-learning en nuttige links

Nuttig e-learning tools of tutorials beschikbaar voor open data management:

- [https://www.ted.com/talks/pieter\\_colpaert\\_open\\_data\\_to\\_create\\_power\\_for\\_the\\_many\\_not\\_the\\_few](https://www.ted.com/talks/pieter_colpaert_open_data_to_create_power_for_the_many_not_the_few)
- Webinarreeks DjustConnect: We produceren voeding...en DATA!

- Sessie 4: Spelregels voor datadelen, hoe zit het juridisch? (Code of Conduct)
  - <https://www.youtube.com/watch?v=RmVLNUVggbU&feature=youtu.be>
- <https://lisboncouncil.net/publications/a-roadmap-for-a-fair-data-economy/>
- Open data in precisielandbouw in Nederland: <https://edepot.wur.nl/400223>

## 5 Bronnen

<https://data.europa.eu/en/trening/what-open-data>

<http://opendefinition.org/>

<https://www.vlaanderen.be/digitaal-vlaanderen/onze-oplossingen/open-data>

<https://www.dataheroes.eu/academy/wat-is-anonimiseren-nu-precies-en-hoe-pas-je-het-toe/>

[https://www.eanonimiseren.nl/data-anonimiseren/?qclid=CjwKCAiApvebBhAvEiwAe7mHSGL8\\_K023Lh5JBwhMjElcdBMHZc](https://www.eanonimiseren.nl/data-anonimiseren/?qclid=CjwKCAiApvebBhAvEiwAe7mHSGL8_K023Lh5JBwhMjElcdBMHZc)

[bhhQMdrbOFVpLTNMNJ9VBw\\_YEgRoCowwQAvD\\_BwE](https://www.eanonimiseren.nl/data-anonimiseren/?qclid=CjwKCAiApvebBhAvEiwAe7mHSGL8_K023Lh5JBwhMjElcdBMHZc_bhhQMdrbOFVpLTNMNJ9VBw_YEgRoCowwQAvD_BwE)

<https://admin.kuleuven.be/icts/onderzoek/beveiligdeomgeving/anonimiseren>

<https://digital-strategy.ec.europa.eu/en/policies/private-sector-data-sharing>

<https://edepot.wur.nl/400223>