



# HOW TO WORK WITH DATA

Ilse Vogel, Maurice Pelt, Arjan Koning, Nanda Piersma.

Data studio / Faculteit Techniek

1.0

30 augustus 2019



## 1. Data verzamelen en verwerken

Data is een bron voor heel veel informatie en in toenemende mate wordt data onderdeel van (onderzoeks)projecten en bedrijfsuitoefening.

Het werken met data geeft verplichtingen en verantwoordelijkheid. Dit rapport geeft een kort overzicht waar je aan moet denken bij onderzoek waar datasets worden gebruikt en geeft inzicht in de belangrijkste zaken rondom privacy en data. Doel is om te helpen omgaan met data in onderzoek en te voldoen aan de regelgeving. Altijd geldt dat afspraken worden vastgelegd in een datamanagementplan (DMP) dat je maakt in overleg met de data eigenaar. Het datamanagementplan maak je voor de start van het onderzoek. Het is een levend plan dat onderhouden moet worden gedurende het onderzoek. Zie hiervoor: <https://rdm.uva.nl/plannen/datamanagementplan/datamanagementplan.html>

Er is een aantal tech reports die je helpen om de data technisch op de juiste plaats te krijgen en om analyses uit te voeren. Zie hiervoor <https://urbananalytics.nl/techreports> of [datastudio.amsterdam](https://datastudio.amsterdam)

### 1.1 Datasets

Er zijn drie manieren om data te verkrijgen:

- Data zelf creëren
- Data van anderen gebruiken
- Open data gebruiken

#### 1.1.1 Data zelf creëren

Als je zelf data verzamelt dan word je **eigenaar** van de data. Je bent verantwoordelijk voor de kwaliteit van de verzamelde gegevens, voor het gebruik (**verwerken en analyseren**) en het verspreiden van de gegevens (wie mag de gegevens inzien en gebruiken).

**Als je data verzamelt van apparatuur** (sensoren, machines etc) dan is de data van de eigenaar van de apparatuur. Ben jij de eigenaar van de apparatuur, dan word je daarmee ook eigenaar van de data. Contractueel kan dit anders afgesproken worden, waarbij de data die een apparaat genereert eigendom wordt van de maker. Bijvoorbeeld als de maker de enige is die iets zinvol met de data kan doen of om commerciële redenen. Let op: de HvA is feitelijk eigenaar van de apparatuur en dus van de data; jij bent de persoon die met de data werkt en daarmee dataverantwoordelijke namens de HvA.

**Als je data verzamelt van zaken die in de openbare ruimte staan of daar plaatsvinden** (temperatuur, aantal mensen in groene ruimtes, vuilnis aangetroffen op straat) dan zijn deze gegevens van jou (de HvA), je wordt eigenaar van de gegevens, ofwel de dataverantwoordelijke. De meeste openbare ruimtes bevinden zich in de open lucht, maar ook vrij toegankelijke overheidsgebouwen en publieke instellingen kunnen tot de openbare ruimte worden gerekend.

**Als je data verzamelt van een proces** (bijvoorbeeld tellingen van mensen werkend in een dienstrooster, tellingen van aantallen auto's die onderdeel zijn van een ritplanning, of observaties van een productieproces) dan is de data eigendom van het uitvoerende bedrijf. Klantgegevens zijn onderdeel van de afspraken die het bedrijf met zijn klanten heeft gemaakt over het gebruik van data. Je hebt alleen te maken met het bedrijf zelf en niet met de klanten van dit bedrijf.



Data verzamelen betekent niet dat je zomaar alles mag verzamelen of dat je alle data kunt gebruiken. De verzamelde data moet in de juiste verhouding staan met het doel van het onderzoek. Eindeloos taxi's op een standplaats tellen zonder doel is niet toegestaan. Ook moet je de **mensen** over je de informatie verzamelt vooraf vragen en informeren. Zie hoofdstuk twee over privacy.

## 1.1.2 Data van anderen gebruiken

Data van anderen gebruiken betekent dat jij geen eigenaar bent, je bent dan **verwerker** van data. De data kan fysiek aan je worden overgedragen (via een gegevensdrager zoals een cdrom of USB stick, of via downloadsoftware zoals wetransfer, API of ftp). De data kan ook op een extern platform in te zien zijn, maar niet te downloaden. Verwerker ben je alleen als je data daadwerkelijk downloadt.

Als je verwerker bent, is de dataleverancier vaak de eigenaar van de data, maar als verwerker heb je wel verantwoordelijkheden. Zowel de dataeigenaar als de verwerker stelt een verantwoordelijke persoon aan. De dataeigenaar bepaalt uiteindelijk hoe wordt omgegaan met de data.

Voor de verwerker geldt het volgende:

- De verwerker zorgt ervoor dat de data niet gedeeld kan worden met onbevoegden.
- De dataeigenaar moet zorgen dat een **geheimhoudingsverklaring** (NDA, Non Disclosure Agreement) getekend wordt als privacygevoelige informatie wordt verwerkt. Alleen degenen die een NDA hebben getekend mogen de data analyseren voor de periode waar ze toestemming voor hebben. Daarbij dient gezorgd te worden dat de onderzoekers alleen toegang krijgen tot het hoogstnoodzakelijke om hun onderzoek te doen. Projectleider, data analisten en studenten zijn geïnformeerd over het niet zomaar delen van alle data. Alleen wat echt nodig is en niet zonder toestemming van de eigenaar.
- De data moet zo worden opgeslagen dat de data **niet geëxporteerd** kan worden. Dit betekent dat je data bij voorkeur niet op een laptop bewaart die kwijt kan raken (het HvA gebouw verlaat), of dat de data wordt gedeeld en onnodig fysiek op laptops van bijvoorbeeld studenten en externe personen belandt. Moet je wel op laptops werken, maak dan afspraken om de data samen te verwijderen zodra deze niet meer nodig is en leg deze afspraken vooraf vast als onderdeel van de NDA.
- Data bewaren is vaak nodig voor de verantwoording van onderzoek. De afspraken daarover leg je vast in een **datamanagementplan** dat je maakt in overleg met de eigenaar. Zie hiervoor: <https://rdm.uva.nl/plannen/datamanagementplan/datamanagementplan.html>
  - De dataverantwoordelijke bepaalt hoelang en waar data mag worden bewaard, met inachtneming van regels van bewaartermijn zoals bij (wetenschappelijk) onderzoek gebruikelijk is. Een bewaarbeleid moet worden opgesteld waar de noodzaak van de opslag wordt uitgelegd.

## 1.1.3 Open data gebruiken

Open datasets worden in groten getale gedeeld op websites en bevatten veel informatie die je kunt gebruiken in onderzoek. In de precieze definitie van open data is vastgelegd dat:

- De data is **openbaar**;
- Er berust **geen auteursrecht** of andere rechten van derden op;
- De data zijn **bekostigd uit publieke middelen**, beschikbaar gesteld voor de uitvoering van die taak;



- De data voldoen bij voorkeur aan ‘**open standaarden**’ (geen barrières voor het gebruik door ICT-gebruikers of door ICT-aanbieders);
- Open Data is bij voorkeur **computer-leesbaar**, zodat zoekmachines informatie in documenten kunnen vinden.

Open datasets van particulieren of van bedrijven zijn dus strikt genomen geen “open” data, maar vallen in de categorie “gedeelde” data.

Open data is altijd voorzien van tags met daarop de datum, de bron en uitleg van de data. De bron en de datum moet je altijd gebruiken bij de verantwoording van het datagebruik. Er is geen eigenaarschap van de data en bij gebruik van deze data ben je dus wettelijk gezien “verwerker”. Als je de data downloadt en opslaat dan ben je “host” van de data. Deze mag je vrij verspreiden, als je de bron en datum maar meegeeft.

### **Gedeelde data**

Gedeelde data blijft eigendom van dataleverancier en als gebruiker ben je ook weer “verwerker”. Je mag de data officieel niet delen zonder toestemming van de eigenaar. Je kunt wel zonder toestemming van de eigenaar verwijzen naar de download locatie. De eigenaar heeft waarschijnlijk gekozen voor het openbaar delen van de data om niet telkens toestemming te hoeven geven voor het verder delen van de data. Verwijs dus naar de download locatie en ga bij voorkeur niet de datasets zelf delen.

### **Social media data**

Een speciale situatie waar het laatste woord nog niet over is gezegd: social media data (bijvoorbeeld twitter data). Dit is een vorm van gedeelde data, deze worden vaak alleen live gedeeld en je kunt de berichten downloaden. Je kunt niet alle historische berichten downloaden. Dat resulteert erin onderzoekers die de data gaan “scrapen”, dat is over langere tijd live downloaden en lokaal bewaren. Twitter heeft in zijn voorwaarden staan dat je niet alle twitter berichten mag downloaden in een eigen database, maar je mag wel op een tag (onderwerp) zoeken en gericht alle berichten met die tag gedurende lange tijd downloaden. Zo heb je toch een dataset in beheer die niet van jou is, deze mag je wel analyseren, maar als dataset niet delen zonder toestemming van het social media bedrijf.

## **2. Algemene verordening gegevensbescherming (AVG)**

De Algemene verordening gegevensbescherming (AVG) gaat over het rechtmatig omgaan met persoonsgegevens. De AVG is ook wel bekend onder de Engelse naam: General Data Protection Regulation (GDPR). Deze wet is sinds mei 2018 van kracht. De belangrijkste bepalingen uit de AVG zijn als volgt samen te vatten:

De wet is bedoeld om personen te beschermen.

Als je met persoonsgegevens werkt heb je verantwoordelijkheden. Je moet kunnen aantonen dat je organisatie in overeenstemming met de AVG handelt. De kernactiviteiten die van belang zijn om aan de AVG te voldoen zijn globaal samen te vatten in drie werkwoorden: **vragen, informeren, vastleggen.**

### **VRAGEN**

Als je persoonlijke gegevens verzamelt dan moet je de betrokkenen op de hoogte stellen dat je gegevens verzamelt en elke persoon moet toestemming geven dat je de gegevens opslaat en gebruikt.

### **INFORMEREN**



De persoon van wie de persoonsgegevens worden verwerkt moet op de hoogte zijn van de identiteit van de organisatie die persoonsgegevens opslaat en verwerkt (de zogeheten verwerkingsverantwoordelijke) en van het doel van de gegevensverwerking.

## VASTLEGGEN

Persoonsgegevens mogen alleen verzameld worden met een gerechtvaardigd doel. Dat doel moet welbepaald zijn en vooraf expliciet zijn omschreven. Het doel waarvoor een organisatie de persoonsgegevens gaat verwerken moet verenigbaar zijn met het doel waarmee de persoonsgegevens zijn verzameld. Je kunt dus niet zomaar alles verzamelen.

De wet heeft de regels dus flink aangescherpt voor wat betreft het opslaan van persoonsgegevens en geeft meer directieve richtlijnen over het informeren van personen over te verzamelen persoonsgegevens. Maar ook: het verzamelen moet een doel hebben en in de juiste verhouding staan tot dit doel.

### 2.1 Rollen

Binnen de AVG heeft de autoriteit persoonsgegevens (AP) verschillende rollen gedefinieerd. Iedere rol heeft een aantal rechten en plichten.

- de verwerkingsverantwoordelijke
- de functionaris gegevensbescherming
- de verwerker
- de betrokkene
- de Autoriteit Persoonsgegevens als toezichthouder

Een betrokkene is degene van wie informatie wordt verwerkt. Deze persoon heeft voornamelijk rechten.

De meeste onderzoekers op de hogeschool zijn verwerkers, namens wie één persoon verwerkingsverantwoordelijke is.

Verwerkers hebben plichten om namens de instelling aan de AVG te voldoen, de verwerkingsverantwoordelijke is namens hen verantwoordelijk om dit ook vast te leggen. Binnen een organisatie hoort een functionaris gegevensbescherming aangesteld te zijn.

Voor de HvA is dat momenteel Martijn de Hamer ([Functionarisgegevensbescherming@hva.nl](mailto:Functionarisgegevensbescherming@hva.nl)).

### 2.2 Proportionaliteitsvereiste en aan subsidiariteitsvereiste

Een van de belangrijke aspecten van de AVG waar je aan zult moeten voldoen zijn (met twee moeilijke woorden) het proportionaliteitsvereiste en aan subsidiariteitsvereiste.

Het **proportionaliteits**vereiste betekent dat je niet zomaar alles mag verzamelen.

**Subsidiariteit** betekent dat altijd gekeken moet worden of het beoogde doel dat voor de verwerking is vastgesteld ook op een minder ingrijpende manier en/of met minder ingrijpende middelen kan worden bereikt.

### 2.3 Zelf data verzamelen

- Voor dit beleid moet je een privacy statement ophangen (stickers, poster, eventueel voorlichting door een persoon) op de plek waar je de data verzamelt, want je moet voldoen aan de informatieplicht, hierin moet in begrijpelijk Nederlands staan wat gedaan wordt met de



gegevens en waarom, en wat hun rechten zijn. Mensen moeten de informatie eerst kunnen lezen of horen voordat hun gegevens worden gebruikt: ze moeten ervoor kunnen kiezen om geen gegevens af te staan. Mensen moeten zich dus kunnen onttrekken aan de dataverzameling en moeten het kunnen inzien als ze daarnaar vragen, en ze hebben het recht om de data die hen betreft te corrigeren. Dit moet allemaal in het statement terugkomen.

- Je moet kunnen aantonen dat je geldige toestemming van mensen hebt gekregen om hun persoonsgegevens te verwerken. En het moet net zo makkelijk zijn om de toestemming in te trekken als dat je de toestemming kan geven.
- Let op de proportionaliteit (de privacy inbreuk moet evenredig zijn aan het nagestreefde doel), dus het belang en doel van het onderzoek moet afgewogen worden tegen het opslaan van privacygevoelige informatie. Dit kun je bespreken met een AVG functionaris.

### 2.3.1 Voldoen aan proportionaliteitsvereiste en aan subsidiariteitsvereiste

Hou bij het verzamelen van data altijd rekening met de betrokkenen en het doel van het onderzoek.

Verzamel niet langer en niet meer (onnodige) gegevens dan noodzakelijk voor het onderzoek.

Als voorbeeld: niet 24 uur, 7 dagen per week WiFidata opslaan, maar alleen dat wat echt nodig is. Dus alleen wanneer de winkel open is. Dus van 9-18 uur data verzamelen. En alleen ip-adressen van mensen die in de winkel zijn opslaan en niet de ip adressen van mensen die buiten de winkel zijn. Want anders sla je ook ip-adressen op van mensen die naast of boven de winkel wonen en dat is tamelijk ingrijpend. Dit kan voorkomen worden door je WiFipunten anders in te stellen.

Om te voldoen aan het vereiste van subsidiariteit mag je namelijk alleen het hoogst noodzakelijke verzamelen.

Daarnaast moeten er geen andere minder invasieve middelen bestaan om hetzelfde doel te bereiken. Want als er een minder invasieve methode bestaat die hetzelfde doel bereikt, dan ben je verplicht die te gebruiken ook al is die methode veel duurder.

Als je verantwoordelijk bent over de data, ben je ook verantwoordelijk voor eventuele verdere verwerkingspartijen en een subverwerkingspartij. Dus als je verwerker het weer uitbesteedt aan een andere verwerker, dan heeft die zich ook te houden aan de overeenkomst, want ook daar ben je verantwoordelijk voor. Dus zorg dat je niet zomaar met je data in India of China eindigt en dat het daar zichtbaar is voor mensen.

- Je bent verantwoordelijk of er een back-up van ruwe data wordt opgeslagen bij de verwerker en bij jezelf en voor hoe lang.
- Je bent verantwoordelijk over hoe de data gehasht dan wel geanonimiseerd wordt bij de verwerker of bij jezelf of waarom niet.
- Je bent verantwoordelijk voor wie van de verwerkers bij de data kan en wat ze kunnen zien.

### 2.3.2 Tips om makkelijker te voldoen aan AVG

- Zorg voor privacy by design; houdt bij het ontwerp van je producten en diensten er rekening mee dat de privacygevoelige gegevens goed worden beschermd.

En zorg voor privacy by default; neem technische en organisatorische maatregelen dat je alleen persoonsgegevens verwerkt die echt noodzakelijk zijn. Maar ook bijvoorbeeld dat je het vakje “ja ik wil meedoen aan het onderzoek” niet al door jou als bedrijf is aangevinkt.



Zelfs als je het opslaan van persoonsgegevens probeert te voorkomen (bijvoorbeeld door data te anonimiseren), analyse van de data kan ervoor zorgen dat je toch personen op identiteit kunt herleiden uit een dataset.

Een voorbeeld hiervan is de aanwezigheid van mensen in lokalen door het registreren van WiFidata. Via het Cisco systeem worden alle inlog(pogingen) op de WiFi connector points geregistreerd met een unieke individueel apparaat identifier. Zelfs als je de unieke identifier van elk apparaat anonimiseert dan nog kun je een apparaat door het gebouw volgen in de tijd. Met kennis van de kantoren en de roosters kun je dan identificeren van wie het apparaat is.

Het kan dus zijn dat een dataset op zichzelf niet privacygevoelig is, maar dat in combinatie met andere dataset en analyse wel wordt. ip-adressen of WiFi data zijn in combinatie met locatiegegevens opeens privacy gevoeliger, aangezien je mensen daadwerkelijk zou kunnen gaan volgen. Bij het koppelen kan data dus privacygevoelig worden. Bij twijfel, check dit bij de data steward van je faculteit (Voor de faculteit is dit Maurice Pelt, m.m.j.m.pelt@hva.nl) of met de functionaris gegevens bescherming (voor de HvA is dat momenteel Martijn de Hamer ([Functionarisgegevensbescherming@hva.nl](mailto:Functionarisgegevensbescherming@hva.nl))).

Om privacy issue goed te kunnen bespreken, maak je voor het consult met de data steward een overzicht van de kolommen die je hebt, met een duidelijke beschrijving erbij, zoals hieronder is weergegeven.

Kolomnaam	Voorbeeld	Beschrijving
CDR_ID	3564035	Unieke waarde voor een CDR record.
Start_datetime	2018-10-01T07:38:09+02:00	Starttijd van laadsessie
End_datetime	2018-10-01T07:41:20+02:00	Eindtijd van laadsessie
Duration	00:03:11	Lengte van laadsessie in hh:mm:ss
Volume	0,0100	Geladen hoeveelheid in kWh

*Voorbeeld van data van een laadsessie voor elektrische voertuigen (fictief)*

### Anonimiseren aan de bron

Een praktische tip is om bij het verkrijgen van data van anderen (een dataleverancier) goede afspraken te maken over welke data jij nodig hebt en accepteert. Het komt regelmatig voor dat een dataleverancier zelf geen zicht heeft op de (privacy) gevoeligheid van de data of vindt dat jij daar verantwoordelijk voor bent (“jij haalt wel even de adressen en namen eruit toch?”).

Het is beter om aan te dringen dat de dataeigenaar alleen data deelt die relevant is voor het onderzoek en zelf de privacy check doet en oplost. Als je de privacygevoelige informatie langer opslaat, dan moet je daar een goed onderbouwd argument voor hebben.

### 3. Meer informatie

Op de volgende websites vindt je meer informatie:

- De Research Data Management portal van de UvA en HvA is het beste startpunt voor meer informatie: <https://rdm.uva.nl/>
- Het Landelijk Coördinatiepunt Research Data Management bevat handreikingen over verschillende thema's: [www.lcrdm.nl](http://www.lcrdm.nl)
- De A-Z lijst van de HvA bevat informatie over auteursrechten, intellectueel eigendom, privacy en delen van data: <https://az.hva.nl/medewerkers/techniek/a-z/az.html?t=onderzoek>



- Meer informatie over privacy wetgeving is te vinden op <https://autoriteitpersoonsgegevens.nl/nl/over-privacy/wetten/algemene-verordening-gegevensbescherming-avg>