

# 1. Datamanagement en privacy

## **Stap 1 Verkennen, onderzoeksvraag en hypothese**

- 1.1 Keuze type data verzamelen
- 1.2 Hergebruik van data
- 1.3 Data van sociale media

## **Stap 2 Onderzoeksplan**

- 2.1 Privacy respondenten waarborgen
- 2.2 Gevoeligheid
- 2.3 Ethische toetsing

## **Stap 3 Uitvoeren en resultaten**

- 3.1 Informed consent
- 3.2 Gegevensverzameling en databoek
- 3.3 Data analyse in kwantitatief onderzoek
- 3.4 Data analyse in kwalitatief onderzoek?

## **Stap 4 Conclusies, verklaren en presenteren**

- 4.1 Welke resultaten selecteren? – kwantitatief
- 4.2 Welke resultaten selecteren? – kwalitatief
- 4.3 Data bewaren
- 4.4 Open Science

## **Introductie module datamanagement en privacy**

In deze module behandelen we integriteitsvragen die samenhangen met datamanagement en privacy. We gebruiken de stappen van de onderzoekscyclus als hulpmiddel om de vragen te bespreken. Aan de hand van verschillende opdrachten maak je kennis integriteitsvragen die belangrijk kunnen zijn voor de praktijk van jouw onderzoek(-spraktijk). We maken gebruik van casusistiek om te laten zien hoe deze vragen in de praktijk kunnen voorkomen. Hoe zorg je er bijvoorbeeld voor dat participanten in onderzoek onherkenbaar zijn (en blijven)? Wat zijn belangrijke regels en richtlijnen op het gebied van privacy? Na het doorlopen van de verschillende onderwerpen heb je een goed beeld van de belangrijkste onderwerpen rond datamanagement en privacy.

De leerdoelen voor deze module zijn:

- Je kunt benoemen welke integriteitsvragen m.b.t. datamanagement en privacy kunnen ontstaan in een onderzoekscyclus en op welke manier die relevant zijn in een onderzoeksproject
- Je weet welke eisen er aan informed consent worden gesteld vanuit Europese regelgeving (de AVG) en wat dit betekent voor het verzamelen, bewaren en analyseren van data in een onderzoek
- Je kunt voor je eigen onderzoek benoemen welke onderwerpen het meest relevant (kunnen) zijn en hoe je het onderzoek op een verantwoorde manier dient uit te voeren

## Stap 1: Verkennen, onderzoeksvraag en hypothese

### 1.1 Keuze type data verzamelen

In je onderzoek kun je verschillende soorten data verzamelen. Sommigen van jullie doen onderzoek dat gebaseerd is op literatuurstudies, algoritmen, experimenten in het lab, anderen doen onderzoek waar mensen bij betrokken zijn doordat je ze interviewt, vragenlijsten voorlegt, observeert of taken oplegt.

Bij kwalitatief onderzoek neem je bijvoorbeeld interviews af, organiseer je focus groep interviews of observeer je in het veld. Bij kwantitatief onderzoek kun je vragenlijsten uitzetten of experimenten doen. Elk type data verzameling kan leiden tot eigen vraagstukken op het gebied van integriteit, die te maken kunnen hebben met bijv. privacy, reproduceerbaarheid (een onderzoek herhalen) en (on)zorgvuldigheid. In onderstaande tabel[1], staan voorbeelden van typen dataverzameling.

[1] Tekst in de tabel komt van Dataverzamelingsmethoden voor je scriptie - 24editor

Type dataverzameling	Methode/discipline (verkorte versie van: <a href="#">Dataverzamelingsmethoden voor je scriptie - 24editor</a> )
<b>Experimenteel onderzoek</b>	Met een experimenteel onderzoek worden gegevens verzameld van proefpersonen in een 'gecontroleerde situatie', dat wil zeggen in een situatie vrij van invloeden van buitenaf (zoals een laboratorium). Een experimenteel onderzoek is een kwantitatieve dataverzamelmethode waarmee je het effect meet van de manipulatie van een variabele op een (verwachte) uitkomst. Deze methode wordt vaak gebruikt in de psychiatrie, medische- en natuurwetenschappen.

Type dataverzameling	Methode/discipline (verkorte versie van: <a href="#">Dataverzamelmethode voor je scriptie - 24editor</a> )
<b>Enquête</b>	<p>Met een enquête, ook wel survey-onderzoek genoemd, worden op systematische wijze vragen gesteld aan meestal een groot aantal personen. Deze methode wordt veel gebruikt in bijvoorbeeld afstudeeronderzoeken. Een surveyonderzoek is een gestructureerde kwantitatieve dataverzamelmethode die vaak ingezet wordt om beschrijvende- en verklarende onderzoeksvragen te beantwoorden. Met gestructureerd wordt bedoeld dat de vragen vooraf vaststaan, waarbij geen ruimte is voor flexibiliteit.</p>
<b>Monitor</b>	<p>Een monitoronderzoek is een kwantitatieve dataverzamelmethode waarmee je een onderzoek over een bepaald onderwerp regelmatig gaat herhalen. Hierdoor worden gegevens verzameld waarmee ontwikkelingen op een bepaald terrein gevolgd kunnen worden. Om te kunnen spreken van een monitoronderzoek, moeten de gegevens van onderzoeken op verschillende tijdstippen met elkaar worden vergeleken. Daarnaast is het van belang dat op elk meetmoment dezelfde meetinstrumenten worden ingezet.</p>
<b>Documentanalyse</b>	<p>Met een inhoudsanalyse of documentenanalyse ga je de inhoud van bestaande informatie analyseren. Om zo efficiënt mogelijk te analyseren, is het handig om vooraf een aantal kenmerken, oftewel variabelen, te bepalen van de documenten die je onderzoekt. Een documentenanalyse is een kwalitatieve dataverzamelmethode en is verkennend van aard. Deze methode is niet alleen in te zetten voor tekstmateriaal zoals een boek of wetenschappelijk artikel, maar ook voor beeld- en geluidsdocumenten.</p>
<b>Literatuuronderzoek</b>	<p>Met een literatuuronderzoek ga je op zoek naar wetenschappelijke informatie die betrekking heeft op de onderwerpen die centraal staan in de probleemstelling van jouw onderzoek. Het literatuuronderzoek is een kwalitatieve dataverzamelmethode waarmee je niet alleen boeken, maar ook (wetenschappelijke) artikelen, papers en publicaties kunt gebruiken als bron. Het literatuuronderzoek geldt als fundering voor je onderzoek en wordt weergegeven in het theoretisch kader van je scriptie.</p>

Type dataverzameling	Methode/discipline (verkorte versie van: <a href="#">Dataverzamelingmethoden voor je scriptie - 24editor</a> )
<b>Interviews</b>	Met een interview verzamel je informatie door respondenten vragen te stellen. Aan de hand van een interview kan je goed iemands beleving of motivatie achterhalen. Een interview is een kwalitatieve dataverzamelingmethode, met uitzondering van een gestructureerd interview.
<b>Focusgroep</b>	In een focusgroep of groeps gesprek worden meerdere mensen tegelijk bevroegd. Een groeps gesprek stimuleert de discussie tussen groepsleden, waardoor je makkelijk argumenten en motieven kunt genereren. Een groeps gesprek is een kwalitatieve dataverzamelingmethode. Deze methode is bijvoorbeeld geschikt om meningen, belevingen en barrières te achterhalen of om ideeën te toetsen.
<b>Observatie</b>	Een observatieonderzoek is een methode waarmee bepaalde gedragingen van personen kunnen worden waargenomen en bestudeerd. Deze waarnemingen worden vervolgens genoteerd in een logboek of databestand. Een observatieonderzoek is een kwalitatieve dataverzamelingmethode waarop meerdere varianten bestaan. Een observatie wordt veelvuldig in scripties gebruikt om 'hoe' of 'waarom' vragen te beantwoorden.
<b>Case study</b>	Een casestudy is een <b>kwalitatieve dataverzamelingmethode</b> waarmee je slechts één onderzoekseenheid, één case, gaat onderzoeken in zijn 'natuurlijke omgeving'. Deze onderzoekseenheid kan bijvoorbeeld een persoon, organisatie of systeem zijn. Een casestudy is breed toepasbaar en vaak worden meerdere <b>dataverzamelingmethoden</b> gecombineerd (triangulatie): observaties, interviews en bureauonderzoek. Daarom wordt ook wel gesproken van een onderzoeksdesign in plaats van methode. Een casestudy is goed inzetbaar voor beschrijvend- en verklarend onderzoek.
<b>Secundaire analyse</b>	Met een secundaire analyse verzamel je gegevens die al eerder door een onderzoeker zijn vergaard en geanalyseerd. Met deze dataset kun je misschien niet direct jouw onderzoeksvraag beantwoorden, maar het kan wel dienen als hulpmiddel. Een bureauonderzoek kan zowel een kwantitatieve als kwalitatieve dataverzamelingmethode zijn. Het verschil in deze dataverzamelingmethoden zit voornamelijk in de aard van de gegevens die verzameld worden. Het is een vereiste dat de bronnen die je gebruikt voor je secundaire analyse valide en betrouwbaar zijn.

## Opdracht

*Individuele opdracht (antwoorden kunnen vervolgens plenair worden besproken, met name vraag 4 en 5). Vragen voor je eigen onderzoek:*

1. Welk soort data ga je verzamelen? Kwalitatief of kwantitatief, labresultaten, historische bronnen of wat anders?
2. Hoe ga je je de data verzamelen?
3. Bekijk de [website van Universiteit Leiden](#) en de [digitale overheid](#), en kijk op de website van jouw hogeschool. Zoek uit welke regels en richtlijnen van belang zijn voor jouw specifieke onderzoek. Schrijf op hoe je deze het best toe kan passen.
4. Zie je op dit moment mogelijke problemen met het volgen, toepassen en begrijpen van regels voor jouw onderzoek?
5. Zoek uit waar je bij jouw hogeschool terecht kan voor ondersteuning op het gebied van regelgeving.

### 1.2 Hergebruik van data

Vaak maak je als student gebruik van datasets die anderen hebben verzameld, om analyses mee te doen. Bijvoorbeeld: je gebruikt interviews met leerkrachten over motivatie van leerlingen na de COVID pandemie om deze te vergelijken met interviewdata uit een onderzoek dat al voor de COVID tijd is verzameld. De onderzoekers die de data verzamelden, hadden daarbij een bepaald doel voor ogen, hebben toestemming gevraagd voor het verzamelen van de gegevens, en bepaalde methodes gebruikt om de gegevens te verzamelen. Bij [hergebruik](#) van deze data is het daarom van belang om na te denken over een aantal vragen over beschikbaarheid, doel, kwaliteit, herkomst en toestemming van respondenten.

1. Kan je de bestaande data opnieuw gebruiken (zijn ze beschikbaar)?
2. Wat was het doel van dataverzameling?
3. Wat zou je kunnen zeggen over de kwaliteit van de dataset?
4. Wat is de herkomst van de data en wie is eigenaar? (Heb je toestemming nodig van de data-eigenaar?)
5. Is hiervoor opnieuw toestemming nodig van de respondenten/proefpersonen? Wat is er over hergebruik in het informed consent opgenomen?

## Opdracht

*Bij jouw eigen afstudeeronderzoek is waarschijnlijk de stage instelling eigenaar van de data. Toch is het belangrijk dat je zelf ook stilstaat bij een aantal vragen over data, dat je algemene regels kent en weet waar hulp te vinden is. Beantwoord onderstaande vragen samen. Als je bepaalde vragen niet kan beantwoorden, bespreek deze dan met je stagebegeleider.*

1. Met welk doel zijn gegevens verzameld? Zijn de data voor jouw doel ook bruikbaar?
2. Wanneer heb je toestemming nodig om data voor andere onderzoeksdoeleinden te gebruiken bij mens gebonden onderzoek?
3. Hoe kan je de kwaliteit en betrouwbaarheid van een dataset beoordelen?
4. Wat is de herkomst van de data? Wie is eigenaar van data?
5. Zoek uit waar je bij jouw hogeschool terecht kan als je vragen hebt over hergebruik van data.

### 1.3 Data van sociale media

Internet en sociale media zijn een rijke bron van onderzoeksdata: er zijn tal van mogelijkheden om datasets te maken over bestaande informatie; blogs, Facebook posts; Tweets. Een belangrijke vraag is of en hoe je de privacy kunt waarborgen en hoe je op de juiste manier omgaat met verzamelde data. In het achtergrondmateriaal staan bronnen met richtlijnen en regelgeving. Mag je immers zomaar blogs of tweets gebruiken in je onderzoek?

## Opdracht

**Casus:** *Stel, je bent geïnteresseerd om voor je onderzoek de mening van studenten over de nieuwe basisbeurs te horen. Je gaat op Facebook en Instagram een post zetten met de vraag om reacties van mensen op de volgende stelling: ‘een basisbeurs moet in eerste levensbehoeften kunnen voorzien (wonen, eten, reizen) en moet dus omhoog’. Je zet erbij dat je de posts gebruikt voor je scriptieproject. De reacties stromen binnen, ook reacties die nogal bot of zelfs agressief zijn.*

Bespreek met elkaar en schrijf op:

1. **Wat zijn relevante overwegingen als je data verzamelt via [sociale media](#) en wat vraagt om reflectie?**
2. **Waar geven mensen toestemming voor als zij op social media iets posten? Mag je dan zomaar hun data gebruiken of zijn daar voorwaarden aan verbonden?** [leef je in, bijvoorbeeld door na te denken wat je er zelf van vindt dat jouw gegevens worden gebruikt zonder dat je precies weet waarvoor. [Lees jij bijvoorbeeld cookiestatements en toestemmingsformulieren?](#)]
3. **Zijn er manieren om privacy en toestemming online beter te waarborgen?**

## Stap 2 Onderzoeksplan

### 2.1 Privacy respondenten waarborgen

Bij het schrijven van je onderzoeksplan is het belangrijk dat je nadenkt over de manier waarop je de privacy van respondenten gaat waarborgen. Dat doe je in de eerste plaats door je te houden aan de regels uit de Europese privacy wetgeving, namelijk de Algemene gegevensverordening ([AVG](#)).

De volgende stap is het waarborgen van geïnformeerde toestemming van respondenten of proefpersonen: informed consent genoemd (voor meer informatie zie 4.1). Tijdens het schrijven van je onderzoeksplan is het goed om vast na te denken op welke manier privacy een rol speelt in jouw onderzoek en hoe je op een verantwoorde manier met [persoonsgegevens](#) om gaat. Twee dingen zijn belangrijk: 1) herleidbaarheid 2) dataminimalisatie.

Om privacy te waarborgen is het van belang dat gegevens niet te herleiden zijn naar individuen. Herleidbare gegevens in het interview moet je anonimiseren of pseudonimiseren bij kwalitatief onderzoek.

Persoonsgegevens worden weggelakt (zodat je ze niet meer kan zien) uit interviews (plaatsnamen, organisaties waar men werkt etc.).

In kwantitatief onderzoek kunnen demografische gegevens uit een survey worden geanonimiseerd. Het is bij kwantitatief onderzoek belangrijk mogelijkheden tot re-identificatie te voorkomen. Het combineren van gegevens kan ervoor zorgen dat respondenten alsnog te herkennen zijn. Bij kleine vragenlijsten in organisaties zijn gegevens al snel te herleiden als een aantal kenmerken van een respondent worden gecombineerd, zoals leeftijd en woonplaats.



Een andere belangrijke stap is nadenken over data minimalisatie: verzamel je alleen wat nodig is voor je onderzoeksvraag? Vervolgens zorg je dat persoonsgegevens van deelnemers (zoals emailadressen) niet in hetzelfde document staan als de gegevens die je hebt verzameld en dat je interview een code krijgt.

## Opdracht

### Lees onderstaande tekst uit een niet-bestaand interview

*‘Wat is uw favoriete eetgelegenheden?’ Ik eet graag bij Smaeck hier in Weteringen, want dat is lekker dichtbij. Er komen veel leeftijdgenoten, net gesettelde dertigers zeg maar, die af en toe zich lekker willen laten verwennen. Het is er een gezellige boel. De eigenaar, Piet Scheffers zorgt dat er altijd verse en regionale producten in de maaltijden worden verwerkt. Dat vind ik heel verantwoord en duurzaam. Ik heb er onlangs nog met mijn collegaatjes van Renova bouw gegeten. Super gezellig en lekker. Ik vind het ook leuk om via social media dit restaurant een beetje te supporten, door het op Facebook of Instagram te vermelden dat ik er ben geweest en wat er op het menu stond.’*

Bespreek in groepjes:

1. Op welke manieren kan je privacy waarborgen in de casus? Wat moet hier weggelakt?
2. Hoe waarborg jij privacy in je afstudeer onderzoek? Waar moet je rekening mee houden?
3. Wat zijn wettelijke regels waar je je aan moet houden?
  - a. Hou houd je persoonsgegevens apart van je data?
  - b. Wat betekent anonimiseren en pseudonimiseren van data en hoe doe je dat? Is er een verschil tussen anonimiseren en pseudonimiseren?
  - c. Hoe kan je re-identificatie vermijden?

[antwoorden: weggelakt moet worden: Smaeck, Weteringen, Piet Scheffers, Renova bouw.]



## 2.2 Gevoeligheid

Bepaalde gegevens van personen worden in de Europese privacywetgeving (AVG) gezien als ‘[gevoelig](#)’ en vallen daarom onder specifieke [verwerkingsvoorwaarden](#):

- Gegevens over ras of etnische afkomst, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen
- Vakbondlidmaatschap
- Genetische gegevens; biometrische gegevens die verwerkt worden om een persoon te identificeren
- Gezondheidsgegevens
- Gegevens die gaan over iemands seksueel gedrag of seksuele gerichtheid

Voorbeelden van onderzoek naar gevoelige onderwerpen zijn bijvoorbeeld:

- o onderzoek onder de moslimgemeenschap naar donaties
- o forensisch onderzoek naar data op computers
- o terrorismelessen op basisscholen
- o onderzoek naar ‘slaan in de opvoeding’

Bij het doen van onderzoek naar gevoelige onderwerpen is het belangrijk gegevens goed op te slaan, en het is daarnaast ook belangrijk om respondenten voorzichtig te benaderen. Het kan bijvoorbeeld kwetsbaar zijn om over bepaalde zaken te spreken in een interviews, zoals over ‘een kind slaan in de opvoeding’, en als mensen zich als groep onveilig of gestigmatiseerd voelen, of bang zijn dat hun gegevens niet anoniem verwerkt worden, zijn ze sneller geneigd sociaal wenselijke antwoorden te geven (of in te vullen bij een vragenlijst). Dit kan invloed hebben op de betrouwbaarheid van je resultaten; je resultaten geven dan immers geen betrouwbaar beeld van wat men daadwerkelijk doet of vindt.

# Opdracht

## Bespreek met elkaar:

- Welke (van de hierboven genoemde) gegevens vinden jullie het meest gevoelig?
- Maak een lijstje op volgorde van gevoeligheid (meest-minst).
- Zijn er nog andere gegevens die jullie gevoelig vinden (financiële gegevens, gegevens over gezinssamenstelling, opleiding of iets anders)?
- Bespreek vervolgens per onderdeel wat gevolgen kunnen zijn voor personen als deze gegevens niet goed beschermd worden.

## 2.3 Ethische toetsing van mensgebonden onderzoek

In veel disciplines worden menselijke proefpersonen gebruikt. In de medische context, sociale en andere wetenschappen worden studies uitgevoerd waarbij personen vrijwillig deelnemen aan experimenten, nieuwe behandelingen, geneesmiddelenonderzoek, enquêtes, interviews of etnografische studies. Onderzoek met mensen kun je niet zomaar doen: het moet ethisch verdedigbaar zijn. Vaak wordt een studie met mensen voorafgaand aan een ethische commissie voorgelegd, waarbij niet alleen de methode van onderzoek moet worden voorgelegd, maar een afweging tussen risico's, belasting en baten van de studie wordt gekeken. Daarnaast moet toestemming op een zorgvuldige manier geregeld zijn en moet informatie over het verzamelen en bewaren van data worden beschreven.

Veel instellingen, ook hogescholen, hebben tegenwoordig hun [eigen ethische commissie](#), of speciaal voor [dierproefonderzoek](#). Voor onderzoek in de gezondheidszorg bestaat er zelfs een wettelijk verplichte ethische toets, die door een METC, een [Medisch Ethische Toetsingscommissie](#), wordt uitgevoerd.

Vier waarden spelen een rol in het maken van ethische afwegingen bij mensgebonden onderzoek: vrijwilligheid, niet-schaden, rechtvaardigheid en weldoen. Deze principes kunnen tegen elkaar worden afgewogen om tot een beslissing te komen.

Ethische commissies maken een afweging of onderzoek toelaatbaar is. In de praktijk zullen de meeste commissies hun oordeel baseren op de volgende overwegingen:

- Is informed consent goed geregeld (en voldoet het aan de AVG regelgeving)?
- Is het gegevensbeheer goed geregeld?
- Staat de relevantie van het onderzoek in verhouding tot de belasting voor de deelnemers en de risico's?
- Gaat het onderzoek zelf over gevoelige onderwerpen of kwetsbare groepen deelnemers? Is het gerechtvaardigd om het onderzoek te doen? Zijn er voorzorgsmaatregelen genomen?
- Als misleiding (respondenten wordt niet de (hele) waarheid verteld, bv of een medicijn werkt) deel uitmaakt van een experiment: worden de deelnemers naar behoren (evt achteraf) geïnformeerd?
- Is er een vergoeding voor deelnemers nodig? Zo ja, is die redelijk? En leidt dit niet tot deelname om oneigenlijke/financiële redenen?

## Opdracht

*Deze opdracht is voor studenten die zelf een onderzoeksvoorstel hebben geschreven.*

1. Vorm in groepjes van 4 een ethische commissie en beoordeel één voor één elkaars onderzoeksvoorstel.
2. Ga bovenstaande vragen af om tot een evaluatie te komen en wees specifiek in je beoordeling; **wat verstaan jullie onder een risico en onder belasting? Wanneer is een groep deelnemers kwetsbaar?**
3. Bereid per groep een korte presentatie voor waarin je je advies geeft over de onderzoeksvoorstellen en jullie afweging o.b.v. de antwoorden op de vragen wordt toegelicht.

## Stap 3 Uitvoeren en resultaten

### 3.1 Informed consent

Het krijgen van informed consent van je respondenten is een belangrijk onderdeel van je onderzoek. Respondenten hebben [rechten](#). Door toestemming te geven maken respondenten zelf een bewuste afweging om mee te doen aan je onderzoek. Voor het maken van deze afweging hebben respondenten informatie nodig over jouw onderzoek:

1. Waar gaat het onderzoek over?
2. Wat wordt verwacht van de respondent/participant?
3. Welke gegevens worden verzameld?
4. Hoe worden gegevens bewaard? (anoniem of niet)
5. Hoe worden de gegevens verwerkt? Wordt er gepubliceerd?
6. Worden gegevens hergebruikt?
7. Wie voert het onderzoek uit? Wie heeft toegang tot de gegevens?
8. Wie is de contactpersoon en hoe is deze te bereiken? Wat bij klachten?
9. Een respondent mag zich altijd terugtrekken uit het onderzoek, zonder opgave van reden.

Deze informatie schrijf je in begrijpelijke taal in een brief, die de respondent vooraf dient te lezen en begrijpen. Het tweede deel van de brief bestaat uit de toestemmingsverklaring die ondertekend dient te worden door de respondent. Je vraagt daarbij toestemming voor:

- Verzamelen en verwerken van gegevens
- Archiveren van de data
- Eventuele (anonieme) publicatie (bv als scriptie of wetenschappelijk artikel)
- Eventueel hergebruik van de data

Als iemand (een respondent) zelf niet in staat is toestemming te geven moet dit gedaan worden door een wettelijk vertegenwoordiger, of bij minderjarigen door een ouder of voogd. Soms is een aparte brief nodig voor de vertegenwoordiger (denk aan een brief voor kinderen en een voor de ouders). Wanneer je onderzoek doet met groepen – collectief consent - kan het lastig zijn als een deel van de groep (of ouders) wel toestemming geeft voor bijvoorbeeld opnames, en een ander deel niet. Bij kwetsbare groepen is het belangrijk extra zorgvuldig over IC na te denken.

## Opdracht

*Schrijf een informatiebrief voor je eigen (of een fictief) onderzoek en neem daarin bovengenoemde punten op.*

*Beoordeel vervolgens de brief van een medestudent en vraag feedback op je eigen brief. **Wat is goed en wat kan er beter?***

**Voor docenten:** [aanvullend materiaal: voorbeeld IC brief + wie controleert IC in de praktijk? Dit kan bijvoorbeeld door een steekproef]

### 3.2 Gegevensverzameling en databoek

Verantwoord onderzoek doen houdt in dat je zorgvuldig te werk gaat bij het [verzamelen](#) van gegevens. Elke keuze die je maakt, elk onderdeel van je experiment, elk artikel dat je opneemt of uitsluit van een review, en elke stap in het bepalen van de opzet van je studie is relevant. In sommige vakgebieden zijn er strikte voorschriften voor het bijhouden van lab-notities, of een [logboek](#), maar in andere vakgebieden is dit misschien wat minder duidelijk.

Nuttige informatie om bij te houden over je gegevensverzameling is bijvoorbeeld:

- Wat je gedaan hebt
- Waarom je het deed
- Hoe je het gedaan hebt
- Waar materialen te vinden zijn
- Wat er gebeurde (en wat niet)
- Je interpretatie
- Bijdrage van anderen
- Wat is de volgende stap en hoe ben je tot dat besluit gekomen

Het doel van databoeken doe je om meerdere redenen:

1. je houdt goed bij welke keuzen je maakt tijdens het onderzoeksproces en hoe je tot die keuzes bent gekomen (vaak is onderzoek een minder gestructureerd proces dan het lijkt in de artikelen die eruit voortkomen)
2. je stelt anderen in staat je keuzes bij te houden en je te helpen bij de herziening ervan
3. je stel anderen in staat je gegevens te gebruiken. Ze maken de gegevens ook beter reproduceerbaar en helpen bij de verificatie van resultaten.

## Opdracht

*Bedenk wat er in het datalogboek voor je eigen onderzoek zou moeten staan aan de hand van de punten die worden genoemd in de tekst.*

**Welke keuzes zijn daarbij voor jouw onderzoek het meest belangrijk om goed bij te houden (wat zijn aspecten waar anderen vragen over zouden kunnen hebben)?**

### 3.3 Data analyse in kwantitatief onderzoek

Voor het analyseren van data zijn verschillende [methoden](#). Het is belangrijk de methodes die je gebruikt en de motivatie daarvoor helder te omschrijven en vooraf vast te leggen. Zo voorkom je een onderzoeksbias, namelijk dat je de analyse en resultaten gunstiger presenteert of dingen weglaat. Dit noemen we ook wel confirmation bias. Bij kwantitatief onderzoek is het belangrijk dat resultaten kunnen worden gerepliceerd; dit betekent dat als een onderzoeksofzet een tweede keer wordt uitgevoerd, dezelfde resultaten naar voren komen. Replicatie is belangrijk, momenteel is er een [replicatiecrisis](#) gaande.

In kwantitatief onderzoek kan juist analyseren P-hacking en HARKing helpen voorkomen. P-hacking is de neiging patronen te ontdekken die statistisch significant lijken, terwijl er feitelijk geen effect is [zie [filmpje](#): en [website](#)]. HARKing lijkt op P-hacking, en is de afkorting voor Hypothesis After Result is Known; een twijfelachtige bezigheid omdat het idee in onderzoek is eerst een hypothese op te stellen en deze te testen. Niet andersom. P-hacking en HARKing vergroten het risico op fouten, en verkleinen de kans om een onderzoek te repliceren.

## Opdracht

*Bij het analyseren van je data dien je na te denken over betrouwbaarheid van je data en de gebruikte methoden. Bespreek deze vragen die kunnen opkomen bij het analyseren van je data:*

1. **Sluit je outliers uit van je analyse?** Outliers zijn uitschieters, punten die ver afwijken van de rest van de data, die data sterk kunnen vertekenen.
2. **Welke statistische methode gebruik je voor de analyse?**
3. **Kan er bias in je coderingen zitten? Kan er sprake zijn van P-hacking of HARKing?**
4. **Zijn je experimenten/is je dataverzameling voldoende/correct uitgevoerd?**
5. **Heb je voldoende verschillende perspectieven uit de literatuur weergegeven in je literatuurstudie?**

*Welke van deze vragen zijn toepasbaar in jouw onderzoek? Hoe zou je ze beantwoorden?*

### 3.4 Data-analyse in kwalitatief onderzoek

Voor het analyseren van data zijn er ook verschillende methoden. Het is belangrijk de methodes die je gebruikt en de motivatie daarvoor helder te omschrijven. **Welke stappen heb je doorlopen tijdens je analyse?** Ook hier geldt dat je een confirmation bias moet zien te voorkomen. Een confirmation bias ontstaat als je niet onafhankelijk naar je materiaal kijkt maar op zoek bent naar bevestiging van je aannemers. Bij het analyseren van interviewdata bv kun je open axiaal coderen, waarbij je niet tevoren vaststelt welke categorieën je aanbrengt. Of je zorgt dat twee mensen coderen en elkaar checken om een te nauwe focus of fouten te voorkomen.

## Stap 4 Conclusies, verklaren en presenteren

### 4.1 Welke resultaten selecteren? - kwantitatief

Wanneer je je onderzoek gaat rapporteren en presenteren maak je bepaalde keuzes; **welke resultaten selecteer je? Waarom vind je deze resultaten belangrijk en andere minder?** Bij het analyseren, rapporteren en presenteren van data kun je veel verschillende integriteitskwesaties tegenkomen.

Je statistische analyses kunnen bv teveel gericht zijn op het vinden van significante verschillen.. Of je gaat hypothesen aanpassen op basis van de data die je gevonden hebt. Je kunt ook je resultaten iets te positief brengen, of maar een deel van het verhaal vertellen. Dit kan tot foute interpretaties leiden over je onderzoek. Je kunt je conclusies ook wat 'opblazen', waardoor ze groter en belangrijker lijken dan ze feitelijk zijn op basis van je data. Zeker als je met stakeholders werkt, kan de verleiding soms groot zijn om het allemaal iets gunstiger te presenteren, of juist kritiek op een organisatie achterwege te laten (zie ook de module Conflict of Interests).

Een bekend voorbeeld zijn geneesmiddelenstudies: als ze gesponsord zijn door de farmaceutische industrie is de kans aanwezig dat studies die niet succesvol waren nooit gepubliceerd worden of dat de positieve resultaten iets sterker aangezet worden dan wenselijk.



## 4.2 Welke resultaten selecteren? - kwalitatief

Wanneer je je onderzoek gaat rapporteren en presenteren maak je bepaalde keuzes; **welke resultaten selecteer je? Waarom vind je deze resultaten belangrijk en andere minder?** Bij het analyseren, rapporteren en presenteren van data kun je veel verschillende integriteitskwesaties tegenkomen. Het trekken van een overhaaste conclusie kan problematisch zijn.

## 4.3 Data bewaren

Het is belangrijk dat onderzoeksgegevens, data, goed worden opgeslagen. Dit kan op een repository. Een repository is een centrale plaats waar computergegevens, data, software is opgeslagen. Het is belangrijk dat deze repository goed georganiseerd is, zodat data terug te vinden is. Hogescholen kunnen gebruik maken van een dienst zoals [Surfdrive](#), of een eigen server hebben.

Voor het bewaren van data gelden bepaalde [termijnen](#).

## Opdracht

*Zoek op en bespreek wat belangrijke bewaartermijnen zijn voor verschillende soorten onderzoek (interviews / medisch wetenschappelijk onderzoek / experimenten met dieren)*

- **Waarom zijn er bewaartermijnen?**
- **Wat zijn voor en nadelen van data eeuwig opslaan?**

Hieronder worden voor en nadelen gegeven. Maak een lijst met voordelen, en een lijst met nadelen. Bespreek vervolgens met elkaar wat een voordeel is en wat een nadeel, en waarom.

- **Voordelen:** Open science, cultuur waarin onderzoekers elkaars data kunnen gebruiken; meerdere onderzoeken kunnen gedaan worden met dezelfde data (kostenbesparing); .....
- **Nadelen:** kosten, duurzaamheid (energieverbruik van opslag, hoeveel versies sla je op?). Ongewenst gebruik (bepaalde onderzoeksresultaten gebruikt in oorlogen)

*Zoek uit hoe het bewaren van data op jouw hogeschool geregeld is en waar je hulp kan vinden als je vragen hebt. **Heeft jouw hogeschool een FAIR data beleid?***

#### 4.4 Open science

Hoe vrij beschikbaar je je resultaten maakt, is een keuze. [Open science](#) is een beweging waarin we stil gezet worden bij het open en beschikbaar stellen van onderzoeksdata en resultaten, maar ook over het transparant werken in algemene zin. Hoewel je als student waarschijnlijk weinig invloed hebt op open science, geven we een paar voorbeelden hoe open science relevant kan zijn in je onderzoek:

1. Het is de trend om 'open access' te publiceren in tijdschriften en rapporten op bv websites van je instelling te plaatsen. Op die manier zijn jouw resultaten voor anderen snel en gemakkelijk te vinden. Maar niet alles kan open en online. Vertrouwelijke gegevens en transcripties van interviews zijn daarvoor niet altijd geschikt.
2. Het open beschikbaar stellen van je data doe je via een repository, veel onderwijsinstellingen hebben deze ter beschikking, denk bijvoorbeeld ook aan de hbo kennisbank.
3. **Wil je je data wel in een tijdschrift publiceren, maar denk je dat anderen er sneller behoefte aan hebben deze tot hun beschikking te hebben?** Dan kun je ook aan een preprint denken: dat zijn artikelen die nog geen peer review hebben gehad, en waarvan de data dus niet gecheckt zijn door peers uit het veld. In toenemende mate plaatsen onderzoekers hun artikelen vast als preprint. Een [preprint](#) is een voorpublicatie van een artikel, voordat de peer review heeft plaatsgevonden.

### Opdracht

- Bekijk het filmpje: [Open Science: what, how & why?](#)
- Bespreek mogelijke voor- en nadelen van Open Science met elkaar.