

Lectoraat
Responsible IT



Hogeschool
van Amsterdam

Handreiking voor verantwoord datagebruik

Arjan Koning

Nanda Piersma

Hogeschool van Amsterdam

Versie 13-06-2022

Creating Tomorrow

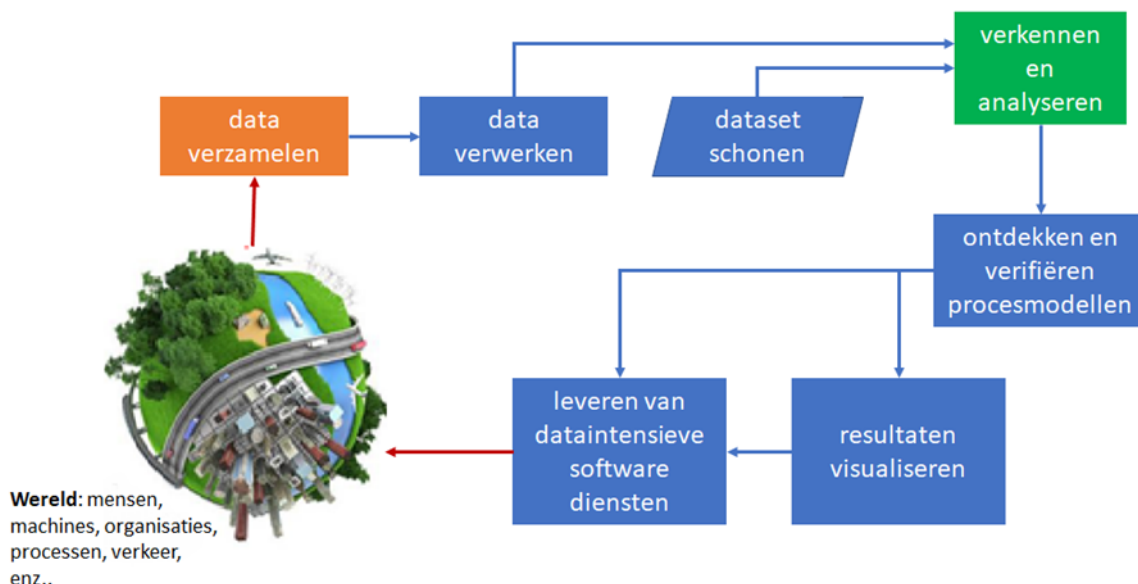
Inhoud

Introductie	3
1. Data vraagstuk	4
Wenselijkheid data-analyse	4
Belastbaarheid deelnemers dataproject	5
Belangen van deelnemers aan het dataproject	5
Ethische Context	6
CHECK DATAVRAAGSTUK	7
2. Data kwaliteit	8
Essentiële dataniveau	8
Bias	8
Privacy	9
Duurzaamheid	9
CHECK DATAKWALITEIT	10
3. Datamanagement	11
Eigenaarschap van de data	11
Opslag	12
Toegang	13
Archivering	13
Data delen en hergebruik	13
CHECK VERANTWOORD DATA MANAGEMENT	15
4. Bestaande protocollen en richtlijnen over datamanagement	16
FAIR principes	16
FACT-principes	17
ROBUST	17
SHARED	18
DEDA	18
TADA	18
5. Referenties	19
Over ons	22

Introductie

Het gebruik van data voor het inrichten en optimaliseren van dagelijkse- en bedrijfsprocessen heeft een enorme vlucht genomen. Het doel van het gebruik van data is om betere beslissingen te nemen, verbeterde, gepersonaliseerde services aan te bieden of beter inzicht te krijgen.

Een data project bevat de volgende onderdelen:



Vrij naar Wikipedia, 2014

Figuur 1 Dataproject

Het gebruik van data vraagt een andere manier van denken, van het anders inrichten van het bedrijfsproces en van de expertise benodigd in het team om succesvol data gedreven te zijn. In de Godatadriven survey 2019-2020 geeft 51% van de geïnterviewde bedrijven aan dat hun grootste probleem bestaat uit de kwaliteit van de data; ze hebben niet de juiste informatie om data gedreven te kunnen werken. (GoDataDriven, 2020) Dit staat in schril contrast met de enorme hoeveelheid data die wereldwijd wordt verzameld en bewaard.

Recente aandacht voor transparantie van dataprocessen en ethische richtlijnen geven een nieuwe dimensie aan de taken van een datamanager en data analisten. Er wordt veel geschreven over ethische kaders bij data gedreven activiteiten, met voorbeelden waar data wordt misbruikt ((Zuboff & Schwandt, 2019) (O'Neil, 2016), (Buitenweg, 2021), (Stikker, 2019)) en voorbeelden van onbedoelde effecten door datagebruik, vaak in combinatie met artificiële intelligentie ((Broussard, 2019), (Crawford, 2021)). Ook uitsluiting van (data-) infrastructuur en van publieke services voor bepaalde groepen mensen, profilering en surveillance van personen, discriminatie door biases in data zijn al veelbesproken onderdelen van het publieke debat over digitale technologie (Kantayya, 2020). Digitale identiteit is nog niet goed juridisch en ethisch vastgelegd en dat maakt dat vooral Tech-bedrijven bijna ongelimiteerd persoonlijke data mogen verwerven en gebruiken. Ook wordt in toenemende gevallen het verplicht delen van persoonlijke data een voorwaarde om een essentiële service te kunnen gebruiken, zoals een Microsoft-account om een laptop en de bijbehorende

software te gebruiken, of een benodigd Google-account om een moderne auto te kunnen rijden en navigatie te gebruiken.

Als tegenbeweging komen er nu richtlijnen, checklists en wetten om datagebruik verantwoord te maken. Twee grote bewegingen zijn **FAIR** (findable, accessible, interoperable en reusable) en **FACT** (Fair, accurate/accountable, controllable en transparant). Er ontbreekt een dimensie in de uitvoering van data projecten, die meer is dan FACT en FAIR principes toepassen. Het **ROBUST**-principe (Resilient, Open, Beneficial, User-oriented, Secure, Trustworthy) is daartoe ontwikkeld en ook het **SHARED**-principe (Sustainable, Harmonious, Affective, Relevant, Empowering, Diverse), die meer op het algeheel gebruik van data en het vertrouwen van burgers zijn gericht. Een overzicht wordt behandeld in hoofdstuk 4.

Deze handreiking kijkt naar het datagebruik als geheel, het vraagstuk waar de data bij wordt gebruikt, de manier waarop de data wordt ingezet en het management van de data.

Het doel van dit rapport is om een gebruiker, de datamanager en de dataproject ontwikkelaars bewust te laten worden van de **acties** om verantwoord datagebruik te realiseren, om handreikingen geven om op een verantwoorde, juridische, ethische en kwalitatief goede manier met data om te gaan. Het omvat een verwijzing van de bestaande detail checks, maar is vooral bedoeld om een geïntegreerd begrip van een verantwoord datagebruik te geven. Deze worden behandeld in hoofdstuk 1 t/m 3.

De aspecten van verantwoord datagebruik die in dit rapport worden behandeld zijn weergegeven in figuur 2:

DATA VRAAGSTUK	DATA KWALITEIT	DATA MANAGEMENT	
WENSELIJKHEID DATA ANALYSE	ESSENTIEEL DATA NIVEAU		EIGENAARSCHAP
BELASTBAARHEID	PRIVACY	BIAS	OPSLAG
BELANGEN DEELNEMERS			ARCHIVERING
ETHISCHE CONTEXT	DUURZAAM		DELEN/ HERGEBRUIK

Figuur 2: Dimensies van verantwoord datagebruik

1. Data vraagstuk

Traditionele datamanagementhandleidingen besteden nauwelijks aandacht aan het data-vraagstuk, terwijl hier juist de basis van verantwoord datagebruik wordt gelegd. Welk vraagstuk gaan we op proberen te lossen? Past het binnen onze doelstellingen en ethische context? Kunnen de gegevens niet ook voor een ander, ongewenst doel worden gebruikt of hebben ze ongewenste neveneffecten?

Wenselijkheid data-analyse

Een eerste overweging om te maken is of het vraagstuk wel om een data-analyse vraagt; of de geplande data-analyse de meest gepaste manier is om het vraagstuk op te lossen.

De belastbaarheid van de mensen is een belangrijke ethische afweging voor experimenten waar

mensen bij betrokken zijn. Meestal is de data zonder toestemming van mensen verkregen, of vanuit voorwaarden die geen opt-out functie hebben (het recht om niet mee te doen aan de dataverzameling). Verantwoord werken begint dus bij het op te lossen probleem. Hoe kan ik antwoord op de vraag krijgen? Als data nodig mocht blijken om het probleem op te lossen, welke data is er dan nodig? En als ik data verzamel, wat kan er nog meer met de data worden gedaan, willen we dat?

Voorbeeld

Bij het monitoren van drukte met gebruik van camerabeelden wordt data verzameld waar mensen herkenbaar in beeld worden gebracht. Met anonimiseringsalgoritmes (blurring) worden de mensen onherkenbaar, en kan de drukte worden gemeten. Maar is dit wel de beste methode om de drukte te meten? Drukke monitoren kan ook met geluid, sensoren, of een slim draaihek. De camerabeelden kunnen ook aanwezigheid van een individu vaststellen en zelfs het gedrag van individuen.

Er is dus een eerste check nodig naar de juistheid (adequaat) van dataoplossing en de effectiviteit van dataoplossing.

Belastbaarheid deelnemers dataproject

De belastbaarheid van en (potentiële) impact op de mensen is een belangrijke ethische afweging voor experimenten waar mensen bij betrokken zijn. Doen we mensen geen pijn, laten we ze geen activiteiten ondernemen die (te) belastend zijn of die hun integriteit schaden? Dezelfde afweging is ook van toepassing op het gebruik van data; het betreft vaak data van personen, waarvan het gebruik mogelijk onnodig is voor het vraagstuk en die in extreme situaties ook belastend kan zijn.

Er worden vaak te veel datakarakteristieken verzameld. Denk aan de individuele aankoopdata in een supermarkt, waaruit persoonskenmerken kunnen worden afgeleid zoals leeftijd, geslacht, postcode en soms zelfs familieverband en opleidingsniveau. Deze data is niet nodig voor de transactie in de supermarkt, maar wordt gebruikt voor data-analyses van koopgedrag. De klant zal dan andere aanbiedingen krijgen, voor sportsokken, of een Netflix abonnement op basis van klantprofielen. Maar ook zijn er voorbeelden van mensen die bijvoorbeeld een “blijje doos” kregen toegestuurd omdat de data uitwees dat de klant zwanger is, iets wat bij de familie in kwestie nog niet bekend was.

Een ander voorbeeld is het moeten verstrekken van (persoons)gegevens om een online service te kunnen gebruiken, zoals online aankopen van tickets voor een evenement. Dit is feitelijk belastend, zeker als er gegevens worden gevraagd die niet nodig zijn voor de service-activiteit, maar alleen voor analyse of profilering.

Belangen van deelnemers aan het dataproject

De data-analyse zal een **doel** hebben dat verschillend kan zijn voor specifieke deelnemers. De inzichten die de data opleveren kunnen ook anders zijn voor verschillende deelnemers. Het is belangrijk om voorafgaand aan de data-analyse vast te leggen wie de analyse op welke wijze kan gebruiken.

Voorbeeld

Er wordt door een school een app gebouwd om studenten te helpen een vrije studieplek in een (druk) schoolgebouw te vinden. Daarbij wordt gebruik gemaakt van Wi-Fi data, vanuit elk access point in het gebouw wordt gemeten hoeveel contacten er worden gemaakt door mobiele telefoons

en laptops. De data wordt geanonimiseerd opgeslagen. De app geeft real-time met een heatmap (kleurencode) weer wat de drukte is in het gebouw (per verdieping, en zelfs per zone op de verdieping). De gebruiker kan aan de kleur zien waar het rustig is om daar een studieplek te vinden. Het vraagstuk was echter waar vrije studieplekken zijn, en niet waar het druk is. Er wordt dus andere informatie verzameld en in de app gegeven dan het vraagstuk nodig heeft. De app zou voor de studenten alleen de rustige locaties moeten laten oplichten, het liefst nog met een mogelijkheid om de beschikbare studieruimte te reserveren. Voor gebouwbeheer is het waarschijnlijk wel belangrijk om de drukke plekken in beeld te hebben, bijvoorbeeld voor schoonmaak.

Duidelijkheid over de belangen voor elke deelnemer aan het project kan voorkomen dat er een mismatch is tussen de data oplossing en het vraagstuk, maar daarmee kan ook in een vroeg stadium andere doelen vanuit de data worden geïdentificeerd, in het bijzonder ongewenste doelen. Zo is de app in het schoolgebouw in deze vorm ook te gebruiken als handhavingstool om te zien of mensen op bepaalde plekken in het gebouw verblijven, in het bijzonder als dat niet is toegestaan.

Ethische Context

We leven in een samenleving waar data en AI een steeds belangrijkere rol spelen in relatie tot onze menselijke waarden. We realiseren ons dat normen en waarden tijd- en plaats afhankelijk zijn. Zo hebben we, naast privacywetgeving als AVG/GDPR, te maken met uitgedragen waarden, met de SDG's van de VN, of met de gemeente Amsterdam die het Tada manifest (Tada, 2021) heeft opgesteld. Organisaties adopteren waardenkaders en gebruiken deze om de identiteit en de activiteiten van de organisatie te richten.

Elk (data) project zal moeten passen binnen de waardenkaders van de organisatie. Een check daarop lijkt overbodig, maar is dat zeker niet. Veel organisaties willen openheid en transparantie van hun bedrijfsprocessen, maar hebben een data management protocol waarbinnen data nooit toegankelijk is. Als duurzaamheid (klimaatbestendig werken) een waarde van de organisatie is, dan kunnen er vraagtekens worden gezet bij het uitvoeren van een dataproject dat nodeloos gebruikt maakt van zware cloudtoepassingen en grote hoeveelheden data.

Niet alleen de waarden van een organisatie (waar staan we voor), ook de ethische code van het bedrijf is een context voor data vraagstukken. Albert Heijn/Delhaize heeft in de ethische code staan: "Persoonlijke gegevens van klanten en medewerkers mogen uitsluitend voor legitieme bedrijfsdoeleinden worden gebruikt, voor zover wettelijk toegestaan." Een data project zal dus altijd binnen de kaders van een legitiem bedrijfsdoel en binnen wettelijke regels moeten plaatsvinden.

De vraagstukken die beantwoord moeten worden zijn ook altijd ingebed in de regelgeving en de normen en waarden van de maatschappij waarin geleefd wordt. De Amerikaanse wet is echt anders dan de Europese, of de Chinese. Met data kunnen we dus niet zomaar alles doen.

CHECK DATAVRAAGSTUK

Wenselijkheid data-analyse

- a. Wat is data voor het vraagstuk dat opgelost dient te worden?
- b. Is de analyse om het vraagstuk op te lossen gepast?
- c. Welke alternatieven zijn er om het vraagstuk te kunnen beantwoorden (zonder data of met andere data)?

Belastbaarheidscheck

- a. Welke (delen van de) data zijn nodig voor de analyse van het vraagstuk?
Maak onderscheid in
 - Essentieel
 - Mogelijk interessant/ Nice to have
 - Onnodig/ongewenst?
- b. Is het noodzakelijk dat (al) de data wordt bewaard? Is het mogelijk om data te verwijderen?
- c. Is er een mogelijkheid om als deelnemer terug te komen op eerder gegeven toestemming / Kun je bij nader inzien afzien van het gebruik van de door jou geleverde data?
- d. Kan de data worden gebruikt voor ongewenste doelen?

Belangen deelnemers

- a. Bepaal de gebruikers en de eigenaar van elk bestand van de data-analyse
- b. Bepaal het doel van de data analyse voor elke deelnemer
- c. Check (samen met de betrokkenen) andere toepassingen en doelen met dezelfde data

Ethische context

- a. Past het datavraagstuk bij de missie en visie van de organisatie waarbinnen/waarvoor je deze uitvoert?
- b. Hoe verhoudt de analyse zich tot de door de organisatie geadopteerde (ethische) richtlijnen, zoals TADA, social development goals of donut economie principes?
- c. Zijn er internationale regels en waardensystemen die invloed hebben op het datavraagstuk?

2. Data kwaliteit

Als de voorwaarden voor het uitvoeren van het datavraagstuk zijn vastgesteld, dus de wenselijkheid van de data-analyse, de belastbaarheid door het gebruik van data, de belangen van de deelnemers en de ethische context van de organisatie, dan kunnen we aan het werk met de datasets zelf. Juist in het gebruik van datasets komen de principes van verantwoord werken met data tot leven. Data-integriteit en datakwaliteit is onderdeel van goed datagebruik en daarmee van verantwoord datagebruik. Denk daarbij aan kwaliteitschecks op ontbrekende gegevens, niet relevante outliers, niet betrouwbare data, enzovoorts (Baer, 2019). Met een bril van verantwoord datagebruik zal de kwaliteitscheck uitgebreider worden, met name wat betreft het bepalen van essentiële datakarakteristieken, privacy issues en bias in de dataset. We bespreken ook het energieverbruik van data, als onderdeel van de kwaliteit van data.

Essentiële dataniveau

Voor het datavraagstuk zijn al de belastbaarheid en de wenselijkheid van de data-analyse benoemd. Er is dus de overtuiging dat de data moet worden verzameld en geanalyseerd en dat de juiste karakteristieken worden verzameld. Het instrumentarium waarmee de data worden verzameld kan naast essentiële data ook andere data bevatten die noodzakelijkerwijs meekomen met het instrumentarium waarmee de data is verzameld. Maak onderscheid tussen een bronbestand en een dataset waar de data-analyse op wordt uitgevoerd. Door niet essentiële data uit het bronbestand weg te filteren in de dataset die wordt gebruikt voor de analyse voorkom je mogelijk ongewenst datagebruik.

Bias

In de gebruikte of te verzamelen data kan een ongewenst vooroordeel zitten (bias). Er zijn meer dan 300 verschillende soorten bias zijn in menselijke gedrag (Bowles, 2018). Daarmee ontkomt een dataset meestal niet aan afwijkingen, scheve weergaves, vooroordelen, discriminatie, misrepresentatie en nog veel andere vormen van bias. Een bias kan op allerlei manieren ontstaan, zelfs in de instellingen van de apparatuur om data te verzamelen kan een afwijking zitten. Het gebruik van datasets met een bias zorgt ervoor dat vooringenomen eigenschappen ingebed worden in de data-analyse en dus in de uitkomsten van de analyse.

Het identificeren van bias in datasets wordt uitgebreid beschreven in Responsible AI (Agarwal, 2021). Het is gebaseerd op een vergelijking van de uitkomsten van een data-analyse met de kenmerken van de dataset die wordt gebruikt in de analyse. De kenmerken in de dataset kunnen worden opgesplitst in onafhankelijke kenmerken en kenmerken met persoonlijke, sociale of raciale kenmerken, de beschermdde kenmerken genoemd. Met de beschermdde kenmerken kunnen bevoorrechte groepen worden geïdentificeerd, afhankelijk van het probleem dat wordt onderzocht. Bekende voorbeelden van bias in datasets zoals de Amazon case waar vrouwen met een grotere kans afgewezen werden door sollicitatiesoftware. In dit geval is gender het beschermdde kenmerk en de mannen vormen de bevoorrechte groep. Bias of fairness metrics die hieruit kunnen worden berekend meten de afhankelijkheid van de uitkomsten (positief of negatief) en de bevoorrechte groepen. In een eerlijke data-analyse zou de uitkomst onafhankelijk moeten zijn van de bevoorrechte groepen binnen de beschermdde kenmerken. Ook de onafhankelijke kenmerken kunnen gecorreleerd zijn met beschermdde kenmerken (zoals postcodes met sociale status). Ook kan er sprake zijn van oververtegenwoordiging van een bepaald (bevoorrechte) groep in de data, waardoor de data-analyse wordt verstoord. Bij het vergaren of verwerken van data is dus het gewenst om meerdere bias-checks te doen, dit is een vrij technische maar noodzakelijke activiteit, zie bijvoorbeeld ook (Baer, 2019) en (Barocas, 2022).

Privacy

Het gebruik van privé gegevens is strikt gebonden aan het doel van het data vraagstuk en moet in ieder geval altijd gedaan worden met toestemming van de mensen die in de dataset voorkomen (informed consent). Het hergebruik van de gegevens om andere doelen na te streven is dan ook niet toegestaan (Vogel & al, 2018). Ook het vastleggen van data die tot individuen terug te leiden is mag niet zonder toestemming. Tracking is het volgen van eenzelfde ID eenheid (aangegeven met een persoonsnaam of geanonimiseerd) over meerdere tijdsdimensies of plaatsbepalingen, waardoor identificatie kan plaatsvinden. Bij de wetgeving (AVG) zijn ook datasets met de mogelijkheid van persoonsidentificatie door tracking verboden.

van Dale

duur·zaam

(bijvoeglijk naamwoord, bijwoord; vergrotende trap: *duurzamer*, overtreffende trap: *duurzaamst*)

- 1 lang durend
- 2 weinig aan slijtage of bederf onderhevig
- 3 het milieu weinig belastend

Duurzaamheid

Er wordt vandaag de dag massaal data verzameld en de informatie daaruit wordt weer in duizenden andere databestanden gekopieerd. Datacenters gebruiken veel stroom, zo hebben ze uitgebreide koelinstallaties nodig. Het aantal providers dat “groene” energie gebruikt en die zuinigere dataverwerking toepassen groeit (zie bijvoorbeeld www.thegreenwebfoundation.org), maar het gebruik van energiezuinige rekencentra

alleen is niet voldoende. Het is daarom verstandig om je af te vragen welke data je verzamelt, bewaart en gebruikt in termen van duurzaamheid. Het verwerken van data kost ook energie, algoritmes zonder onnodige loops zullen efficiënter werken en duurzamer zijn. Duurzaam is hier dus ook een kwalitatief meetinstrument.

Het woord duurzaam heeft meerdere betekenissen. We bespreken hier duurzaamheid in termen van energiegebruik en belasting van het milieu. Als de data hergebruikt kan worden dan is dat duurzamer dan dat de data weer moet worden samengesteld. Aan de andere kant moet de data dan weer bewaard worden. Simpele rekenregels zijn beschikbaar om de CO-2 uitstoot per GB data te berekenen (Stanford Alumni Association, 2021).

Verantwoord datamanagement begint daarom ook vanuit het duurzaamheidsaspect aan de start van een data gedreven procesaanpak: welke data hebben we nodig en welke niet?

Om bij te dragen aan een langere levensduur van een data verwerkend systeem zal ook inzichtelijk moeten zijn hoe het werkt. Transparantie en onderhoudbaarheid zijn zo onlosmakelijk aan elkaar verbonden.

CHECK DATAKWALITEIT

Essentieel data niveau

- Als je zelf data verzamelt, check of de meetapparatuur geen ongewenste data verzamelt.
- Als je zelf data verzamelt, check of de meetapparatuur geen groepen uitsluit (bijvoorbeeld een zwarte huid, of kleine mensen).
- Als je zelf data verzamelt, maak twee databestanden: een bronbestand met alle data (dus essentiële en mogelijk extra data) en maak daaruit een werkbestand met alleen essentiële data.
- Als je databestanden verwerft van anderen, maak afspraken om alleen essentiële data te verwerven met een mogelijkheid om na overleg en met redenen toegang te krijgen tot het bronbestand.

Bias

- a. Check op over/ondervertegenwoordiging van bepaalde groepen
- b. Check op afwijking in karakteristieken voor bepaalde groepen
- c. Check op ongewenste indeling op karakteristieken (Agarwal, 2021).

Privacy

- a. Check in de dataset welke persoonsgegevens aanwezig zijn.
 - Verwijder of anonimiseer onnodige persoonsdata uit de dataset (Piersma, 2021)
 - Informeer alle personen die in de dataset voorkomen met benodigde persoonsdata volgens het principe vragen, informeren en vastleggen.
- b. Bij het gebruik van datasets met meerdere dimensies (bijvoorbeeld tijd-plaats): check of personen niet kunnen worden geïdentificeerd door tracking.
- c. Bij het gebruik van meerdere datasets waarin eenzelfde persoon (anoniem) voorkomt: check of door het combineren van meerdere datasets persoonsidentificatie kan plaatsvinden.

Duurzaam

- a. Wat is het energiegebruik van dataopslag en van de data-analyse?
- b. Hoe kan de dataset verder worden gebruikt om de opslag ervan te rechtvaardigen?
- c. Wat is de beste dataset format voor duurzaam datagebruik?
- d. Wat is de noodzakelijk bewaartermijn van de dataset, en wat is het mechanisme van verwijdering?

3. Datamanagement

Het verantwoord omgaan met de datasets als onderdeel van het dataproject, het datamanagement, heeft zowel een technische als een management dimensie. Er zijn voor beide dimensies al veel richtlijnen en werkwijzen. De handreiking op dit gebied is daarom een verwijzing naar de bestaande protocollen met hier en daar een duiding van verantwoord gebruik van datasets. Technisch gezien is verantwoord datamanagement goed beschreven door de FAIR principes. De management dimensie omvat een uitbreiding van de bestaande managementafspraken tussen partijen.

Eigenaarschap van de data

Hardware in bezit van een partij die data verzamelt geeft automatisch eigenaarschap van de dataset aan de eigenaar van de hardware. Een sensor die temperatuur meet maakt een dataset die eigendom is van de eigenaar van de sensor. Bij sensing devices voor persoonlijk gebruik, zoals een Fitbit apparaat, die via online apps/software data opslaan in cloud-omgevingen is de software organisatie die de app beheert de eigenaar van de data. Afspraken met individuele gebruikers worden vrijwel altijd eenzijdig door de leveranciers opgelegd in de verkoopvoorwaarden of via het verplichte gebruik van specifieke apps om de data in te zien. De bezitter van het individuele Fitbit horloge kan de data dan wel inzien, maar heeft door het gebruik van de app het eigenaarschap overgedragen aan de software organisatie van de app. Daarmee wordt eigenaarschap van software gekoppeld aan eigenaarschap van data, in plaats van hardware-data eigendom relatie. Dit principe staat onder druk; mag Fitbit mijn hartslagdata gebruiken voor andere doeleinden dan het delen van mijn sportprestaties aan mijzelf?

Eigenaarschap van data kan via licentieafspraken met hardware of software partijen worden afgesproken, bijvoorbeeld licentiehouders van bedrijfssoftware (een personeelsadministratie, of Wi-Fi netwerk-gebruiksdata). Hierbij wordt via afspraken in de licenties met de softwareleverancier afgesproken wie dataeigenaar is. Dit kan per datakarakteristiek verschillen, een Wi-Fi toegangspunt slaat alle connecties op tijd en plaats op, met identifiers voor de connecting devices. De hardware producent (licentiehouder) heeft slechts zicht op connectievolumes per zone in een gebouw, de klant heeft zicht op alle details van elke gemaakte connectie op elke individuele locatie in het gebouw.

Online-services van BigTech partijen zoals Twitter of Facebook bepalen dat het (gratis) gebruik van de software betekent dat een gebruiker de data afstaat en dat de datasets van het BigTech-bedrijf worden gevoed met de data van de gebruikers van de service. Het gebruik van de software impliceert een akkoord van de gebruiker met de voorwaarden (informed consent). Ook mogen deze partijen volgens hun voorwaarden koppelingen maken tussen verschillende services (Fitbit data worden gekoppeld aan Facebook-gebruik, en aan de contactenlijst van Google-mail). Daarmee verrijken de BigTech-bedrijven de datasets in hun eigendom en vormen zij een datamacht die grote invloed heeft.

Omdat data een relatief nieuw onderwerp is de regelgeving nog niet in staat om ongewenst gebruik van data en datamacht te voorkomen. De AVG (Algemene Verordening Gegevensbescherming) valt onder de juridische kaders en zegt dat er geen data mag worden gedeeld waarin privacygegevens staan zonder toestemming van de betrokkene. Ook mag er geen data worden gedeeld waaruit persoonlijke gegevens kunnen worden afgeleid. De BigTech methode van informed consent door het gebruik van een gratis service is juridisch houdbaar, maar staat ethisch onder druk.

Op terreinen waarvoor (nog) geen juridische kaders bestaan of waar die onvoldoende bruikbaar zijn, kunnen morele regels verder gaan. Bijvoorbeeld als het gaat om het doorverkopen van gegevens, kan het sterk van de context afhangen of dit moreel laakbaar is. (Vallor, 2016)(p. 17).

“Ook op het terrein van privacy, dataficatie (al het gedrag van burgers vastleggen in gegevens) en dataveillance (surveillance op basis van gegevens) liggen ethische uitdagingen. Deze morele regels kunnen ook eenvoudig op het hergebruik van gegevens worden toegepast.” (Custers, 2019)

Het voorkomen van dataficatie en dataveillance bij datavraagstukken kan plaatsvinden via een check op de gebruikte hardware en software.

Opslag

De opslag van data lijkt evident, de datasets staan ofwel op een kennisdrager in eigen eigendom (laptop, externe harde schijf, server), ofwel op een cloudomgeving (datacenter). Ook staat de data wel op de sensing devices (zoals een smart camera of een Tesla computersysteem) of in een analyse omgeving (IBM Watson, R studio), meestal ook in een cloudomgeving.

Dataopslag moet aan een aantal voorwaarden voldoen om verantwoord genoemd te kunnen worden, te weten veilig, duurzaam, en gecontroleerd.

Data kan op meerdere manieren veilig worden opgeslagen. Niet alleen mag niet iedereen de data zien (wanneer het bijvoorbeeld om privé- of bedrijfsgegevens gaat) maar ook mag de data zonder wijzigingshistorie niet worden aangepast (vervuild) en daarmee onbruikbaar worden.

Veiligheid wordt op twee manieren geïnterpreteerd.

1. Het mag niet ongeautoriseerd toegankelijk zijn.
2. De data mag niet kwijt raken of ongeautoriseerd gewijzigd worden.

Veiligheid staat daarmee in nauw verband met autorisatie. Er moet een functionerend autorisatiesysteem zijn én de data moet dusdanig veilig worden opgeslagen om te voorkomen dat dat de beoogde autorisatieregels omzeild kunnen worden. Dit geldt tijdens het gehele proces van de dataverwerking van bron, via eventuele tussenbestanden tot de uiteindelijke dataset waaruit de conclusies worden getrokken (informatie) waarop vervolgacties worden gebaseerd.

Zonder alle regels van security op te noemen, zijn er standaard werk-etiquettes zoals wachtwoord vereisten, niet laten slingeren/delen van inloggegevens, het verwijderen/uitzetten van verlopen accounts, die altijd gehanteerd moeten worden. Voor data in clouds en op servers met externe connectiemogelijkheden is cyber security bijzonder actueel. Elke eigenaar is gehouden aan het nemen van voldoende maatregelen om de data te beschermen voor onrechtmatig toegang en gebruik. Om de data niet kwijt te raken zal een backup systeem geregeld moeten zijn.

Onder duurzame dataopslag verstaan we naast het energievraagstuk voor dataopslag ook het gegeven dat niet data niet onnodig (lang) wordt bewaard of onnodig gedupliceerd wordt.

- a. Kunnen mensen of algoritmes die bij de data moeten komen deze ook vinden en benaderen?
- b. Is de data opgeslagen met de juiste metadata¹ (FAIR)?

¹ Metadata zijn gegevens die de karakteristieken van een dataset beschrijven, zoals auteur, de datum van schrijven, de uitgever, het aantal pagina's en de taal waarin de gegevens zijn opgesteld.

Toegang

De toegang tot de data wordt in een autorisatiesysteem geregeld. Er wordt gebruik gemaakt van aan individuen gekoppelde gebruikersprofielen en (iedere) toegang wordt gemonitord. Duurzaam datamanagement op het gebied van toegang bestaat uit het goed uitvoeren van de bestaande data management protocollen. Ook een eerlijk toegankelijkheidsprotocol voor alle belanghebbenden is onderdeel van verantwoord data management.

Archivering

Er zijn diverse, in de wet vastgestelde, regels (zoals AVG) die bepalen hoe lang data mag of moet worden bewaard.

Onderzoekdata moeten **minimaal** 10 jaar bewaard blijven na datum van de eerste publicatie van het onderzoek (of einde onderzoek). Tot persoon herleidbare gegevens hebben echter een bewaartermijn van **maximaal** 10 jaar (met een uitzondering voor persoonsgegevens uit medische dossiers, waarvoor max 15 jaar staat, met daarop weer de uitzondering voor projecten waar dit echt niet anders kan (longitudinaal onderzoek). De beschrijvingen van datasets (metadata) moeten echter langer blijven bestaan. FAIR Data Point (Fair data team, 2022) is een metadata service die toegang geeft tot metadata volgens de FAIR principes. *Eenheden waarin de data wordt opgeslagen zijn algemeen aanvaarde eenheden, zie daarvoor ook de go-fair website (GO FAIR, 2021).*

De archiveringsfase van een dataproject vraagt om een update van het essentiële dataniveau (welke gegevens uit de data moeten we archiveren), van toegang (voor wie vervalt toegang en hergebruik), en van duurzaamheidschecks (hoeveel energie verbruiken de alternatieve opslagmogelijkheden voor archivering).

Data delen en hergebruik

Er kunnen drie verschillende vormen van hergebruik van datasets worden onderscheiden, afhankelijk van de vraag of het doel en de context bij het hergebruik afwijken van het doel en de context bij het oorspronkelijke gebruik van de gegevens:

- Data recycling: gegevens worden hergebruikt voor hetzelfde doel in dezelfde context (maar door een andere partij).
- Data repurposing: gegevens worden hergebruikt voor een ander doel dan waarvoor ze oorspronkelijk zijn verzameld, maar het hergebruik vindt wel plaats in dezelfde context.
- Data recontextualisation: de gegevens worden hergebruikt in een nieuwe context, anders dan de context waarin ze oorspronkelijk zijn verzameld.

Openbare en open data

Openbare data is een vorm datamanagement waarbij de persoonlijke informatie of bedrijfswaarde gegevens worden omgevormd tot niet-individueel-traceerbare data, waarbij meerdere partijen inzicht kunnen krijgen in de informatie uit de data van anderen. Deze datasets worden openbaar gepubliceerd, en mogen worden gebruikt onder verwijzing naar de eigenaar (soms is toestemming voor gebruik noodzakelijk). De regelgeving verschilt per land verschilt. In de EU is het bijvoorbeeld niet toegestaan om van staatswege gegevens te moeten vrijgeven aan de overheid, terwijl dat in de VS wel mogelijk is.

Open data bestaat uit openbare datasets waarbij de persoonlijke informatie of bedrijfswaarde gegevens worden omgevormd tot niet-individueel-traceerbare ruwe data. Open datasets zijn **openbaar** en 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. Deze datasets mogen worden gebruikt onder verwijzing naar de eigenaar (soms is toestemming voor gebruik noodzakelijk). De bronvermelding is hierbij eerder een kwaliteitstoets van de data, en geen toestemmingsmechanisme. Open datasets maken bij voorkeur gebruik van FAIR principes om de vindbaarheid te verhogen en hergebruik te vergroten.

Data delen

In het hoofdstuk Data Vraagstuk is al stilgestaan bij de mogelijke gevaren van het gebruik van data voor andere vraagstukken, doelen of door andere partijen. Maar hergebruik van data en data delen is ook een best practice wat betreft duurzaamheid en kwaliteitsverbetering van algoritmes en services. Het samenvoegen van MRI scans data van meerdere ziekenhuizen zal tot betere tumor detectie algoritmes kunnen leiden, met alle gezondheidsvoordelen. Ook is het verzamelen van dezelfde data(kenmerken) door meerdere partijen minder duurzaam, dan het opnieuw gebruiken van dezelfde dataset. In academische studies zijn gedeelde datasets zelfs noodzakelijk voor verificatie en kwaliteitschecks van gepresenteerde modellen en resultaten.

Openbare en open data zijn vrij te gebruiken en daarmee is de eigenaar juridisch niet meer verantwoordelijke voor het verantwoorde gebruik daarvan. Verantwoord data management bij hergebruik van datasets is gericht op het voorkomen van ongewenst gebruik van bestaande datasets. De data eigenaar is juridisch wel verantwoordelijk voor hergebruik van beschikbaar gestelde niet-openbare datasets. Data deel protocollen geven duidelijke richtlijnen over eigenaarschap, toegang, en intellectueel eigendom van niet openbare data. Ethisch gezien kan de oorspronkelijke eigenaar in beide gevallen alleen maar inschatten hoe de data gebruikt kan worden (zie checks). Voor niet openbare datasets kan de eigenaar protocollen uitbreiden met afspraken over verantwoord datagebruik.

CHECK VERANTWOORD DATA MANAGEMENT

Eigenaarschap

- a. Bij datasets van anderen: Wie is de eigenaar van de gebruikte dataset?
- b. Bij datasets van anderen: Wat zijn de voorwaarden voor gebruik (hardware, software, licentie)?
- c. Als je zelf data verzamelt: Via welke hardware en software partijen wordt de data verzameld en beschikbaar gemaakt? Wat betekent dit voor eigenaarschap en datamacht?
- d. Welke juridische wetgeving (land) is van toepassing op de datasets, de hardware voor opslag en de software voor toegang?

Opslag

- a. Veilig, duurzaam: Waar staat de data?
- b. Gecontroleerd: Wie beheert de data?
- c. Is de metadata toegankelijk en leesbaar?

Toegang

- a. Wie beheert de data (ook na afloop van het dataproject)?
- b. Worden de beheerders ook opgevolgd?
- c. Zijn de dataprotocolen bekend bij alle belanghebbenden van het datavraagstuk?
- d. Is er sprake van eerlijke datatoegang voor alle belanghebbenden van het datavraagstuk?

Archivering

- a. Wat is de bewaartermijn van de data? Vinden ook acties plaats als deze termijnen actie vragen, door wie?
- b. Update van het essentiële dataniveau (welke gegevens uit de data moeten we archiveren)?
- c. Update van toegang (voor wie vervalt toegang en hergebruik)?
- d. Duurzaamheid: hoeveel energie verbruiken de alternatieve opslagmogelijkheden voor archivering?

Hergebruik

- a. Is er duidelijkheid over het vraagstuk waarvoor de data wordt hergebruikt?
- b. Is het mogelijk om data te verwijderen of toe te voegen, waardoor de dataset wordt gecompromitteerd?
- c. Is er een mogelijkheid om terug te komen op eerder gegeven toestemming van gebruik van de geleverde datasets?

4. Bestaande protocollen en richtlijnen over datamanagement

Het werken met data heeft een grote vlucht genomen en gelukkig is bij velen doorgedrongen dat onverantwoord met data omgaan consequenties kan hebben. Voor een wereldwijde meta-analyse van ethische richtlijnen voor digitale activiteiten, zie (AI Ethics lab, 2022). De Digital society (The digital society, 2022) heeft een aantal dataprotocollen op een rijtje gezet. De protocollen worden vaak aangeduid met een reeks beginletters die stuk voor stuk weer een (Engels) woord vormen. Hier worden ze kort genoemd.

FAIR principes

FAIR (Findable, Accessible, Interoperable en Reusable) in het Nederlands: Vindbaar, Toegankelijk, Uitwisselbaar, Herbruikbaar (GO FAIR, 2021), (Wilkinson, Michel Dumontier, & al, 2016).

Hoewel het woord “fair” zelf al een ethische lading heeft zijn de FAIR principes vooral voorwaardenscheppend voor kennis delen. FAIR gaat niet over wat er met de data gebeurt, maar zorgt ervoor dát de data gecontroleerd (her)gebruikt kan worden.

Findable / Vindbaar

De eerste voorwaarde om onderzoeksgegevens te kunnen hergebruiken is om ze te kunnen vinden. Metadata zijn gegevens die de karakteristieken van de data in de dataset beschrijven als structuur, context, betekenis, onderlinge relaties, locatie, status, eigenaarschap, oorsprong, etc. De metadata van de gegevens moeten eenvoudig te vinden zijn (voor zowel mensen als computers).

Metadata die door computers gelezen moeten kunnen worden zijn essentieel om datasets en services te kunnen vinden. Daarbij moeten de gegevens een eenheid hebben die algemeen gebruikt wordt.

1. (Meta) data hebben een “unique en persistent identifier”
2. De data worden uitgebreid beschreven met metadata.
3. De metadata bevatten zelf de identifier van de dataset die ze beschrijven.
4. (Meta)data worden opgeslagen in een doorzoekbare bron.

De (meta)data worden op een plek opgeslagen die algemeen bekend is en goed bereikbaar is.

Accessible / Toegankelijk

De gebruiker van de dataset moet via de metadata duidelijk hebben hoe de data kan worden opgehaald. De identificatie- en autorisatieprocedures staan vermeld.

1. (Meta)data kunnen worden opgehaald op basis van identifiers via een standaard communicatieprotocol.
 - a. Het protocol is open, gratis, and universeel te implementeren;
 - b. Het protocol maakt waar nodig een authenticatie- en autorisatieprocedure mogelijk.

Zaken als waar je moet zijn om autorisatie aan te vragen (een site, e-mailadres) zouden hier bij moeten staan. Het is niet gezegd dat deze data ook gratis of per se voor iedereen toegankelijk hoeft te zijn, maar de metadata moeten deze gegevens wel bevatten.

2. Metadata zijn toegankelijk en blijven toegankelijk, ook als de onderliggende data niet meer beschikbaar is.

Interoperable / Uitwisselbaar

Om data te kunnen combineren moet de data gelezen kunnen worden door algemeen beschikbare software, maar ook met tools waarmee je die data verwerkt en in opslaat.

1. Zet de (meta) data daarom in een formele, algemeen bekende taal om kennis te delen.
2. Gebruik voor (meta)data “vocabularies” die ook de FAIR principes volgen (Fair Data Point, 2021).
3. De verwijzingen in de (meta)data worden kwalitatief beschreven met gebruikmaking van de identifiers.

Reusable / Herbruikbaar

Het uiteindelijke doel van FAIR is om (onderzoeks-)gegevens zo goed mogelijk te hergebruiken. De de (meta) data moeten daarom zo goed mogelijk beschreven worden zodat ze in andere toepassingen kunnen worden gebruikt.

Als aan de eerste drie letters van FAIR wordt voldaan is de vierde letter eigenlijk al klaar. FAIR zegt daarover:

(Meta)data worden uitgebreid beschreven met een veelheid van precieze en relevante attributen.

- 1.1. (Meta)data worden gepubliceerd met een heldere en toegankelijke data gebruik licentie.
- 1.2. De herkomst van de (meta)data wordt duidelijk beschreven.
- 1.3. (Meta)data voldoen aan de standaarden van het domein waarin ze gebruikt worden.

De metadata zijn dus het technische hart van FAIR principes van datamanagement. Het ondersteunt ook duurzaam gebruik van data op de tweede duurzaamheidsduiding: het gebruik en hergebruik van dataset. Bij het maken van een dataset (de verzameling) is deze technische exercitie dus belangrijk als hergebruik en delen van de data belangrijk is. Daarmee is het FAIR principe dus een belangrijke tool voor verantwoord datamanagement.

FACT-principes

FACT staat in dit geval voor Fair(eerlijk), Accurate (nauwkeurig), Confidential (vertrouwelijk), Transparent (transparant).

Eerlijk: Oneerlijke conclusies worden vermeden, ook al zijn ze correct berekend op basis van beschikbare data en modellen.

Nauwkeurig: Berekende antwoorden worden gegeven met een gegarandeerd niveau van nauwkeurigheid, zodat misleidende conclusies vermeden worden.

Vertrouwelijk: Resultaten worden bereikt op een veilige en gecontroleerde manier zonder dat geheime (privé- of zakelijke) informatie bekend wordt gemaakt.

Transparant: Berekende antwoorden kunnen op een zodanige manier worden begrepen en verhelderd dat ze onbetwistbaar en dus betrouwbaar zijn.

ROBUST

Resilient (veerkrachtig), Open, Beneficial (nuttig), User-oriented (gericht op de gebruiker), Secure (veilig), Trustworthy (betrouwbaar).

SHARED

“Wil een digitale samenleving open en democratisch blijven, dan moet de informatietechnologie op de mens gericht zijn en berusten op de SHARED-principes. Deze principes, die variëren van duurzaamheid tot diversiteit, vormen de bouwstenen voor een beter leefbare wereld.

Sustainable (duurzaam): De betrokkenheid van burgers moet zodanig worden gestimuleerd en georganiseerd dat ze blijft bestaan na de lancering van een project of de ontwikkeling van een programma.

Harmonious (harmonieus): Burgers moeten worden betrokken op manieren die inclusief zijn en niet in tegenspraak zijn met bestaande wetgeving, sociaal beleid en/of normen van goed burgerschap en beschaafd gedrag.

Affective: Burgers moeten zodanig worden betrokken dat uiteenlopende emotionele investeringen en zorgen van mensen met betrekking tot technologie en data worden erkend en gerespecteerd.

Relevant: Die burgers moeten betrokken worden die rechtstreeks te maken hebben met en de gevolgen ondervinden van de relevante aspecten en activiteiten.

Empowering: Burgers moeten zodanig worden betrokken dat ze een beter inzicht krijgen in ontwikkelingen en in de intellectuele en praktische hulpmiddelen waarmee ze zich hierover een mening en oordeel kunnen vormen. Ook moeten ze waar mogelijk toegang krijgen tot de technische hulpmiddelen om er invloed op uit te oefenen.

Divers: Burgers moeten worden betrokken op basis van erkenning van en ruimte voor de verschillende dimensies van diversiteit, waaronder gender, etniciteit, klasse, handicap, seksualiteit en religie.”

(Van de website the digital society (REFERENTIE)).

DEDA

De Ethische Data Assistent/Data Ethics Decision Aid (DEDA) is een instrument om ethische problemen te herkennen in dataprojecten, ontwikkeld door de Universiteit Utrecht en de Gemeente Utrecht. (Utrecht Data School, 2022).

TADA

Tada is het een data-manifest dat ontwikkeld is en mede ondertekend is door de Gemeente Amsterdam en is er op gericht dat iedere belanghebbende gelijkwaardig voordeel geniet (Tada, 2021).

Inclusief

Onze digitale stad is inclusief. We houden rekening met de verschillen tussen individuen en groepen, zonder gelijkwaardigheid uit het oog te verliezen.

Zeggenschap

Data en technologie moeten bijdragen aan vrijheid van bewoners. Data zijn dienend. Om het leven vorm te geven naar eigen inzicht, zelf informatie te verzamelen, kennis te ontwikkelen, ruimte te vinden om jezelf te organiseren.

Menselijke maat

Data en algoritmen hebben niet het laatste woord. menselijkheid gaat altijd voor. We laten ruimte voor onvoorspelbaarheid. Mensen hebben het recht om digitaal vergeten te worden. Zo blijft er altijd ruimte voor een nieuwe, schone start.

Legitiem en gecontroleerd

Bewoners en gebruikers hebben zeggenschap over de vormgeving van onze digitale stad. De overheid, maatschappelijke organisaties en bedrijven faciliteren dit. Zij monitoren de ontwikkeling en de maatschappelijke gevolgen.

Open en transparant

Welke data worden verzameld? Waarvoor? En met welke uitkomsten en resultaten? Daarover zijn we altijd transparant.

Van iedereen - voor iedereen

Data die overheden, bedrijven en andere organisaties uit de stad genereren en over de stad verzamelen zijn gemeenschappelijk bezit. Iedereen kan ze gebruiken. Iedereen kan er voordeel van hebben. Hier maken we gezamenlijk afspraken over.

5. Referenties

- Agarwal, S. (2021). *Responsible AI: Implementing Ethical and Unbiased Algorithms*. Cham, Switzerland: Springer.
- AI Ethics lab. (2022, 06 13). <https://aiethicslab.com/big-picture/>. Retrieved from aiethicslab.com: <https://aiethicslab.com/big-picture/>
- Baer, T. (2019). *Understand, Manage, and Prevent Algorithmic Bias*. New York: Apress.
- Barocas, H. e. (2022, 06 13). *Fairness and machine learning*. Retrieved from <https://fairmlbook.org>: <https://fairmlbook.org/pdf/fairmlbook.pdf>
- Bowles, C. (2018). *Future Ethics*. London: NowNext Press.
- Broussard, M. (2019). *Artificial Unintelligence*. Cambridge: MIT Press.
- Buitenweg, K. . (2021). *Datamacht en tegenkracht : hoe we de macht over onze gegevens kunnen terugkrijgen*. Amsterdam: De Bezige Bij.
- Crawford, K. (2021). *Atlas of AI*. New Haven and London: Yale University Press.
- Custers, B. (2019, 11). *behoorlijk-datagebruik-in-de-openbare-ruimte.pdf*. Retrieved from www.digitaleoverheid.nl: <https://www.digitaleoverheid.nl/wp-content/uploads/sites/8/2019/11/behoorlijk-datagebruik-in-de-openbare-ruimte.pdf>
- Dignum, V. (2019). *Responsible Artificial Intelligence, How to Develop and Use AI in a Responsible Way*. Springer. Retrieved from Virginia Dignum, Responsible Artificial Intelligence, How to Develop and Use AI in a Responsible Way, 2019, Springer.

- Fair data team. (2022, 05 21). *Metadata provider*. Retrieved from github.com/FAIRDataTeam/FAIRDataPoint-Spec#metadata-provider-api: <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec#metadata-provider-api>
- GO FAIR. (2021, 06 02). <https://www.go-fair.org/fair-principles/>. Retrieved from <https://www.go-fair.org>: <https://www.go-fair.org/fair-principles/>
- GoDataDriven. (2020). <https://godatadriven.com/topic/data-ai-survey-2019-20/>. Retrieved from <https://godatadriven.com>: <https://godatadriven.com/topic/data-ai-survey-2019-20/>
- Green Web Foundation. (2022, 06 13). www.thegreenwebfoundation.org. Retrieved from www.thegreenwebfoundation.org: www.thegreenwebfoundation.org
- Kantayya, S. (Director). (2020). *Coded Bias* [Motion Picture].
- Manovich, L. (2020). *Cultural analytics*. Cambridge: MIT Press.
- O'Neil, C. (2016). *Weapons of math destruction*. London: Penguin Books.
- Piersma, N. (2021, 12 16). *Anonimiseren van datasets*. Retrieved from www.hva.nl/subsites/nl/kc-fdmci/lectoraten/lectoraat-responsible-it/documenten/anonimiseren-van-datasets.html: <https://www.hva.nl/subsites/nl/kc-fdmci/lectoraten/lectoraat-responsible-it/documenten/anonimiseren-van-datasets.html?origin=fpYBevADQTCjoNaW6loniA>
- Schut, G. (2020, 03 13). *nieuwe-analyse-van-energieverbruik-datacenters-stelt-gerust*. Retrieved from <https://www.technischweekblad.nl>: <https://www.technischweekblad.nl/artikelen/tech-achtergrond/nieuwe-analyse-van-energieverbruik-datacenters-stelt-gerust>
- Shin, T. (2021, 06 02). *real-life-examples-of-discriminating-artificial-intelligence*. Retrieved from <https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070>: <https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070>
- Stanford Alumni Association. (2021, 06 11). <https://medium.com/stanford-magazine/carbon-and-the-cloud-d6f481b79dfe>. Retrieved from <https://medium.com/stanford-magazine>: <https://medium.com/stanford-magazine/carbon-and-the-cloud-d6f481b79dfe>
- Stikker, M. &. (2019). *Het internet is stuk maar we kunnen het repareren*. Amsterdam: de Geus.
- Tada. (2021, 06 02). tada.city/. Retrieved from <https://tada.city/>: <https://tada.city/>
- The digital society. (2021, 06 02). <https://www.thedigitalsociety.info/nl/over/dataprincipes/>. Retrieved from <https://www.thedigitalsociety.info>: <https://www.thedigitalsociety.info/nl/over/dataprincipes/>
- The digital society. (2022, 06 13). www.thedigitalsociety.info/about/data-principes/. Retrieved from [/www.thedigitalsociety.info](http://www.thedigitalsociety.info): <https://www.thedigitalsociety.info/about/data-principes/>
- Utrecht Data School. (2022, 06 13). dataschool.nl/deda. Retrieved from dataschool.nl: <https://dataschool.nl/deda>
- Vallor, S. (2016). *An Introduction to Data Ethics*.
- Vogel, I., & al, e. (2018). *How to work with data*. Amsterdam: HVA.

- Wilkinson, M. D., Michel Dumontier, I. J., & al, e. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Nature*, <https://www.nature.com/articles/sdata201618>.
- Zheng, C. W. (2014). <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02>. Retrieved from www.microsoft.com: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/UrbanComputing-zheng-tist2014.pdf>
- Zuboff, S., & Schwandt, K. (2019). *The age of surveillance capitalism*. London: Profile Books.

Over ons

Responsible IT, Hogeschool van Amsterdam

Responsible IT heeft als doel het ontwikkelen van digitale technologie die op een verantwoorde manier wordt ingezet voor verantwoorde doelen. De beweging om IT systemen te gebruiken en te ontwerpen op een mensgerichte manier, vanuit publieke waarden is een reactie op de bestaande praktijk van de (veelal) Amerikaanse Big Tech bedrijven en meestal Aziatische toepassingen op het gebied van overheidscontrole systemen. Data sensing, datamanagement en datagebruik is een van de onderzoeksterreinen van de responsible IT beweging met als doel om op een verantwoorde manier data te verwerven, te delen en gebruik te maken van data. Data, en de algoritmes die de data analyseren, worden alleen gebruikt als deze ook te vertrouwen zijn in de brede zin van het woord (Engels: reliable en trustworthy), wanneer mensen begrijpen wat de analyse betekent en wanneer er een meerwaarde is om op een data gedreven manier naar een probleem te kijken. Datagebruik "omdat het kan" leidt vaak tot onverantwoorde situaties, waarin de privacy van mensen wordt geschonden, waarin mensen worden benadeeld of zelfs gediscrimineerd en waarin de macht van data-eigenaren leidt tot misbruik. Responsible IT en het onderdeel verantwoord datamanagement is werken met data waarop vertrouwd kan worden en die bijdraagt aan een betere wereld.