

# White Paper Verantwoord Datagebruik door wetenschappelijke uitgeverijen

10 februari 2023

## Inleiding

Wetenschappelijke uitgeverijen zijn (ook) leveranciers van digitale producten waarin veel en veelsoortige data worden verwerkt. Er blijkt een maatschappelijke behoefte om inzicht te krijgen in de deze verwerkingen en welke eventuele risico's daarmee gepaard gaan. Dat blijkt onder andere uit het [rapport](#) van het Rathenau Instituut 'Naar hoogwaardig digitaal onderwijs'.

Zowel op Nederlands als Europees niveau neemt ook de politieke aandacht voor regelgeving omtrent het veilig en verantwoord gebruiken en verwerken van data een vlucht. Dit werkt ook door in regelgeving, zoals (aankomende) Europese wetgeving als de Data Act, Data Governance Act, Digital Markets Act en Digital Services Act. Onderdeel van de discussie is de vraag hoe bedrijven omgaan met data. De kritiek richt zich vaak op de "datahonger" van deze bedrijven, de afhankelijkheid die zij creëren, en dat gebruikersdata wordt gebruikt en verhandeld voor bijvoorbeeld marketing of AI, zonder dat de gebruiker dit weet.

MEVW hecht eraan dat de opslag en het gebruik van data op verantwoorde wijze gebeurt. Dit betekent concreet dat wetenschappelijke uitgeverijen transparant zijn over welke data zij opslaan en op welke manier zij de data verwerken en gebruiken. In dit stuk wordt het volgende uiteengezet:

- De belangrijkste types van data die uitgeverijen opslaan;
- Wat uitgeverijen doen met data;
- Waar het eigenaarschap en de verantwoordelijkheid ligt waar het gaat om het verwerken (zoals verkrijgen, opslaan, gebruiken en delen) van data;
- De positie van MEVW ten aanzien van verantwoord datagebruik.

Bedenk daarbij dat niet alle wetenschappelijke uitgeverijen alle onderstaande soorten data verwerken. Dit stuk schetst wel de kaders waarbinnen wetenschappelijke uitgeverijen opereren.

## 1. Data types

Wetenschappelijke uitgeverijen verwerken naast de artikelen zelf verschillende soorten data. Dit white paper heeft als doel om meer duidelijkheid te geven over hoe en waarom deze data types worden verwerkt. Het betreft de volgende data types:

- Onderzoeksdata zijn de gegevens die via wetenschappelijk onderzoek verkregen zijn. Bijvoorbeeld: bestanden met seismische informatie of ingevulde vragenlijsten.
- Publicatiedata zijn gegevens die betrekking hebben op wetenschappelijke artikelen. Bijvoorbeeld: lijsten van auteurs met hun affiliaties of de citaties in artikelen.
- Textmined data zijn de gegevens die verkregen zijn door wetenschappelijke artikelen automatisch te verwerken tot gestructureerde informatie. Bijvoorbeeld: lijsten van artikelen die verwijzen naar een bepaalde chemische stof.
- Gebruik(er)sdata zijn gegevens over het gebruik (soms de gebruiker) van wetenschappelijke informatie. Bijvoorbeeld: log files van artikelen die vanuit een instelling gedownload zijn.
- Persoonlijke data zijn alle gegevens die naar de gebruiker kunnen leiden of aan de gebruiker gekoppeld zijn. Bijvoorbeeld een telefoonnummer of e-mailadres.
- Geaggregeerde data zijn samengevoegde data die ontstaan door het weglaten van onderliggend details. Bijvoorbeeld: data over graanproductie van alle landen ter wereld kan worden geaggregeerd tot data over graanproductie van alle continenten.

## 2. Wat doen wetenschappelijke uitgeverijen met data?

Wereldwijd zijn er duizenden databases van allerlei kennis- en onderzoeksinstituten. En omdat voortschrijdend inzicht inherent is aan wetenschappelijk onderzoek, komen er dagelijks veel nieuwe wetenschappelijke onderzoeksresultaten (incl. de bijbehorende datasets) bij. Wetenschappelijke kennisinfrastructuren zijn dus omvangrijk en dynamisch.

Een van de primaire taken van wetenschappelijke uitgevers is het voor onderzoekers en onderzoeksinstellingen toegankelijk, vindbaar en bruikbaar maken van de enorme hoeveelheid wereldwijd beschikbare onderzoeksresultaten. Daarmee vervullen uitgevers een onmisbare schakel in het opbouwen, onderhouden en uitbreiden van wetenschappelijke kennisinfrastructuren.

### 2.1. Onderzoeksdata

Onderzoeksdata zijn de gegevens, metingen en observaties die ten grondslag liggen aan de theorie en of argumenten zoals gepresenteerd in een onderzoeksartikel. Deze worden door wetenschappers voornamelijk opgeslagen in dataopslagplaatsen of van hun kennisinstelling of in (nationale of internationale) discipline-specifieke dataopslagplaatsen. Uitgevers hebben geen invloed op of zeggenschap over het beheren van deze data en de opslagplaatsen, maar spelen wel een belangrijke rol in:

1. Het stimuleren van wetenschappers of faciliteren om hun onderzoeksdata ergens op te slaan;
2. Het digitaal linken van de onderzoeksdata aan de bijbehorende artikelen.

### 2.2. Publicatiedata

Publicatiedata wordt gegenereerd in het publicatieproces en speelt een belangrijke rol in het vindbaar maken van onderzoeksartikelen. Deze data is gestructureerd (bijvoorbeeld: *auteurs, titel, samenvatting, tijdschrift, citaties, jaar*) en wordt door uitgeverijen aan elkaar beschikbaar gemaakt om zo publicaties toegankelijk, vindbaar en bruikbaar te maken. Zo kan bijvoorbeeld geanalyseerd worden welke artikelen bepaalde andere artikelen citeren en welke wetenschappers in een bepaald gebied werken om zo inzichten te verstrekken zoals welke samenwerking de wetenschap vooruit zou kunnen helpen.

### 2.3. Textmined data

Data verkregen door textmining wordt steeds relevanter in wetenschappelijk onderzoek omdat het een doeltreffend en efficiënte manier is om relevante informatie uit grote hoeveelheden artikelen inzichtelijk te maken en bijvoorbeeld geïnformeerde beslissingen te nemen over onderzoeksrichtingen en samenwerkingen. Een concreet voorbeeld hiervan is het versneld identificeren van welke specifieke medicijnen mogelijk geschikt zijn voor hergebruik in de behandeling van een chronische of zeldzame ziekte.

### 2.4. Gebruikers- en gebruiksdata

Gebruikersdata is data over de gebruiker zelf en gebruiksdata is data over het gebruik van een applicatie; die laatste soort wordt soms verwerkt op het niveau van personen of instellingen. Beide soorten data kunnen door de uitgever worden verzameld om persoonlijke services te leveren (zoals het maken van aanbevelingen of voor marketingdoeleinden) of om de producten te kunnen verbeteren. Daarnaast is het verzamelen en verwerken van gebruik(er)sdata noodzakelijk om te waarborgen dat de website goed blijft functioneren. Zo kan de oorzaak van een probleem op de website liggen bij een specifiek gebruik of specifieke gebruiker. Soms zijn ze ook nodig voor het beveiligen van websites en voor het in rekening brengen van de informatie die via de website verkregen is. Bijvoorbeeld wanneer de uitgever onderzoek moet kunnen doen naar een gebruiker van de website door zich niet houdt aan de contractvoorwaarden. Voor zover persoonsgegevens worden verwerkt, gebeurt dat uiteraard in het kader van de AVG.

## 2.5. Persoonlijke data

Persoonlijke data zijn alle gegevens die naar de gebruiker kan leiden of aan de gebruiker gekoppeld is, zoals een telefoonnummer of e-mailadres. Dit kan bijvoorbeeld nodig zijn voor het aanmaken van een profiel of account op de website van de uitgeverij. Uitgeverijen gebruiken persoonlijke data alleen wanneer daar toestemming voor is gegeven en binnen het wettelijke kader van de Europese AVG. Welke persoonlijke data precies wordt opgeslagen en op welke manier dit gebruikt kan worden door de uitgeverij wordt uiteengezet in de privacyverklaring van de desbetreffende uitgeverij.

## 2.6. Geaggregeerde data

De genoemde persoonlijke data kan -mits daar toestemming voor is gegeven- ook worden gebruikt om data te aggregeren, waarbij ze wordt geanonimiseerd; gegevens van een individuele gebruiker zijn in geaggregeerde vorm dus niet te achterhalen. Afhankelijk van de afspraken die de uitgeverij maakt met haar klanten (zoals kennisinstellingen) wordt geaggregeerde data opgeslagen, ingezet en met hen gedeeld. Het aggregeren van data heeft een waardevolle maatschappelijke rol. Zo geven uitgeverijen inzichten in (en overzichten van) trends in onderzoek in individuele kennisinstellingen tot trends in wetenschap in de bredere samenleving. Enkele voorbeelden:

- Geaggregeerde data kan inzichtelijk maken hoe de gender- en diversiteitsverhoudingen in een onderzoeksveld liggen, om daarmee eventueel gewenste interventies te onderbouwen.
- Geaggregeerde data kan laten zien welke onderzoeksgebieden in opkomst zijn, om zo eventuele wijzigingen in toewijzingen van onderzoeksgelden te motiveren.
- Het aantal onderzoeken die werken aan de VN *Sustainable Development Goals* (SDGs) kan in kaart gebracht worden, en leiden tot bijstelling van onderzoeksrichtingen.

## Koppelen van verschillende types data

Naast het verwerken, verrijken, en aanbieden van data van bovenstaande types, zijn uitgeverijen ook actief in het aanbrengen van koppelingen tussen data van verschillende types, om zo de informatie beter vindbaar en bruikbaar te maken voor wetenschappers en om organisaties te voorzien van relevante inzichten. Voorbeelden hiervan zijn:

- Het koppelen van informatie, tussen het wetenschappelijke artikel de bijbehorende datasets en de wetenschappelijke artikelen onderling (bijvoorbeeld omdat het een de ander citeert). Hiervoor zijn brede standaarden ontwikkeld die alom gebruikt worden;
- Wetenschappelijk onderzoek en patenten worden gekoppeld aan auteurs en subsidiegevers, zodat de impact van subsidies inzichtelijker kan worden gemaakt;
- Het beschikbaar maken (in downloadbare format) van grote datasets, nadat ze gestructureerd en verrijkt zijn met unieke identifiers, wat verdere analyse en interoperabiliteit mogelijk maken.

## 3. Eigenaarschap en verantwoordelijkheid op data

Een belangrijk aandachtspunt ten aanzien van het verzamelen, de opslag, en de verwerking van data (van elke van de genoemde types) door wetenschappelijke uitgeverijen richt zich op persoonsgegevens van wetenschappers, in elke mogelijke rol (auteur, editor, reviewer, gebruiker, data creator, manager, etc.). Een essentiële vraag op dit gebied is of, en tot in hoeverre, een uitgeverij verwerkingsverantwoordelijke of gegevensverwerker is. De Europese AVG-regels zijn hierbij leidend en naleving van de wet in Nederland wordt getoetst door de Autoriteit Persoonsgegevens (AP). De definitie van de Europese Commissie is als volgt:

- De **verwerkingsverantwoordelijke** bepaalt de doeleinden waarvoor en de middelen waarmee persoonsgegevens worden verwerkt;

- De **gegevensverwerker** verwerkt persoonsgegevens uitsluitend namens de verwerkingsverantwoordelijke.

Binnen deze wettelijke kaders wordt de rol van de uitgeverij vastgelegd in een overeenkomst met haar afnemers. Zij maken concrete afspraken met individuen of (consortia van) kennisinstellingen, zoals universiteiten, hogescholen en bedrijven over de doelen en middelen waarmee persoonsgegevens worden verwerkt. In de overeenkomst wordt precies uiteengezet in welke rol de uitgeverij optreedt als gegevensverwerker of verwerkingsverantwoordelijke. Voor zowel individuen als organisaties geldt dat de uitgeverij -voordat zij hun gegevens delen- transparant is over waar de gegevens wel en niet voor gebruikt mogen worden.

#### 4. Positie MEVW

Volgens MEVW leveren wetenschappelijke uitgeverij met de dataverwerkingen op hun platforms een waardevolle en belangrijke maatschappelijke bijdrage, met name voor de (Nederlandse) onderwijs- en wetenschapsector. Bijvoorbeeld:

- Door de rol van wetenschappelijke uitgeverijen in de keten kunnen kennis en kunde – in de vorm van data – op grote schaal aan elkaar worden gekoppeld, gecentraliseerd, geïndexeerd en toegankelijk worden gemaakt om de wetenschap vooruit te helpen;
- Door middel van geaggregeerde data bieden wetenschappelijke uitgeverijen handvaten om impact / betrokkenheid te meten op/met actuele maatschappelijke ontwikkelingen. Zo krijgen funders bijvoorbeeld inzicht aan welke VN SDGs onderzoeksinstellingen zij bijdragen;
- Door het analyseren van gebruikersdata kunnen uitgeverijen tegemoet komen aan de diverse functionaliteitswensen van gebruikers, zoals het kunnen terugkijken van het eigen zoekgedrag, of het aanbevelen van ander onderzoek op basis van het eigen zoekgedrag;
- In een tijd van des- en misinformatie zijn uitgeverijen poortwachters. Door middel van extensieve validatieprocessen waarborgen zij de (academische) integriteit van de beschikbare kennis en is het te herleiden naar de oorspronkelijke auteur(s).

MEVW onderschrijft tegelijkertijd het belang van heldere (juridische) kaders voor de verwerking van data om zo maatschappelijke waarden als privacy, keuzevrijheid en onafhankelijkheid te waarborgen. In het bijzonder waar het gaat om:

- Transparantie in de 'wat' en 'hoe' uitgeverijen omgaan met het verzamelen, de opslag, en de verwerking van data en het behoud van deze transparantie te betrachten;
- Het waarborgen van de (individuele) privacy en keuzevrijheid van gebruikers;
- Het waarborgen van de keuzevrijheid van afnemers, zoals kennisinstellingen;
- Het verstrekken van informatie van trends in de wetenschap op die vlakken waar ze belangrijke maatschappelijke ontwikkelingen raken;
- Het verder toegankelijk maken van kennis en kunde, zowel binnen als buiten de academische gemeenschap.

In de optiek van MEVW vraagt het borgen van maatschappelijke waarden bij de dataverwerking door wetenschappelijke uitgeverijen om publieke en private samenwerkingen en om het gesprek met publieke stakeholders aan te gaan over welke waarborgen zij wenselijk of noodzakelijk achten waar het gaat over het dataopslag- en gebruik door uitgeverijen. Daardoor:

- Kunnen duurzame oplossingen tot stand komen die passen bij de desbetreffende stakeholders;
- Wordt de 'wat', 'hoe' en 'waarom' van uitgeverijen in hun dataopslag- en gebruik inzichtelijker en dus transparanter gemaakt;
- Kan verdere innovatie in de (Nederlandse) kennisinfrastructuren worden gewaarborgd, ten gunste van de (Nederlandse) wetenschaps- en onderzoeksgemeenschap;
- Wederzijds vertrouwen tussen publieke en private partijen worden gecultiveerd.



Samen werken  
aan kennis en  
ontwikkeling

Op verschillende manieren ondernemen bij MEVW aangesloten wetenschappelijke uitgeverijen hier reeds proactief actie op, bijvoorbeeld:

- Het samenwerkingsverband COUNTER bestaat uit uitgeverijen en universiteiten, waarbinnen precies wordt aangegeven wat wetenschappelijke uitgeverijen delen met de instituten over het geaggregeerde downloadgedrag van gebruikers.
- In het kader van Open Science streven uitgeverijen ernaar om transparantere en inclusievere platformen te realiseren. Een voorbeeld hiervan is [EPDOS](#), waar de Nederlandse onderzoeksgemeenschap (UNL, NFU, VH, NWO en KNAW) samen met Elsevier een overeenkomst hebben gesloten voor de ontwikkeling nieuwe Open Science diensten.