

# Open Data richtlijn voor CREATE data

De onderzoeken die binnen het CREATE-programma worden uitgevoerd, genereren doorgaans waardevolle gegevensverzamelingen. Bij het faciliteren van onderzoek binnen het CREATE-programma geldt als vereiste dat deze gegevens als Open Data beschikbaar worden gesteld. Van deze vereiste kan alleen worden afgeweken indien de *Uitvoeringswet Algemene verordening gegevensbescherming* (AVG) of auteursrechtelijke aanspraken van derden openbaarmaking van zulke data verhindert.

## **Wat is Open Data?**

Onder Open Data verstaan we gegevensverzamelingen die kosteloos en vrij toegankelijk in een machineleesbaar formaat openbaar worden gemaakt.

## **Waarom Open Data?**

Vanuit de samenleving klinkt steeds luider de roep dat de resultaten van publiek gefinancierd onderzoek ook publiekelijk beschikbaar moeten worden gemaakt (cf. Open Access). Dat geldt niet alleen voor artikelen en boeken maar ook voor de onderzoeksgegevens die hieraan ten grondslag liggen. Daarnaast draagt Open Data bij aan een gezonder wetensklimaat doordat het de controleerbaarheid en de repliceerbaarheid van onderzoek bevordert. Verder stimuleert Open Data het hergebruik van onderzoeksgegevens. Binnen het CREATE-programma wordt al enthousiast en op grote schaal gebruik gemaakt van Open Data. En met succes!

Een mooi voorbeeld van het gebruik van Open Data is de ONSTAGE database met daarin alle theateropvoeringen in de Amsterdamse Schouwburg. Voor het samenstellen van deze database hadden we een referentielijst nodig van toneelschrijvers. Door uit de Short Title Catalogue Netherlands alle schrijvers van toneelwerken op te vragen, hadden we deze lijst in twee minuten samengesteld. Zonder Open Data zouden we waarschijnlijk enkele weken kwijt zijn geweest aan het opstellen van zo'n lijst.

Met deze richtlijn willen we het gebruik en de beschikbaarheid van Open Data binnen de geesteswetenschappen dan ook graag een steuntje in de rug geven.

## **Hoe maak je Open Data?**

Het open maken van data is deels een technische kwestie en deels een juridische. Voor de technische kant zijn er tal van infrastructurele voorzieningen, zowel binnen de UvA als binnen landelijk opererende platformen als DANS en CLARIAH. Wat de juridische kant betreft, kan data al open zijn doordat deze zich in het publieke domein bevindt. Als dat niet het geval is, kan data open gemaakt worden middels een licentie waarin aangegeven wordt dat de gegevens door derden vrijelijk gebruikt mogen worden. Licenties kunnen evenwel alleen verstrekt worden door de eigenaar van de data.

## Intellectueel eigendom van onderzoeksdata

Om te bepalen of data open is of open gemaakt kan worden is het van belang om vast te stellen welke eigendomsrechten er op de gegevens rusten. De intellectuele eigendom van onderzoeksdata vloeit voort uit het auteursrecht en/of het databankenrecht. Afhankelijk van de inhoud en de aard van je gegevensverzameling, kan de intellectuele eigendom daardoor verschillende vormen aannemen. Bij geesteswetenschappelijk onderzoek hebben we bovendien vaak te maken met gelaagde dataverzamelingen, waarbij de eigendom per laag kan verschillen.

De intellectuele eigendom die voortvloeit uit het auteursrecht is het bekendst. Deze eigendom wordt door de maker van een werk automatisch verkregen. Daarbij geldt wel dat artikel 7 van de auteurswet bepaalt dat het auteursrecht van een werk dat in loondienst wordt vervaardigd, toekomt aan de werkgever. Als een onderzoeker in dienst van de UvA een gegevensverzameling genereert in het kader van zijn/haar dienstverband, verkrijgt de UvA dus het auteursrecht. De jurisprudentie regelt dat een werk om voor auteursrechtelijke bescherming in aanmerking te komen, het product moet zijn van een creatieve en/of intellectuele inspanning van de maker.

Het databankenrecht regelt de eigendom en exploitatierechten van databanken. De juridische definitie van een databank wijkt enigszins af van de technische. In juridische zin is een databank een methodisch of systematisch geordende verzameling van zelfstandige elementen. Ook niet-digitale gegevensbestanden als kaartenbakken, rolodexen en telefoonboeken zijn in juridische zin databanken, terwijl gegevensverzamelingen in spreadsheets (wellicht) soms niet aan de vereisten voldoen. Voorwaarde voor het verkrijgen van het databankenrecht is dat er een substantiële investering (in geld of arbeid) is gedaan. Het is ter bescherming van die investeringen dat databanken door het databankenrecht beschermd worden. Die bescherming geldt niet voor de zelfstandige elementen (die wel weer onder het auteursrecht kunnen vallen, bijvoorbeeld bij een encyclopedisch werk) maar voor de databank als geheel. Het databankenrecht geldt tot vijftien jaar na de laatste substantiële investering in de databank.

Geen intellectuele eigendom kan worden geclaimd op:

- Gegevens die op zichzelf staan. Bijvoorbeeld op het enkele gegeven dat iemand op een bepaalde datum op een bepaald adres woonde, kan geen intellectuele eigendom geclaimd worden. Indien dit gegeven echter voortvloeit uit een beredenering, dan valt die beredenering – mits vastgelegd – wel onder het auteursrecht.
- Werken waarvan de maker langer dan zeventig jaar overleden is. Dat geldt ook voor reproducties van zulke werken. Op een transcriptie van een werk waarop het auteursrecht verlopen is, berust dus geen auteursrecht. Dat geldt ook voor fotografische reproducties, met inbegrip van infrarood fotografie, digitale scans etc. In deze gevallen geldt dat het werk zich in het publieke domein bevindt.
- Wetten en regelgeving van overheden. Ook in deze gevallen geldt dat het werk zich in het publieke domein bevindt.
- Werken die dusdanig triviaal zijn dat er geen creatieve en/of intellectuele inspanning aan ten grondslag ligt.

## Licenties

Een licentie kun je alleen verstrekken op gegevensverzamelingen waarop je rechten hebt. Uiteraard kun je een specifieke licentie maken voor jouw specifieke dataverzameling, maar dat vereist veel juridische kennis. Het is daarom bijna altijd aan te bevelen om gebruik te maken van generieke licenties zoals de Open Database License en de Creative Commons licenties.

De **Open Database License (ODBL)** is een licentie die specifiek voor databases is opgesteld. Deze licentie heeft alleen betrekking op rechten die voortvloeien uit het databankenrecht en het auteursrecht op de databasestructuur. Eventuele rechten op afzonderlijke elementen in de database worden echter niet door deze licentie gedekt. Deze licentie valt dus alleen te overwegen als er geen rechten berusten op de afzonderlijke elementen van je dataverzameling, of als deze bij anderen berusten. In het eerste geval valt hierbij te denken aan database met archivalia die zich in het publieke domein bevinden. In het tweede geval kun je denken aan de respons op een enquête met open vragen. De rechten op de afzonderlijke antwoorden liggen dan immers bij de respondenten. Een andere reden om voor deze licentie te kiezen is als je voor je gegevensverzameling op grote schaal gebruik maakt van OpenStreetMap gegevens, aangezien deze database onder deze specifieke licentie valt.

De Open Database License kent twee varianten. Een waarbij alle rechten worden vrijgegeven:

<https://opendatacommons.org/licenses/odbl/1-0/>

En een variant waarbij de rechten worden vrijgegeven onder voorwaarde van naamsvermelding:

<http://opendatacommons.org/licenses/by/1.0/>

De Open Database License wordt behalve door OpenStreetMap onder andere gebruikt door de RKD (Nederlands Instituut voor Kunstgeschiedenis) en het public knowledge project Open Food Facts:

<https://world.openfoodfacts.org/>

Bij Open Food Facts wordt in de licentievoorwaarden een expliciet onderscheid gemaakt tussen de licentie op de database en de databasestructuur (ODBL licentie) en de licentie op de afzonderlijke elementen van de database, die onder een Creative Commons licentie worden vrijgegeven.

De **Creative Commons (CC)** licenties zijn destijds in het leven geroepen om makers de gelegenheid te geven om af te zien van rechten die voortvloeien uit het auteursrecht. Tot en met versie 3 van Creative Commons licenties zijn de rechten voortvloeiend uit het databankenrecht maar ten dele gedekt. Vanaf versie 4 kunnen deze licenties voor zowel teksten, media als data gebruikt worden. Het is daarom belangrijk om voor data versie 4 (of hoger) van de licentie te gebruiken.

De Creative Commons licenties bieden de mogelijkheid om voorwaarden te stellen aan het hergebruik van data. Dat kan bijvoorbeeld naamsvermelding (BY) zijn, de verplichting om de data ongewijzigd te laten (Non Derivative, ND), maar ook de voorwaarde dat afgeleide producten onder dezelfde condities moeten worden vrijgegeven (Share Alike, SA) of de voorwaarde dat afgeleide producten niet commercieel mogen worden uitgebaat (Non Commercial, NC). Men kan er ook voor kiezen om data vrij te geven onder een Creative Commons Zero (CC0) licentie vrij te geven, waarmee de data gevrijwaard wordt van alle vormen van auteursrecht. Een licentiekiezer helpt je bij het bepalen van de mogelijkheden:

<https://creativecommons.org/choose/?lang=nl>

Naast licenties voor data en werken waarop rechten berusten, biedt Creative Commons ook de mogelijkheid om bij her distributie van werken aan te geven dat deze zich in het publieke domein bevinden, middels een Public Domain Mark:

<https://creativecommons.org/publicdomain/mark/1.0/>

Een Public Domein Mark is geen licentie, maar een verklaring van de her distributeur dat het werk zich in het publieke domein bevindt.

Indien je samen met je onderzoeksdata werken her distribueert die zich in het publieke domein bevinden, dan kan het nodig zijn dat je de rechten op deze werken los van het gelicentieerde deel van je verzameling op deze wijze markeert. Stel je maakt een gegevensverzameling met door jou geannoteerde teksten van Multatuli. Dan bevinden de originele teksten zich in het publieke domein, maar de rechten op de annotaties liggen dan bij jou. In zo'n geval is het aanbevelenswaardig om de originele teksten met een Public Domain Mark te her distribueren, terwijl je de annotaties onder een Creative Commons licentie vrijgeeft.

Het is overigens altijd toegestaan als eigenaar van de data, om een uitzondering op bovenstaande verplichtingen te maken voor een specifieke partij.

### **Beperkingen, ja of nee?**

Een belangrijk argument tegen het opnemen van aanvullende beperkingen in je licentie is dat je bij het samenvoegen van datasets (bijvoorbeeld als Linked Open Data) te maken krijgt met een stapel licenties die onderling conflicterend kunnen zijn, waardoor datasets *de iure* niet meer interoperabel zijn. Binnen de Open Data-beweging is daarom het credo om zo min mogelijk beperkingen in je licentie op te nemen. Als vuistregel geldt:

- Voor gegevensverzamelingen waarop alleen rechten berusten die uit het databankenrecht voortvloeien: geen beperkingen.
- Voor gegevensverzamelingen waarop ook rechten berusten die uit het auteursrecht voortvloeien: met naamsvermelding.

Voor de argumentatie hieromtrent zie:

<https://www.openaire.eu/research-data-how-to-license/>

en:

[http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf)

In haar diesrede van 2019 heeft rector magnificus Karen Maex overigens enkele kanttekeningen geplaatst bij het vrijgeven van data zonder aanvullende beperkingen. Zo waarschuwt ze voor het scenario dat commerciële partijen middels 'verrijkte' aggregaten van open data, onderzoekers uiteindelijk voor hun eigen open data laten betalen:

<https://www.uva.nl/binaries/content/assets/uva/nl/nieuws-en-agenda/openingstoespraak-karen-maex-dies-2019.pdf>

Wanneer het risico op dergelijke piraterij hoog is, kan overwogen worden om bij wijze van uitzondering een Share Alike clause in de licentie op te nemen.



Dit werk valt onder een Creative Commons Naamsvermelding 4.0 Internationaal-licentie.