

Leidraad bij Datacharter voor Citizen Science

v1.0

Een basisset van principes
ter ondersteuning van open en
interoperabele citizen-sciencedata



Bijdragers

Naam	Organisatie	Betrokken in project	Rol
Mieke Sterken	<i>Adviseur Citizen Science, Scivil & KULeuven</i>	BRITEC, MoMster, CS-Roadmap voor Lokale besturen	Auteur
Ruben Capiou	<i>Business analyst, Digitaal Vlaanderen</i>	Vlaamse basisregisters, OSLO, INSPIRE, digitaliseringstrajecten	Auteur, voorzitter WG
Elien Dewitte	<i>Science officer, VLIZ</i>	SeaWatch, Grote Schelpenteldag	Acknowledged contributor, deelnemer WG
Sien Benoit	<i>PhD student, UGent</i>	Fietsbarometer	Acknowledged contributor, deelnemer WG
Tim Adriaens	<i>Science officer, INBO</i>	Vespa-Watch	Contributor, deelnemer WG
Christophe Stroobants	<i>Teammanager, VMM</i>	CurieuzeNeuzen, Project Zuivere Lucht	Contributor, deelnemer WG
Kristijn Swinnen	<i>Senior researcher, Natuurpunt</i>	waarnemingen.be, Mijn Tuinlab	Contributor, deelnemer WG
Maarten Trekels	<i>Data scientist, Plantentuin Meise</i>	Groene Pioniers	Contributor, deelnemer WG
Nils Walravens	<i>Senior researcher, IMEC-SMIT-VUB</i>	Open Data Charter 'Smart Flanders'	Contributor, deelnemer WG
Jef Van Laer	<i>Adviseur Citizen Science, Scivil</i>	Ondersteuning alle EW-citizen-scienceprojecten	Contributor, deelnemer WG
Annelies Duerinckx	<i>Coördinator Scivil, Scivil</i>	CS-Roadmap voor Lokale besturen	Contributor, deelnemer WG
Mathias De Schrijver	<i>Adviseur Informatiebeleid, Digitaal Vlaanderen</i>	Vlaams Open Data Portaal, Bestuursdecreet	Reviewer
Tom Callens	<i>Adviseur Informatiebeleid, Digitaal Vlaanderen</i>	Vlaamse Geografische Data-Infrastructuur, INSPIRE	Reviewer
Geraldine Nolf	<i>Product owner, Digitaal Vlaanderen</i>	Metadata Vlaanderen, DataPublicatie-proces	Reviewer
Delfien Vandyck	<i>Hoofddocent, UGent</i>	Fietsbarometer	Reviewer
Ziggy Vanlshout	<i>Programmamanager, Digitaal Vlaanderen</i>	Authentieke gegevensbronnen, Vlaamse basisregisters, OSLO	Acknowledged
Pieter Colpaert	<i>Postdoctoraal onderzoeker, UGent-IDLab-imec</i>	Publieke web API's	Acknowledged
Philippe Michiels	<i>Technical & Data Architect, imec City of Things</i>	Vlaamse Open City Architectuur (VLOCA)	Acknowledged
Werkgroep Datamanagement	<i>Stuurorgaan, Vlaams Informatie- en ICT-beleid</i>	nvt	Acknowledged

Dit document, het 'Datacharter voor Citizen Science' opgesteld door de werkgroep datamanagement van Scivil (<https://www.scivil.be/>), wordt uitgegeven onder een **CC BY- 4.0**-licentie. Bezoek <https://creativecommons.org/licenses/by/4.0/> om de licentietekst erop na te lezen.

Lay-out: Femke Vanhellemont ([lafem.](#))

Copywriting: Maarten Corten ([Slagpen](#))

Voorgescreven wijze van naamsvermelding (*attribution*):

Sterken M., Capiou R., Leden van de Scivil-werkgroep datamanagement. 2021. Datacharter voor Citizen Science: Een basisset van principes ter ondersteuning van open en interoperabele citizen-sciencedata. Scivil, 2021.

Legende

	Thema
	Principe
	Voor toepassing, tips & tricks, tools ...
	Praktijkvoorbeeld
Afkortingen- of begrippenlijst	Term toegelicht in afkortingen- of begrippenlijst
<i>Principe</i>	Kruisverwijzing naar principe of thema

Bijdragers	2
Licentie	3
Legende	3
Inleiding	5
Nota: het bekomen van '5-sterren open data'	7

I. OPEN ATTITUDE

1. Streef ernaar je data open op het web te publiceren, of geef een duidelijke en gegronde reden wanneer dit niet kan	10
2. Publiceer je data onder een open licentie die je kiest uit een beknopte, aanbevolen lijst	13
3. Publiceer je onderzoeksresultaten en -bevindingen waar mogelijk in 'Gold Open Access Journals'	17
4. Publiceer, waar mogelijk, ook de software die je ontwikkelt onder open licenties	20
5. Ga actief op zoek naar bestaande open data	23
6. Vraag raad aan ondersteunende diensten	25

II. PRIVACY & ETHIEK

7. Besteed actieve aandacht aan privacy en zorg voor kennisdeling	28
8. Kom met je project tegemoet aan de GDPR-richtlijnen	30
9. Communiceer helder rond intellectuele eigendommen en copyrights	34
10. Weeg je datakwaliteit en -kwantiteit voorzichtig af tegenover ethiek, haalbaarheid en projectdoelen	37
11. Let op voor ethisch ongewenste neveneffecten van het vrijgeven van je data	39
12. Neem een respectvolle, gelijkwaardige houding aan tegenover burgerwetenschappers	41

III. DATAHYGIËNE

13. Stel een datamanagementplan (DMP) op	44
14. Hou rekening met het cumulatiepotentieel van data: 'Behandel een kleine dataset zoals een grote'	46
15. Draag extra zorg voor de kwaliteit van data	48
16. Stel een conceptueel datamodel op voor jouw project	51

IV. DATASTANDAARDEN EN -FORMATEN

17. Bouw verder op bestaande en voor jouw project relevante datastandaarden	55
18. Maak gebruik van machineleesbare, open formaten	58
19. Ken globaal unieke en persistente identificatoren toe aan je data	60
20. Incorporeer je projectdata in het netwerk van Linked Open Data	63

V. METADATA

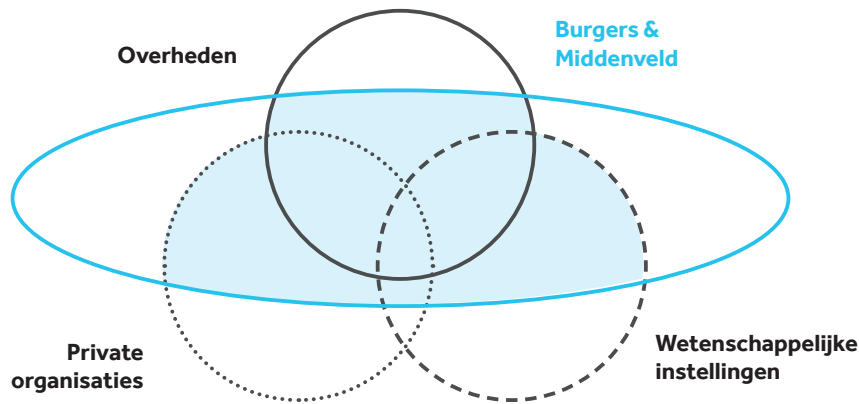
21. Voorzie je data van zo rijk en accuraat mogelijke metadata	66
22. Denk eraan de metadata zo dicht mogelijk bij de bron te capteren	69
23. Ken aan je metadataset een globaal unieke en persistente identicator toe	71
24. Gebruik de juiste standaard voor je metadata	73
25. Maak je metadata, en dus je datasets en onderzoeksresultaten, vindbaar door ze te registreren op een doorzoekbaar portaal	76
26. Maak je citizen-scienceproject vindbaar voor collega's en vrijwilligers	78
Referenties	80
Afkortingenlijst	81
Begrippenlijst	82

Inleiding

Met het Datacharter voor Citizen Science biedt Scivil, het Vlaamse kenniscentrum citizen science (CS), een leidraad aan om bij een citizen-scienceproject tot goed gedocumenteerde data te komen.

Het charter is er voor iedereen die een citizen-scienceproject wil opstarten of eraan wil deelnemen, dus zowel projectinitiatoren als deelnemers (citizen scientists) uit alle sectoren van de zogenaamde 'quadruple helix' (burgers, overheden, wetenschappelijke instellingen, private organisaties).

Een citizen-scienceproject is daarbij elk project dat kennis of inzichten probeert te vergaren op een wetenschappelijk verantwoorde manier, samen met of door burgers.



Figuur 1
De maatschappelijke actoren die betrokken kunnen zijn bij citizen science. Burgers spelen daarin een cruciale rol.

Zowel in de academische als overheidscontext zijn er wereldwijd initiatieven om het algemene dataverkeer gestroomlijnder, efficiënter en veiliger te maken. De internationale FAIR-principes¹ voor wetenschappelijke data (Wilkinson et al. 2016) in de academische wereld. Op Vlaams niveau zijn er de Open Standaarden voor Linkende Organisaties (OSLO) (Buyle et al. 2016) en het Open Data Charter² van Smart Flanders; zij moeten ervoor zorgen dat organisaties hun data meer open en linked maken, conform de Europese richtlijnen (bijvoorbeeld de INSPIRE-richtlijnen voor geografische data, en de Open Data Richtlijn van de Europese Commissie).

Burgers of initiatiefnemers van burgerwetenschapsprojecten weten mogelijk niet welke richtlijnen te volgen, of weten überhaupt niet dat er reglementen bestaan. Immers, zowel academici als overheidsinstanties, middenveldorganisaties, private bedrijven of burgers³ kunnen een citizen-scienceproject opstarten, al dan

niet in partnerschap en al dan niet gesponsord door een private partner, academische partner of overheidsinstantie. Naast het kluwen aan bestaande adviezen en regelgeving door verschillende maatschappelijke regulatoren, komen er bij citizen science bovendien andere uitdagingen kijken die een invloed kunnen hebben op de omgang met data. Onder andere het betrokken houden van burgers, het beschermen van hun privacy, en de vele mogelijke onderzoeksmethoden en datasettypes bemoeilijken het formuleren van een set universele regels voor de omgang met data.

Met dit datacharter voor citizen science wil Scivil een overzicht bieden van de belangrijkste datagerelateerde aanbevelingen die je ook in een citizen-scienceproject best nastreeft. We vertrekken hierbij vanuit de FAIR-dataprincipes, het Open Data Charter (Smart Flanders) en de 10 principes van de European Citizen Science Association (ECSA). Het charter poogt overzicht en reliëf te brengen in alle aspecten die met 'data in citizen science' te maken hebben.

Een basisprincipe hierbij is het streven naar het 'zo open en interoperabel mogelijk' publiceren van gegevens. Meer dan waar ook kan de meerwaarde van citizen science exponentieel toenemen door de principes in dit charter systematisch in acht te nemen. Via open en duidelijk gedefinieerde data kunnen zowel grote als kleine citizen-scienceprojecten op termijn leiden tot belangrijke wetenschappelijke inzichten, doordat hun resultaten kunnen gebruikt en gecombineerd worden met datasets van andere citizen-scienceprojecten, overheden, onderzoeksinstituten of zelfs de private sector.

"Data management is fundamental to the success and long-term impact of a citizen science (CS) project, yet it is often not considered until the end of the project, by which time decisions made at the beginning of the project can limit the choices for data reuse."⁴

¹ Zie <https://www.go-fair.org/fair-principles/>

² <https://smart.flanders.be/open-data-charter/>

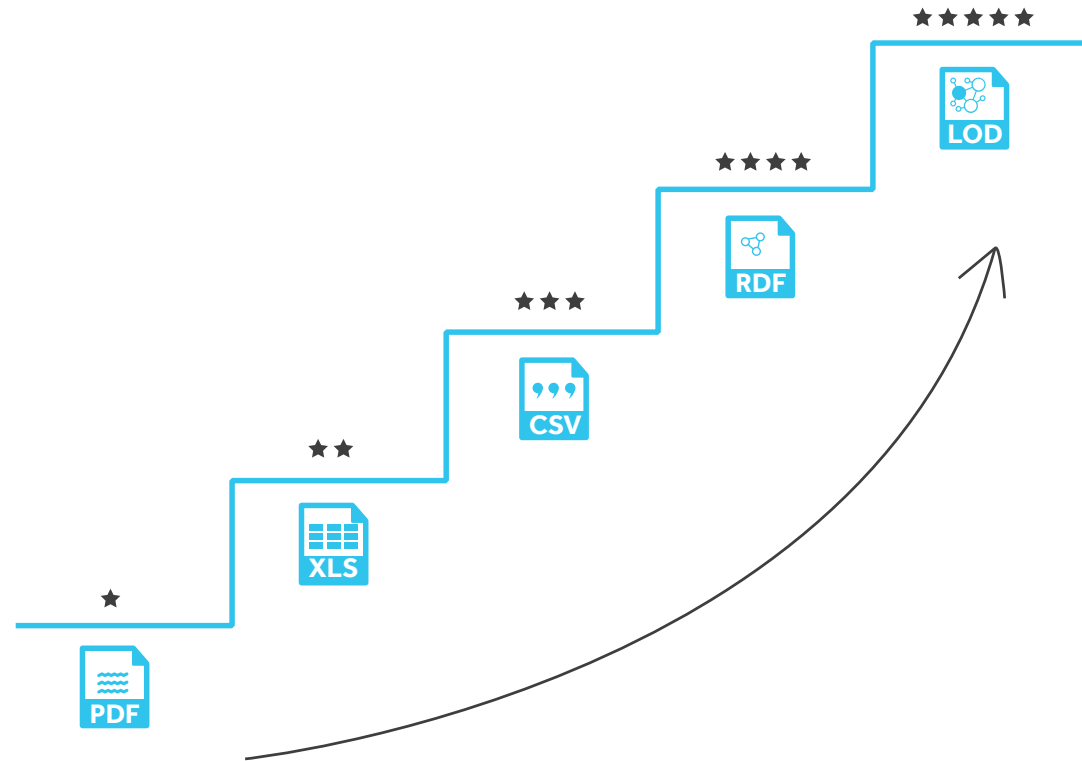
³ Wanneer burgers in de plaats van beleidsmakers initiatieven ontwikkelen, spreekt men over 'grassroots'-initiatieven.

⁴ Bron: <https://alien-csi.eu/working-group-3-wg3/> Vrij vertaald: "Datamanagement is cruciaal voor het succes en de langetermijnimpact van een burgerwetenschapsproject, en toch wordt hier vaak pas aan het einde van het project over nagedacht. Op dat moment kunnen de mogelijkheden voor datahergebruik al ingeperkt zijn door de keuzes die aan het begin van het project gemaakt werden."

Nota

het bekomen van '5-sterren open data'

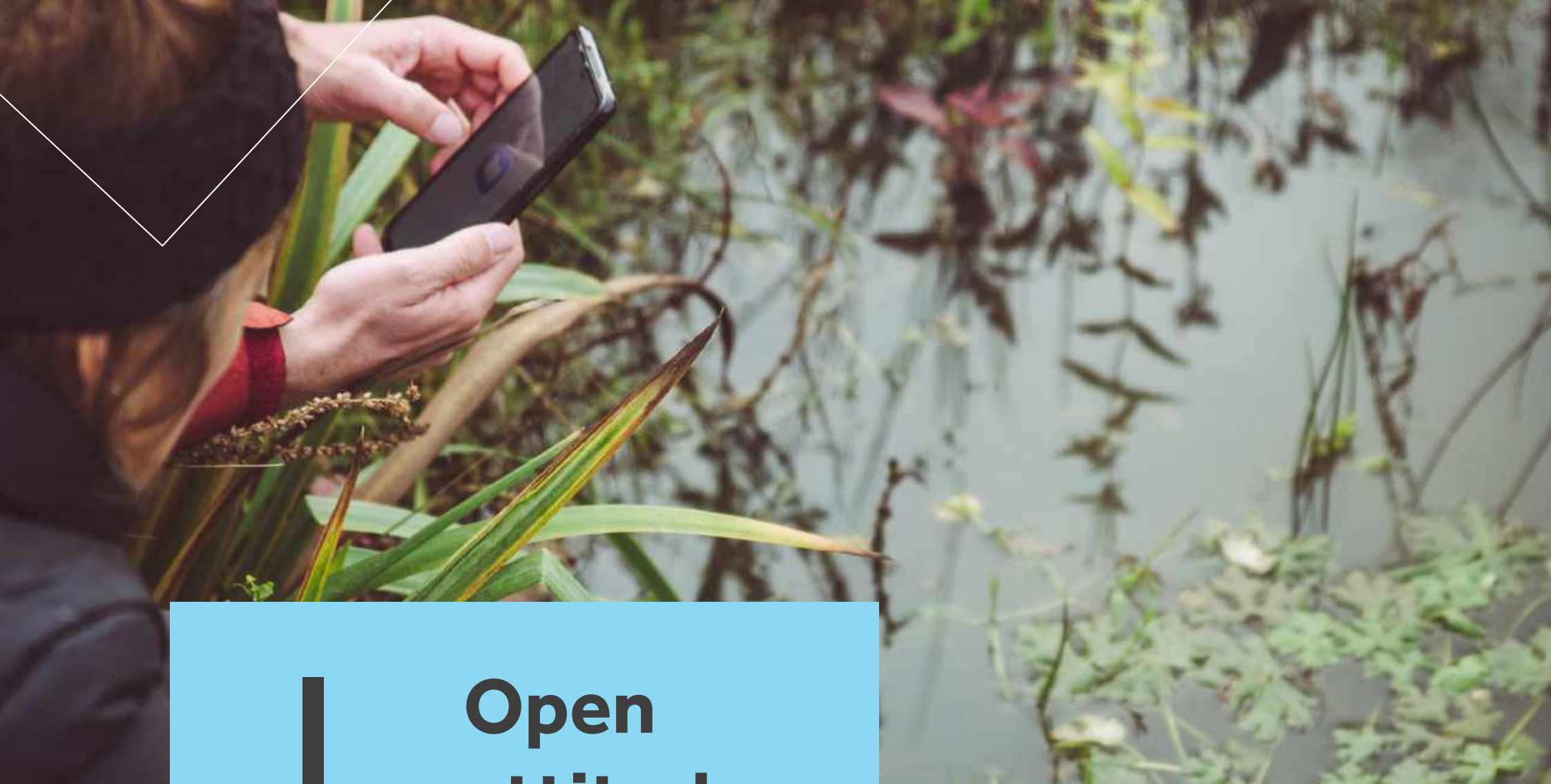
Wereldwijd streeft men steeds meer naar het publiceren van '5-sterren open data', ook gekend onder de term 'Linked Open Data' (LOD). Hierbij structureer en definieer je data zodanig dat je ze eenvoudig digitaal met elkaar kunt linken (binnen de grenzen van privacy- en veiligheidsreguleringen) en waarbij data-uitwisseling eenvoudiger wordt. Sir Tim Berners-Lee, de uitvinder van het wereldwijde web, vertaalde deze zogenaamde **interoperabiliteit** naar een getrapte schaal (zie Figuur 2). Hoe hoger op de trap, hoe makkelijker anderen jouw projectdata kunnen hergebruiken.



Figuur 2
'5 Star Open Data'.

Als je alle principes uit het Datacharter voor Citizen Science naleeft, bekom je 5-sterren open data. Onderstaande tabel illustreert aan welk principe je minimaal moet voldoen om een extra ster te verdienen, en wat dit concreet inhoudt of wat dit betekent voor de gebruikers van de betreffende projectdata.

Sterren	Bereikt na ...	Betekenis
★ aangeboden als open data	<u>principe 2</u>	Je begrijpt dat je data niet mag opsluiten in silo's en dat data pas echt renderen als anderen ze kunnen raadplegen en hergebruiken. Daarom kies je ervoor om jouw projectdata onder een open licentie te delen met de buitenwereld. Mogelijks bied je nog gegevens in een suboptimale, niet-machineleesbare vorm aan (bv. een scan van een tabel met resultaten of een PDF met een oplijsting van meetpunten): daarom verdien je voorlopig (maar) één ster.
★★ gestructureerd en machineleesbaar	<u>principe 18</u>	Je biedt jouw projectdata niet alleen onder een open licentie aan, maar ook op een gestructureerde (en dus machineleesbare) manier . Tekst plaats je in een Word-bestand, cijfergegevens in een Excel en je spiegelreflexcamera leverde hoogkwalitatieve RAW-beelden op. Daarmee zet je de stap naar '2-sterren open data'. Gebruikers kunnen nu de data gaan visualiseren of kunnen er berekeningen mee maken. Maar hebben ze daar de juiste software voor?
★★★ opgeslagen in een open dataformaat	<u>principe 18</u>	Doordat je kiest voor een open dataformaat , hoeven gebruikers niet langer specifieke software te gebruiken of aan te schaffen om met jouw data aan de slag te gaan. Geïnteresseerden kunnen jouw tekst- en spreadsheetdocumenten nu probleemloos en zonder gegevensverlies openen in een (gratis) softwarepakket naar keuze. Je foto's geven niet langer de foutmelding 'unknown file format', wanneer een collega ze opent op zijn computer. Dit krypt jouw data op tot zeer herbruikbare '3-sterren open data'. Stel nu dat een biologiestudent eigen onderzoeksgegevens aan observatie '100' uit jouw dataset wou koppelen (waarneming van een wilde zwaan in de Blaarmeersen op 2021-01-10T15:23:22). Hij verwarde deze observatie echter met observatie '100' uit een andere, Europese dataset (waarneming van een vale gier in de Pyreneeën op 2017-07-05T06:10:05) en trok bijgevolg de verkeerde conclusies. Hoe kunnen we dit in de toekomst voorkomen?
★★★★ uniek identificeerbaar	<u>principe 19</u>	Met jouw projectdata neem je nu de kop van het peloton. Voortaan kan iedereen op ondubbelzinnige wijze linken naar een individueel datapunt uit jouw dataset. Verwarring tussen https://waarnemingen.be/id/observatie/100 en https://nature-observatory.org/id/observation/100 is niet meer mogelijk dankzij de globaal unieke, persistente identificatoren . Ook de betekenis van de verzamelde data kan, zelfs volautomatisch, door andere systemen achterhaald worden via de URI's van de dataconcepten. Je dataset kan nu in het netwerk van Linked Open Data (LOD) opgenomen worden.
★★★★★ gelinkt met externe bronnen	<u>principe 20</u>	Jouw projectdata zijn nu open, gestructureerd, leesbaar door verschillende applicaties, uniek geïdentificeerd en gelinkt met externe bronnen . Gebruikers kunnen vanuit jouw verrijkte dataset andere LOD-bronnen verkennen en krijgen dus meer context bij jouw gegevens. Het is mogelijk om via querytalen als SPARQL intelligente bevragingen uit te voeren op jouw dataset in combinatie met andere bronnen. Jouw projectdata zijn nu volwaardige '5-sterren open data'!



**| ■ Open
attitude**

1. Streef ernaar je data open op het web te publiceren, of geef een duidelijke en gegronde reden wanneer dit niet kan

'Open' betekent dat eender wie jouw data vrij kan consulteren, gebruiken, aanpassen en delen. Er gelden hoogstens voorwaarden om de herkomst en openheid van je data te vrijwaren. 'Op het web' impliceert dat je de data via een webbrowser kunt opvragen.

Praktisch betekent dit dat je gegevens onder een [open licentie \(zie principe 2\)](#) vrijgeeft. Als dit niet mogelijk is, specificeer je uitdrukkelijk de reden voor het niet openstellen van (sommige) data. Tracht in dergelijke gevallen toch om minstens een deel van de gegevens (bv. [metadata of dus informatie over de data](#)) open te stellen. Het uitgangspunt is dus 'zo open als mogelijk, zo gesloten als nodig'.

Merk op: 'open data' betekent niet noodzakelijk dat je de gegevens gratis of meteen online ter beschikking stelt. Behalve voor onderzoeksgegevens en hoogwaardige datasets die onder de [Europese Open-Datarichtlijn](#)⁶ vallen (cfr. art. 10 en 14 van de [PSI-richtlijn](#)), kan je in principe marginale kosten aanrekenen voor vermenig-vul-di-ging, verstrekking en verspreiding. Bovendien kan het noodzakelijk zijn data expliciet aan te vragen. Deze praktijken worden echter afgeraden, omdat ze een extra drempel opwerpen voor de herbruikbaarheid van data.

⁶ Ook gekend als Public Sector Information richtlijn of PSI-richtlijn: <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN>. Artikel 10 van deze richtlijn bepaalt onder welke voorwaarden onderzoeksgegevens voor hergebruik beschikbaar moeten zijn, artikel 14 van de richtlijn bevat de regelingen voor publicatie en hergebruik van hoogwaardige datasets.

Wat wordt gedeeld?

- L Indien mogelijk worden de ruwe, onverwerkte data als open data ter beschikking gesteld. Bij voorkeur worden ook de naverwerkte data, onderzoeksresultaten en -inzichten 'open' gedeeld.
- L In sommige gevallen kunnen gegevens niet onder een open licentie vrijgegeven worden omdat daar een goede reden voor is. Zo kunnen de data gevoelig zijn (bv. locatie van bedreigde diersoorten, ingevoerd via www.waarnemingen.be van Natuurpunt), worden ze gebruikt om inkomsten te garanderen voor het onderhoud van de dataset waarin ze gepubliceerd worden of worden ze omwille van de bescherming van de privacy, confidentialiteit of vooraf vastgelegde copyrights niet ontsloten. Data aggregeren vooraleer deze te publiceren kan een oplossing zijn wanneer de onverwerkte data privacygevoelig zijn (zie ook **principe 7**).
- L In andere gevallen moeten gegevens kosteloos en onder een open licentie vrijgegeven worden. Dat is vanaf juli 2021 het geval voor onderzoeksgegevens die geheel of gedeeltelijk gefinancierd zijn met overheidsmiddelen, vanwege de omzetting van de **PSI-richtlijn**.

Waarom dit principe?

- A **Ethiek**: citizen science is bij uitstek een onderzoeksvorm die steunt op de vrijwillige medewerking van burgers. Het is dan ook correct om hen minstens toegang te verschaffen tot de gegevens die ze zelf gegenereerd of geanalyseerd hebben (zie **principe 12**).
- B **Linked data**: daarnaast is het bij burgerwetenschap minstens zo belangrijk als in regulier wetenschappelijk onderzoek om gegenereerde datasets open en linked te maken, omdat 'vele kleine beetjes' (i.e. een aantal meetwaarden of een kleine dataset waaruit enkel binnen de scope van het project conclusies kunnen getrokken worden) na verloop van tijd samen een unieke, ruimere dataverzameling gaan opleveren waaruit bijkomende conclusies kunnen afgeleid worden (zie **principe 14**).
- C **Wetgeving** legt het op: legt het op conform de **PSI-richtlijn**

Hoe toepassen?

- 1 Bepaal van elke ruwe of naverwerkte dataset of deze (a) een meerwaarde kan betekenen voor andere gebruikers (b) open gepubliceerd mag/moet worden.
- 2 Voorzie de opengestelde data van een licentie uit de beknopte lijst in **principe 2**.
- 3 Maak de opengestelde data bevroegbaar via een web browser, dus plaats deze achter een http(s)- of (s)ftp-link.
- 4 Beschrijf kort welke data niet opengesteld worden, en wat daar de motivatie voor is.

⁵ vrij vertaald naar <http://opendefinition.org>, <https://okfn.org/opendata/> en https://opendatahandbook.org/guide/nl_BE/what-is-open-data/

⁶ Ook gekend als Public Sector Information richtlijn of PSI-richtlijn: <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN>. Artikel 10 van deze richtlijn bepaalt onder welke voorwaarden onderzoeksgegevens voor hergebruik beschikbaar moeten zijn, artikel 14 van de richtlijn bevat de regelingen voor publicatie en hergebruik van hoogwaardige datasets.

⁷ <https://overheid.vlaanderen.be/organisatie/informatiemanagement/omzetting-psi-richtlijn>



Praktijkvoorbeeld

Het project VLINDER⁸ (VLaanderen In De wEEr) van de Universiteit Gent bestaat uit een netwerk van 60 weerstations, die bij onderzoekinstellingen, lokale partners en scholen over heel Vlaanderen verschillende weerparameters (temperatuur, neerslag, luchtdruk, luchtvochtigheid, windsnelheid) monitoren. Ook het project Leuven.cool⁹ (KU Leuven, KMI, klimaatvzw Leuven2030) doet onderzoek naar stedelijke microklimaten via sensoren in de stad en in de tuinen van burgers. Beide projecten koppelen hun data aan het crowdsourcingplatform WOW-BE¹⁰ en stellen het daar ook open beschikbaar. Per weerstation kan eender wie de gegevens rechtstreeks downloaden in verschillende formaten (bv. SVG, PDF, XLSX, CSV). De gebruiksvoorwaarden en informatie over licenties zijn (nog) niet expliciet of duidelijk vermeld, maar vallen te raadplegen op de website van UK MetOffice¹¹.

Het project Nieuws van de Grote Oorlog¹² van Het Archief¹³ gebruikt beelden van kranten uit de Eerste Wereldoorlog. Sommige teksten of foto's in die kranten kunnen echter nog beschermd zijn door auteursrechten, waardoor ze niet zomaar als 'open' kunnen gepubliceerd worden. Om ze toch als '**Linked**' te kunnen publiceren, zijn enkel de

metadata van de kranten onder een open licentie beschikbaar.

De eigenlijke teksten werden dan weer gepubliceerd als linked data¹⁴ waarop de gebruiksvoorwaarden van toepassing zijn.

Zo kunnen de krantenartikelen bijvoorbeeld gelinkt worden met de online Namenlijst¹⁵ van oorlogsslachtoffers (militairen en burgers) van het 'In Flanders Fields'-museum¹⁶. Hoewel deze databanken geen resultaat zijn van citizen-scienceprojecten, kunnen ze toch als bron dienen voor citizen-scienceprojecten in de toekomst.

Ook het citizen-scienceproject Vespa-Watch¹⁷, gefinancierd door het Departement Economie, Wetenschap en Innovatie van de Vlaamse Overheid, geeft haar data systematisch als 'open' vrij. De modaliteiten rond die openheid staan beschreven in het open-databeleid¹⁸ van het Instituut voor Natuur- en Bosonderzoek¹⁹ (INBO), dat dit project leidt.

⁸ <https://vlinder.ugent.be/onderzoekers.html>

⁹ <https://leuven.cool/>

¹⁰ <https://wow.meteo.be/nl/over-wow-be/the-weather-together/>

¹¹ <https://www.metoffice.gov.uk/about-us/legal>

¹² <https://nieuwsvandegrooteoorlog.hetarchief.be/nl>

¹³ <https://www.hetarchief.be>

¹⁴ De website van Meemo, het Vlaams Instituut voor het Archief, bevat uitgebreide informatie over Linked en Open Data: <https://meemoo.be/nl/publicaties/linked-data-publiceren-en-inlezen-waar-begin-je>

¹⁵ <https://database.namenlijst.be/publicsearch/#/search/language=nl>

¹⁶ <https://www.inflandersfields.be/nl/kenniscentrum/namenlijst/>

¹⁷ <https://vespawatch.be/>

¹⁸ <https://www.vlaanderen.be/inbo/opendatabeleid/>

¹⁹ <https://www.vlaanderen.be/inbo/home/>

2. *Publiceer je data onder een open licentie die je kiest uit een beknopte, aanbevolen lijst.*

Licenties zijn in essentie een set van regels die beschrijven onder welke voorwaarden gegevens of kennis kunnen gebruikt worden. Ze vormen het legale kader voor de omgang met gegevens (van alle aard) dat nodig is om verwarring te vermijden rond bv. intellectuele eigendomsrechten die al dan niet op gegevens of informatie rusten. Gebruik licenties die open, internationaal bruikbaar en machineleesbaar zijn.



Op <http://opendefinition.org/licenses/> vind je de volledige lijst met licenties die passen binnen de definitie van 'open' zoals vastgelegd door de Open Knowledge Foundation²⁰. Volgens Groom et al. (2017) zijn de meest courant gebruikte licenties in citizen science deze van Creative Commons CC (hoofdzakelijk CC0, CC-BY, CC-BY-NC) en/of Open Data Commons ODC (hoofdzakelijk PDDL, ODC-BY, ODbL). Ook de CC-licenties CC-BY-SA en CC-BY-NC-SA kunnen gebruikt worden.

Het principe van de **CC-licenties** is dat ze bestaan uit een eenvoudige combinatie van vier restricties voor het delen:

- L **'BY'** betekent dat er moet gerefereerd worden naar de bron via een naamsvermelding;
- L **'NC'** (noncommercial) betekent dat data niet voor commerciële doeleinden mogen gebruikt worden;
- L **'SA'** (share-alike) duidt op de toestemming om afgeleide werken te maken op voorwaarde dat deze vrijgegeven worden onder dezelfde licentie als het originele materiaal;
- L **'ND'** (no derivatives) houdt in dat het originele materiaal niet mag aangepast worden of gebruikt als basis voor afgeleide werken. Deze restrictie is echter in de eerste plaats van toepassing op kunstwerken en (geschreven) media, en wordt hier daarom buiten beschouwing gelaten.

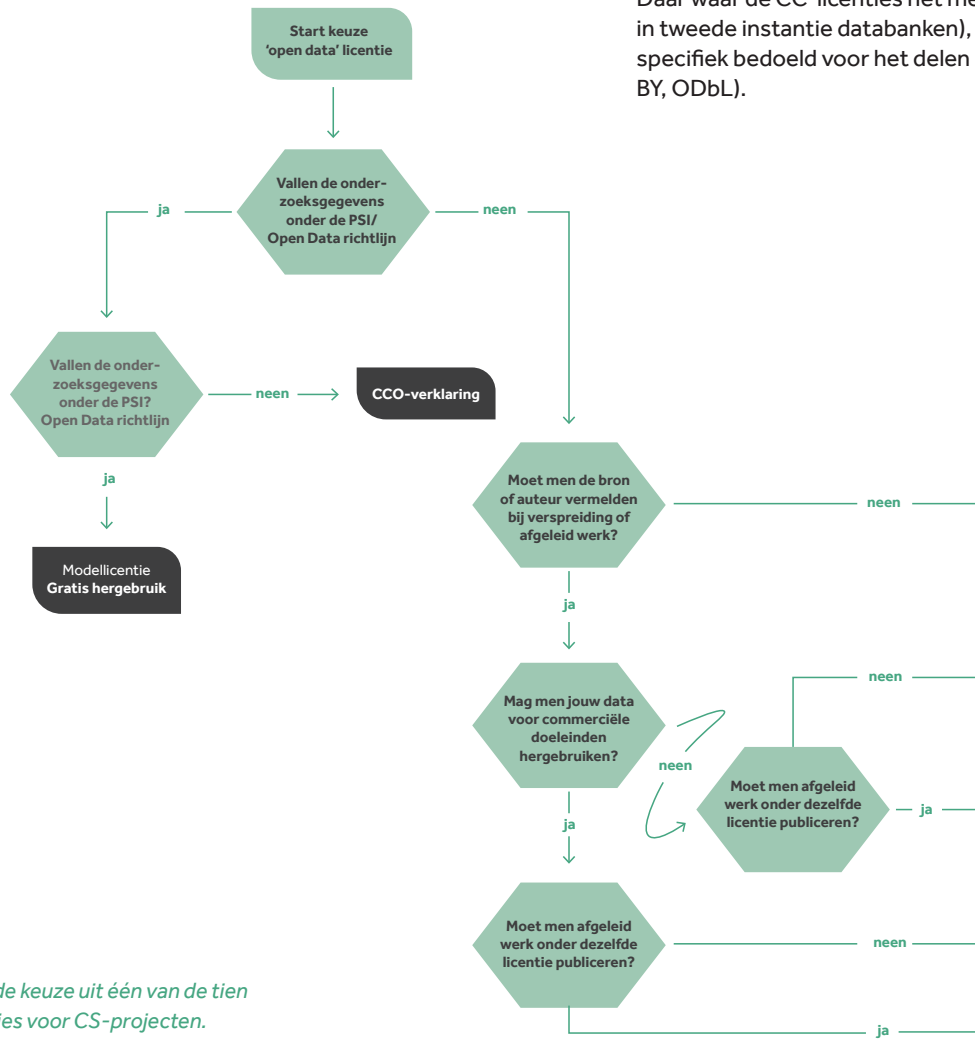
Daarnaast bestaat een publieke-domeinverklaring (CC0) die geen enkele restrictie oplegt.

Tip

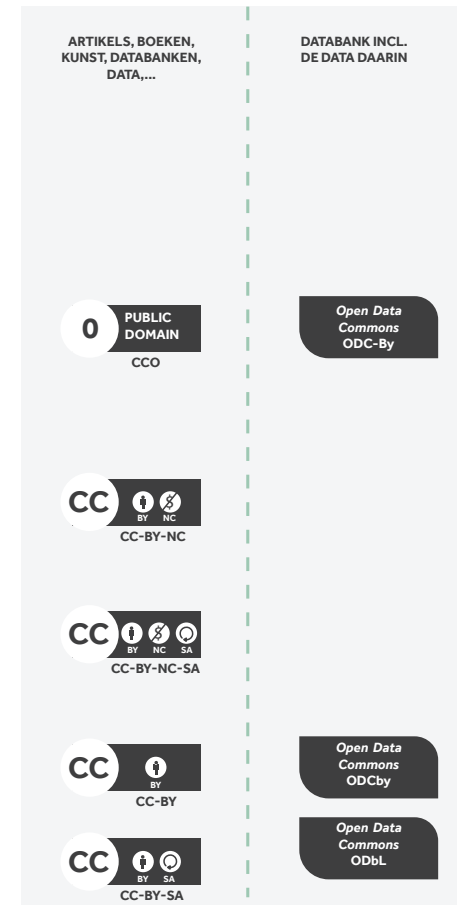
stapsgewijze hulp bij het 'samenstellen' van de meest geschikte CC-licentie krijg je op <https://creativecommons.org/share-your-work/>.

²⁰ <https://okfn.org/>





Daar waar de CC-licenties het meest geschikt zijn voor het delen van data (en in tweede instantie databanken), zijn de **Open Data Commons (ODC)**-licenties specifiek bedoeld voor het delen en beschermen van **databanken** (PDDL, ODC-BY, ODbL).



Figuur 3
Beslisboom voor de keuze uit één van de tien aanbevolen licenties voor CS-projecten.

Vallen je onderzoeksgegevens onder de **PSI/Open Data-richtlijn**²¹?

Dan maak je gebruik van:

- L CC0-verklaring
<https://overheid.vlaanderen.be/cc0-verklaring>, of
- L Modellicentie 'Gratis Hergebruik'
<https://overheid.vlaanderen.be/modellicentie-gratis-hergebruik>

Hieronder vind je de acht overige licenties:

- L CC0 "Publiek domein"
<https://creativecommons.org/share-your-work/public-domain/cc0/>

Merk op: conventies rond wetenschappelijke bronvermeldingen blijven behouden.
- L CC-BY "Naamsvermelding"
<https://creativecommons.org/licenses/by/4.0>
- L CC-BY-NC "Naamsvermelding-NietCommercieel"
<https://creativecommons.org/licenses/by-nc/4.0>
- L CC-BY-SA "Naamsvermelding-GelijkDelen"
<https://creativecommons.org/licenses/by-sa/4.0>
- L CC-BY-NC-SA "Naamsvermelding-NietCommercieel-GelijkDelen"
<https://creativecommons.org/licenses/by-nc-sa/4.0>
- L ODC-PDDL (Public Domain Dedication and License)
<https://opendatacommons.org/licenses/pddl/summary/>
- L ODC-By (Attribution License)
<https://opendatacommons.org/licenses/by/summary/>
- L ODC-ODbL (Open Database License)
<https://opendatacommons.org/licenses/odbl/summary/>



Praktijkvoorbeeld

CC-licenties binnen citizen-scienceprojecten EWI

- L Het project 'Vespa-Watch' (EWI-oproep 2017) publiceert haar data onder de CC0-licentie, omdat het gebruik maakt van het platform iNaturalist. De data van iNaturalist worden gekoppeld aan het wereldwijde open dataplatform GBIF²² (Global Biodiversity Information Facility; zie ook **Principe 5**). Bij GBIF geldt internationaal de afspraak dat gekoppelde gegevens gepubliceerd worden onder een CC0, CC-BY of een CC-BY-NC- licentie, dan wel vrijgegeven worden onder een equivalente licentie, met voorkeur voor de CC-BY (zie 'GBIF Terms of Use'²³).
- L Het project 'AstroSounds' (EWI-oproep 2019) hanteert een 'CC-BY-NC-SA'-licentie.

Tip

Een groot aantal bestaande platformen om waarnemingen te uploaden maakt gebruik van specifieke licenties. Check als deelnemer aan een citizen-scienceproject steeds onder welke licentie jouw ingevoerde gegevens zullen vallen. Omgekeerd is het voor coördinatoren van citizen-scienceprojecten belangrijk deze informatie duidelijk mee te geven aan alle bijdragers bij de start van het project.

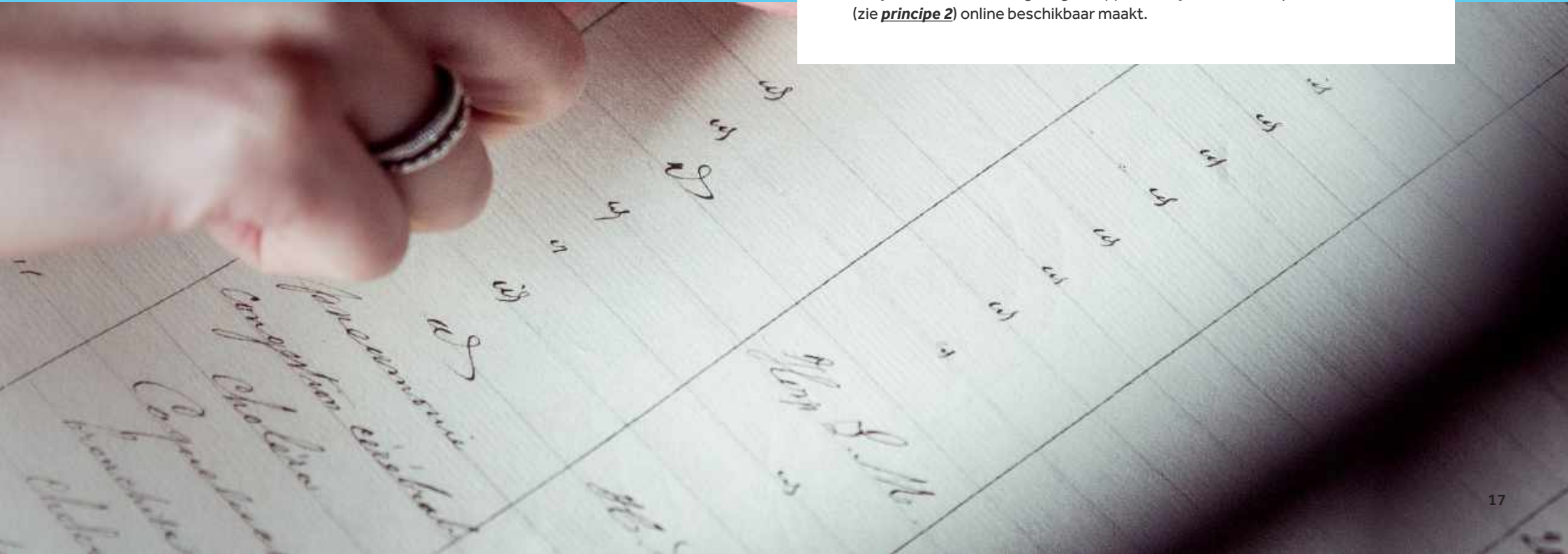
²¹ Zie <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN>. Artikel 10 van deze richtlijn bepaalt onder welke voorwaarden onderzoeksgegevens voor hergebruik beschikbaar moeten zijn. Voor hergebruik van andere bestuursdocumenten (behalve hoogwaardige datasets) kan een vergoeding gevraagd worden en moet desgevallend gebruik gemaakt worden van de modellicentie 'hergebruik tegen vergoeding': zie <https://overheid.vlaanderen.be/modellicentie-hergebruik-tegen-vergoeding>

²² <https://www.gbif.org/>

²³ <https://www.gbif.org/terms>

3. *Publiceer je onderzoeksresultaten en -bevindingen waar mogelijk in 'Gold Open Access Journals'.*

Als je data publiceert, moet je ook de context en betekenis van die data meegeven: het is belangrijk om betrouwbare informatie beschikbaar te stellen over het onderzoeksdoel, de aangewende methoden, de resultaten en hun eventuele betekenis. Als wetenschapper publiceer je je bevindingen bij voorkeur in 'Gold Open Access Journals', die peer-reviewed zijn en open voor iedereen. Is een (open) peer-reviewed artikel geen haalbare kaart? Beschrijf dan je onderzoek in een gedegen rapport, dat je onder een open licentie (zie [principe 2](#)) online beschikbaar maakt.



Academici gebruiken hoofdzakelijk het systeem van 'peer-reviewed' publiceren. Wanneer een onderzoeksartikel naar een tijdschrift wordt gestuurd, lichten (meestal anonieme) externe wetenschappers het artikel grondig door, waarna ze het ofwel goedkeuren, afkeuren, of goedkeuren mits kleine of grote aanpassingen. Het is belangrijk erop te letten dat artikels die claimen 'onderzoek' te publiceren, verschenen zijn in 'peer-reviewed' tijdschriften ('journals'). Deze worden in wetenschappelijke kringen als de meest betrouwbare artikels beschouwd, zeker als ze gepubliceerd worden in tijdschriften die opgenomen zijn in de [Web of Science](#)²⁴ databank (de zogenaamde 'A1-artikels'). Ze hebben ook allen een **DOI-nummer**, en bevatten voor iedere uitspraak die niet rechtstreeks het resultaat is van het eigen onderzoek, een verwijzing (referentie) naar een ander, eerder gepubliceerd artikel of betrouwbare bron.

Steeds meer universiteiten en overheidsinstellingen publiceren hun wetenschappelijke artikels in 'Open Access'. Deze artikels zitten niet verborgen achter betaalmuren of copyright-restricties. 'Open Access' voor publicaties betekent vrij beschikbaar zijn op het internet, zodat eender welke gebruiker de publicaties kan lezen, downloaden, kopiëren, (ver)delen, printen, zoeken of linken, zonder financiële, wettelijke of technische restricties. De enige voorwaarden houden verband met het correct citeren van auteurs en het behouden van de onderzoeksintegriteit ([Budapest Open Access Initiative, 2002](#)²⁵).

Publicaties kunnen op twee manieren Open Access gemaakt worden:

- L door het archiveren van documenten in publieke archieven (= 'Green Open Access')
- L door het artikel rechtstreeks te publiceren in Open Access-tijdschriften (= 'Gold Open Access').

Publieke archieven kunnen instituutgebonden zijn (bv. per onderzoeksinstituting zoals [UGent](#)²⁶ of [INBO](#)²⁷, die de gepubliceerde artikels van de eigen wetenschappers online bijhouden), of topicgebonden (bv. via [preprint-services](#) zoals [ArXiv](#)²⁸). Een internationale lijst van publieke archieven vind je op [OpenDOAR](#)²⁹.

²⁴ <https://apps.webofknowledge.com/>

²⁵ <https://www.budapestopenaccessinitiative.org/read>

²⁶ <https://biblio.ugent.be/>

²⁷ <https://www.vlaanderen.be/inbo/publicaties/>

²⁸ <https://arxiv.org/>

²⁹ <https://v2.sherpa.ac.uk/opensoar/about.html>

³⁰ [PLOS](#) - <https://plos.org/>

³¹ [BMC](#) - <https://www.biomedcentral.com/>

³² <https://www.nature.com/srep/>

³³ [DOAJ](#) - <https://www.doaj.org/>

³⁴ <https://open-research-europe.ec.europa.eu/>

De tweede weg (die van 'Gold Open Access'), die te verkiezen is, loopt via **Open Access Journals** (OAJ). Dit zijn tijdschriften die de gepubliceerde artikels vrij beschikbaar en gratis maken voor iedereen op het internet. Bekende Open Access-uitgeverijen zijn [Public Library of Science](#)³⁰ en [BioMed Central](#)³¹. Ook de groep rond het befaamde wetenschappelijke tijdschrift Nature heeft een Open Access-tijdschrift onder de vleugels: [Scientific Reports](#)³².

Een lijst van Open Access Journals vind je in de [Directory of Open Access Journals](#)³³. Deze website geeft wereldwijd toegang tot meer dan 15 000 Open Access wetenschappelijke wereldwijd, met meer dan 5 miljoen peer-reviewed artikels.

Tip

voor scholen werd recent het allereerste 'peer reviewed' wetenschappelijke tijdschrift in Europa in het leven geroepen: [The Open Schools Journal for Open Science](#). Het tijdschrift aanvaardt wetenschappelijke artikels geschreven door leerlingen van zowel lagere als middelbare scholen, die onder begeleiding van hun leerkrachten wetenschappelijke data produceren, bv. via een citizen science initiatief. Het tijdschrift publiceert in verschillende talen en is ook gratis.

Tip

wordt jouw project door het Europese Horizon Europe programma gefinancierd? Dan kan je jouw onderzoek publiceren in het nieuwe [Open Research Europe publicatieportaal](#)³⁴.

Is je onderzoek niet voldoende sterk om door het review-proces van een wetenschappelijk tijdschrift te geraken? Maak dan toch een duidelijk, volledig rapport met (a) het onderzoeksdoel, (b) een inleiding die het onderzoek kadert in een bredere context, (c) de methoden die gebruikt werden om de data te verzamelen, waaronder ook de gevolgde procedures voor het vrijwaren van een goede datakwaliteit, (d) de resultaten zelf, eventueel in een Annex, (e) de statistische analyses op de resultaten en/of de interpretatie van de resultaten [dit heet 'de discussie'], en vervolgens (f) je conclusies uit het onderzoek. Een dergelijk verslag helpt om aan te tonen dat de verzamelde data van je project betrouwbaar zijn en laat anderen toe om zelf een inschatting te maken van de betrouwbaarheid van je data. Dit vergroot de kans dat anderen je data ook verder zullen gebruiken.



Praktijkvoorbeeld

De publicatie '**Predicting the current and future distribution of the western black-legged tick, *Ixodes pacificus*, across the Western US using citizen science collections**' van Porter et al. (2021) is een A1-publicatie in het Open Access tijdschrift PLOS ONE³⁵, dat publiceert onder een CC-BY licentie.

Ook het A1-artikel 'A checklist recipe: making species data open and FAIR' van Reyershove et al. (2020) is gepubliceerd onder CC-BY licentie in het Open Access tijdschrift Database³⁶.

Het tijdschrift 'Citizen Science, Theory and Practice³⁷ van de Amerikaanse Citizen Science Associatie (CSA³⁸) is geen A1-tijdschrift, maar is wel peer-reviewed en Open Access.

³⁵ <https://journals.plos.org/plosone/>

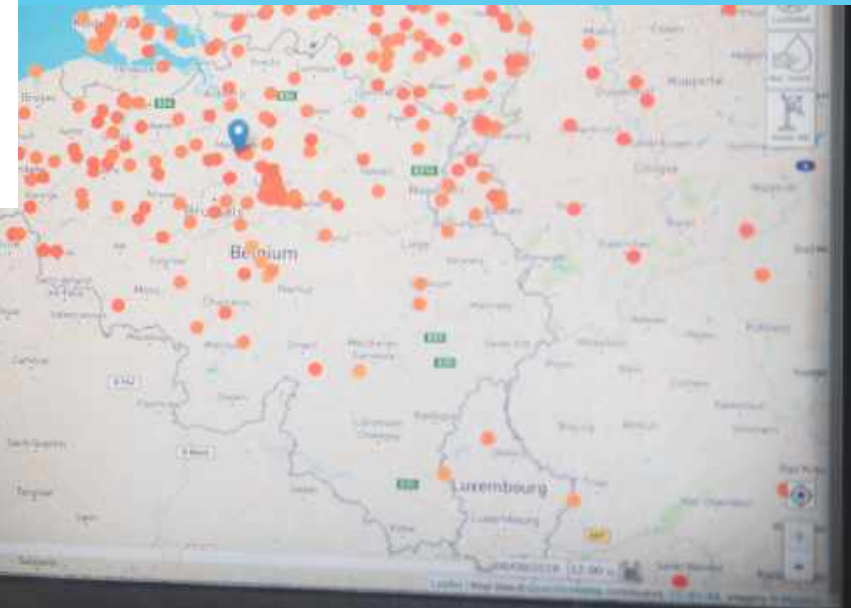
³⁶ <https://academic.oup.com/database/pages/About>

³⁷ <https://theoryandpractice.citizenscienceassociation.org/>

³⁸ <https://www.citizenscience.org/>

4. *Publiceer, waar mogelijk, ook de software die je ontwikkelt onder open licenties.*

Soms wordt er software ontwikkeld voor of door een citizen-scienceproject. Het wordt sterk aanbevolen om ook deze broncode open te publiceren. Er zijn een aantal goede redenen om achterliggende code 'open source' te maken.³⁹ Zo kunnen anderen bijdragen aan de ontwikkeling en testing van de applicatie. De open broncode biedt ook volledige transparantie over hoe met de data wordt omgegaan. Open-sourceapplicaties trekken bovendien een 'community' van geïnteresseerden aan die willen bijdragen aan de oplossing. Je kan ook op jouw beurt gebruik maken van 'open-source libraries' die anderen aanbieden.



³⁹ Lees meer over de voordelen van FOSS op <https://opensource.com/article/17/8/enterprise-open-source-advantages>.

Er bestaan een aantal gekende websites⁴⁰ waarop je broncode open kunt publiceren:

L GitHub

bv. <https://github.com/opendata-stuttgart/sensors-software> (scripts en applicaties), <https://github.com/opendata-stuttgart/luftdaten-local-labs> (website met JavaScript) en <https://github.com/dylanrees/citizen-science> (bronnen)

L GitLab

bv. <https://gitlab.com/micro-iot-platform/projects/temp-humidity-monitor>

L BitBucket

Er bestaat een 'Open Source Project License' aangeboden door moederbedrijf Atlassian: <https://www.atlassian.com/software/views/open-source-license-request>.

L SourceForge

Een populaire website bij open-sourceprojecten: <https://sourceforge.net/>.

L Google Code

Niet zozeer voor publicatie, maar het vermelden waard, omdat Google interessante 'code libraries' open aanbiedt via <https://opensource.google/projects> (bv. code voor data-analyse of visualisatie).

De meest gebruikte licenties⁴¹ voor open software(code) zijn:

L MIT License

<https://choosealicense.com/licenses/mit/>

Een korte en eenvoudige licentie, die anderen toelaat ongeveer alles te doen met jouw software(code), ook het maken en distribueren van gesloten ('closed source') versies, onder een andere licentie, waarbij de broncode niet noodzakelijk wordt vrijgegeven.

L Apache License 2.0

<https://choosealicense.com/licenses/apache-2.0/>

Net zoals bij de 'MIT License' mogen afgeleide werken onder een andere licentie, en desnoods zonder de broncode, worden verspreid. Tegenover MIT bouwt deze licentie wat meer beschermingsclausules in (bv. rond patenten) en is daardoor beter geschikt voor grotere softwareprojecten.

L GNU General Public License v3.0 (GPLv3)

<https://choosealicense.com/licenses/gpl-3.0/>

Deze licentie laat ongeveer alles toe, behalve het maken en verspreiden van gesloten versies. De broncode moet steeds ontsloten worden. Afgeleid werk moet je onder dezelfde licentie aanbieden, en het moet duidelijk zijn welke wijzigingen werden aangebracht.

Meer nuance en toelichting bij de keuze van de meest aangewezen licentie is te vinden op <https://opensource.guide/legal/>.

⁴⁰ gebaseerd op <https://itsfoss.com/github-alternatives/>

⁴¹ gebaseerd op <https://opensource.guide/legal/>



Praktijkvoorbeeld

In het project NoiseTube⁴² (Vrije Universiteit Brussel) wordt aan mensen een app aangeboden waarmee zij geluid (dB) kunnen monitoren met hun smartphone. Het project publiceerde de broncode voor de gebruikte software open op de server van de VUB⁴³, onder een GNU-LGPL 2.1-licentie. Het project staat ook vermeld in Google Code⁴⁴, waardoor de broncode makkelijker vindbaar wordt (naar analogie met principe 25, maar dan voor de software gebruikt bij het project).

Ook in het project VLINDER (Universiteit Gent) wordt de code⁴⁵ achter het dashboard⁴⁶ opgesteld op Github.

Het crowdsourcingplatform DoeDat⁴⁷ (Plantentuin Meise) vormt een uitbreiding op de interface van het open-sourceplatform DigiVol⁴⁸ (The Australian Museum & Atlas of Living Australia); de DoeDat-interface biedt bijvoorbeeld meerdere talen aan. DoeDat heeft haar eigen code⁴⁹ beschikbaar gesteld via Github zodat andere instituten ook een meertalige versie van DigiVol kunnen opzetten.

Bij het citizen-scienceproject Vespa-Watch⁵⁰ (INBO & Honeybee Valley) werden waarnemingen van Aziatische hoornaars geregistreerd en gesynchroniseerd met het platform iNaturalist.⁵¹ In het eindrapport van het project (Schoonvaere et al. 2020) wordt duidelijk omschreven dat de applicatie gebruik maakt van open source software en frameworks, en ook zelf als open source software werd ontwikkeld. De broncode is beschikbaar op <https://github.com/inbo/vespa-watch>. Er werd ook een functionaliteit ontwikkeld om data te synchroniseren met het iNaturalist-platform, die als losstaande open source module (Python library) werd ontwikkeld en beschikbaar is op <https://pyinaturalist.readthedocs.io/>.

Ook het Europese Sensor.community⁵² -netwerk biedt haar software⁵³ voor het uitlezen van luchtkwaliteitssensoren open aan op Github, net als het Zooniverse-platform⁵⁴ dat onder de open Apache 2.0-licentie⁵⁵ staat.

⁴² NoiseTube - <http://www.noisetube.net/help> (website) en <https://code.google.com/archive/p/noisetube/> (broncode)

⁴³ <https://gitlab.soft.vub.ac.be/groups/NoiseTube>

⁴⁴ <https://code.google.com/archive/p/noisetube/>

⁴⁵ <https://github.com/bmesuere/vlinder>

⁴⁶ <https://vlinder.ugent.be/dashboard/?stations=vlinder02&stations=vlinder21&stations=vlinder41>

⁴⁷ <https://www.doedat.be/>

⁴⁸ <https://digivol.ala.org.au/>

⁴⁹ <https://github.com/AgentschapPlantentuinMeise/volunteer-portal>

⁵⁰ <https://vespawatch.be/>

⁵¹ <https://www.inaturalist.org/>

⁵² <https://sensor.community/en/>

⁵³ <https://github.com/opendata-stuttgart/>

⁵⁴ <https://github.com/DiamondLightSource/zooniverse/blob/master/LICENSE>

⁵⁵ <https://www.apache.org/licenses/LICENSE-2.0>

5. Ga actief op zoek naar bestaande open data.

Steeds meer gegevens zijn als open data op het web beschikbaar. Ga actief op zoek naar bestaande open data die relevant zijn voor jouw project. Je bespaart jezelf mogelijk veel tijd en kosten, en het kan je project inhoudelijk sterk verrijken. Gebruik wel betrouwbare bronnen, die bij voorkeur compatibel zijn met courante datastandaarden (zie **principe 17**). Ook de overheid voert een actief open-databeleid: maak daar dus gebruik van. Onderzoekers hebben recht op toegang tot en hergebruik van overheidsinformatie in toepassing van de **PSI/Open-Data-richtlijn**.



Tip

Voor meer informatie over **open data bij de Vlaamse overheid** (charter, actieplan, modellicenties, metadata- en URI-standaarden) kan je terecht op <https://overheid.vlaanderen.be/open-data-bij-de-vlaamse-overheid>.

Tip

Neem een kijkje op volgende dataportalen.



Algemene data	Vlaams Open Dataportaal (VODAP): https://opendata.vlaanderen.be/	https://data.gov.be/nl	https://data.europa.eu/euodp/en/home	https://dataverse.harvard.edu/
			https://www.european-dataportal.eu/	
Geodata	Geopunt: https://www.geopunt.be/	Geo.be: https://www.geo.be/	INSPIRE-dataportaal: https://inspire-geoportal.ec.europa.eu/	
Biodiversiteit				GBIF: https://www.gbif.org/
Genetica				GenBase: https://www.ncbi.nlm.nih.gov/genbank/
Mariene biologie				WoRMS: http://marinespecies.org/
Atmosfeer: broeikasgassen			ICOS: https://www.icos-cp.eu/data-services/about-data-portal	
Mobiliteit		Transportdata: https://www.transportdata.be/nl/		
Satellietbeelden	Terrascope (VITO): https://terrascope.be/nl			

Tip

Een basislijst van internationale open dataportalen vind je op de website van de 'Open Access' uitgeverij PLOS: <https://journals.plos.org/plosone/s/recommended-repositories>.

Tip

Voor het vinden van overheidsgerelateerde **authentieke gegevensbronnen**⁵⁶ op federaal, Vlaams en Waals niveau kan je terecht op <https://dt.bosa.be/nl/gegevensuitwisseling/authentieke-bronnen/overzicht-authentieke-bronnen>.

⁵⁶ Bij de overheid spreekt men van authentieke gegevens als zijnde 'unieke en oorspronkelijke gegevens' over personen en rechtsfeiten, die correct en up-to-date zijn. Men streeft ernaar die gegevens slechts één maal online te publiceren in een authentieke bron, zodat men slechts één maal de gegevens moet aandragen bij een overheidsinstantie, terwijl andere instanties (onder voorwaarden) deze bronnen kunnen raadplegen, en zij dus niet elk apart dezelfde gegevens moeten gaan opvragen bij de burger of bij rechtspersonen.

6. *Vraag raad aan ondersteunende diensten.*

Scivil faciliteert en ondersteunt citizen science in Vlaanderen. Daardoor is Scivil op de hoogte van de laatste ontwikkelingen in citizen science, en kan ze je dus ook op infrastructuren, gespecialiseerde instituten of partners wijzen. De Vlaamse overheid zet ook in op verscheidene initiatieven die allen werken naar het zo interoperabel en open mogelijk ontsluiten van onderzoeks- en overheidsgegevens. De Data Protection Officers van grote bedrijven en organisaties zijn hiervan op de hoogte; consulteer ook hen!



Tip

Naak gebruik van de kennishub die Scivil vormt, zodat je meteen begeleid kan worden naar de diensten en expertise van de vele partners van ons kenniscentrum. Op deze manier hoef je het warm water niet opnieuw uit te vinden en kan jouw project een vliegende start nemen.

Voor vragen rond de communicatie over je gegevens kan je alvast een kijkje nemen in de *Scivil communicatiegids*⁵⁷. Voor specifieke vragen rond privacy en ethiek verwijzen we ook naar *principe 7*.

Tip

Neem een kijkje op de websites van initiatieven als *VODAP*⁵⁸, *OSLO*⁵⁹, *VLOCA*⁶⁰ en *FOSB*⁶¹.



⁵⁷ <https://www.scivil.be/publications>

⁵⁸ Vlaams Open Data Portaal - <https://opendata.vlaanderen.be/>

⁵⁹ Open Standaarden voor Linkende Organisaties - <https://data.vlaanderen.be/>

⁶⁰ Vlaamse Open City Architectuur - <https://vloca.vlaanderen.be/>

⁶¹ Flemish Open Science Board - <https://www.ewi-vlaanderen.be/nieuws/flemish-open-science-board-fosb-opgericht>



II. Privacy & Ethiek

In vergelijking met klassiek wetenschappelijk onderzoek, brengt citizen science sterke, nieuwe uitdagingen rond privacy en ethiek. Dit kan gaan over (a) datakwaliteit en integriteit, (b) delen van resultaten en intellectuele eigendommen, (c) belangenvermenging en (d) inzet van vrijwilligers (zie Resnik et al. 2015). Vragen rond ethiek verschillen ook sterk van project tot project en van domein tot domein. Zo zijn in de medische wetenschappen bijvoorbeeld reeds goede richtlijnen opgesteld voor de verzameling van persoonlijke informatie.

In andere domeinen, zoals geologie of biologie, zijn persoonlijke gegevens vaak niet zozeer onderwerp van onderzoek, maar eerder een neveneffect, doordat citizen science wordt ingezet als methode. Denk bijvoorbeeld aan het risico op het blootgeven van verblijfplaatsen van mensen die locatiegebaseerde metingen doen, of denk aan foto's die deelnemers nemen van pakweg bomen, waar ongewild wandelende kinderen op afgebeeld staan. Als je een citizen-scienceproject initieert, al dan niet in cocreatie met burgers, is het belangrijk om hiermee doordacht om te springen in je onderzoeksopzet en om er ook duidelijk over te communiceren.

7. *Besteed actieve aandacht aan privacy en zorg voor kennisdeling.*

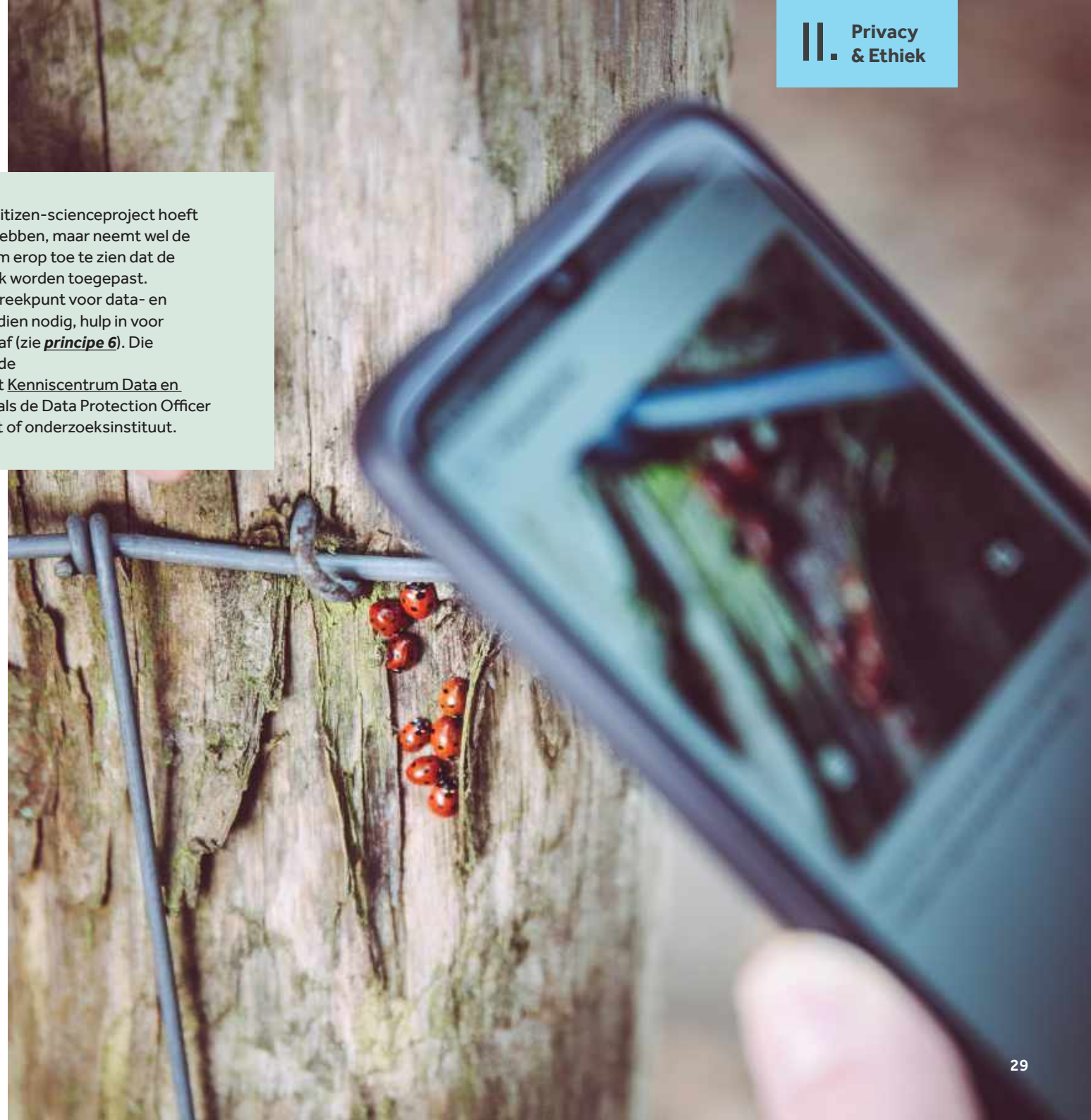
Om tot een correcte databehandeling te komen zet je als projectleider actief in op het informeren en opleiden van je medewerkers en je burgerwetenschappers. Daarbij hoort (a) een duidelijke communicatie over datamanagement, privacy en ethiek bij citizen science, (b) het aanwijzen van een dataverantwoordelijke binnen het project en (c) expliciete aandacht voor dit topic bij aanbestedingen en overeenkomsten met derden.

Burgerwetenschappers moeten een basisinformatiepakket krijgen voor ze aan het project deelnemen, over zowel privacy, rechten en plichten, als de integere omgang met data.



Tip

een dataverantwoordelijke binnen een citizen-scienceproject hoeft geen uitgebreide data-achtergrond te hebben, maar neemt wel de centrale verantwoordelijkheid op zich om erop toe te zien dat de principes uit dit charter zo goed mogelijk worden toegepast. Deze persoon is ook het centrale aanspreekpunt voor data- en privacygerelateerde vragen en roept, indien nodig, hulp in voor informatie en ondersteuning van buitenaf (zie **principe 6**). Die ondersteuning kan je vinden bij Scivil⁶², de Gegevensbeschermingsautoriteit⁶³, het Kenniscentrum Data en Maatschappij⁶⁴, en andere diensten - zoals de Data Protection Officer (**DPO**) die aanwezig is in elke universiteit of onderzoeksinstituut.



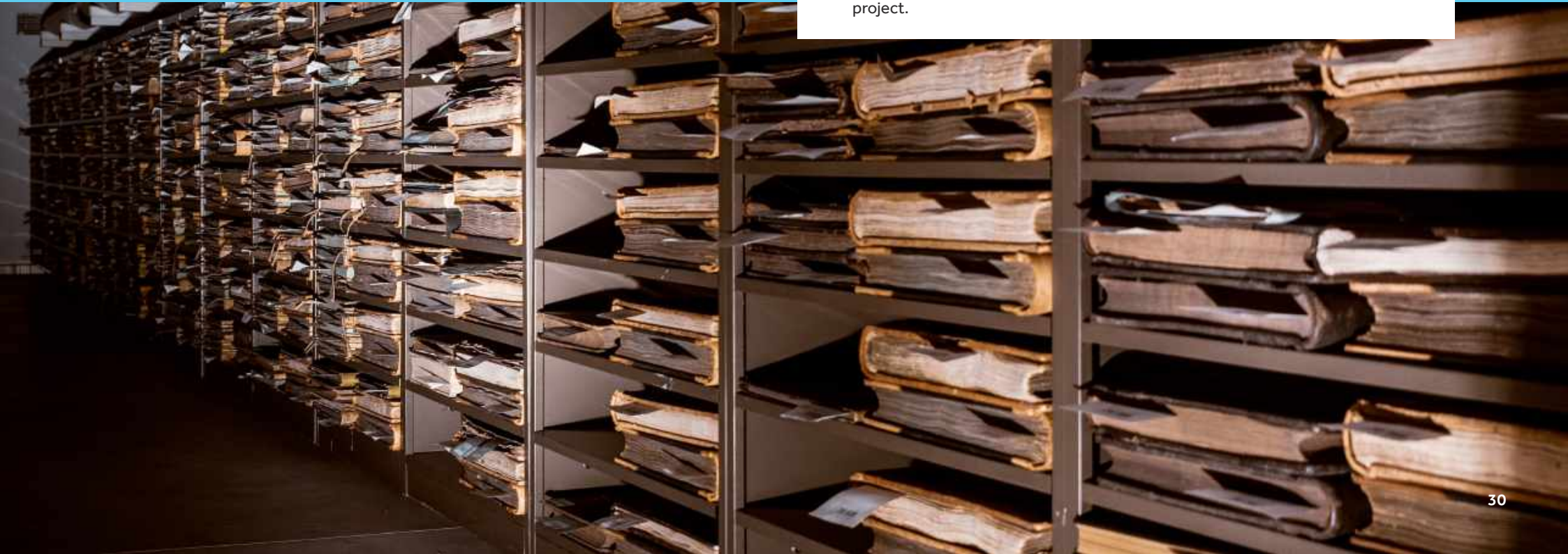
⁶² <https://www.scivil.be/>

⁶³ <https://www.gegevensbeschermingsautoriteit.be/professioneel>

⁶⁴ <https://data-en-maatschappij.ai/>

8. Kom met je project tegemoet aan de GDPR-richtlijnen.

Respecteer de **GDPR**⁶⁵-regels voor omgang met persoonsgegevens van zowel je deelnemende burgerwetenschappers, als van eventuele andere personen waarvan gegevens (on)opzettelijk gecapteerd worden tijdens je project.



Persoonsgegevens

Persoonsgegevens omvatten alle gegevens over fysieke personen, die, al dan niet in combinatie met andere gegevens, kunnen leiden tot de identificatie van een individu. Een goede, gedetailleerde beschrijving van wat persoonsgegevens zijn, vind je op de website van de [Europese Commissie](#)⁶⁶. Hieronder kunnen dus ook zaken zoals de locatie van een persoon, bepaalde medische eigenschappen of digitale gegevens (bv. gebruikersnamen of IP-adressen) vallen.

GDPR

De **GDPR** omvat een aantal basisprincipes, en een overkoepelend bijkomend principe ('verantwoordingsplicht'):

- L Rechtmatigheid (Lawfulness):** het verzamelen, verwerken en ontsluiten van persoonsgegevens moet onder een nationale of internationale wet vallen die het verwerken van die persoonsgegevens toelaat. Een goed overzicht van rechtsgronden vind je op de [website van de Universiteit Gent](#)⁶⁷ (toestemming, algemeen belang, gerechtvaardigd belang, wettelijke verplichting, uitvoering van een overeenkomst, vitaal belang). In sommige gevallen, zoals bijvoorbeeld bij **gevoelige gegevens**, is dit niet voldoende en moet de persoon wiens gegevens verwerkt worden, extra expliciete toestemming geven. De [gegevensbeschermingsautoriteit](#)⁶⁸ (GBA) beschrijft op haar [website](#)⁶⁹ o.a. waaruit de verhoogde bescherming voor dergelijke gegevens bestaat. Welke informatie je best opneemt in een informed consent, vind je [op deze pagina](#)⁷⁰ van de Universiteit Gent.
- L Transparantie (Transparency):** de personen waarvan je project de gegevens verzamelt, moet je vooraf duidelijk, volledig en transparant (dus begrijpbaar) op de hoogte brengen van hoe hun gegevens worden verwerkt, bewaard en/of vrijgegeven of doorgegeven. Ook dat moet wettig gebeuren.
- L Doelbinding (Purpose limitation):** de redenen waarom je persoonsgegevens verzamelt, moeten een specifiek, expliciet en wettig doel dienen. Deze redenen deel je ook mee aan de persoon, waarover je gegevens verzamelt. Ook de wijze van verwerking van de data moet compatibel zijn met het vooropgestelde doel. Dit betekent ook dat je eerder verzamelde gegevens voor een gespecificeerd doel niet mag gebruiken voor een ander doel.

- L Minimale gegevensverwerking (Data minimization):** de hoeveelheid gegevens en de datakwaliteit moet je zo laag mogelijk bij de doelstellingen van het onderzoek houden. Dit betekent dat je enkel data verzamelt die strikt nodig en relevant zijn voor het beoogde doel. Concreet raadt de GDPR hierin ook aan om alle gegevens, waar mogelijk, minstens te **pseudonimiseren**.
- L Opslagbeperking (Storage Limitation):** specificeer voor hoe lang je de gegevens bewaart, en onder welke vorm. Indien je data niet onmiddellijk bij de captering pseudonimiseert, moet je deze data, zodra identificatie van personen niet meer nodig is, zo snel mogelijk intern pseudonimiseren (dus: pseudonisatie zo dicht mogelijk bij de bron). Uitzonderingen op dit principe zijn o.a. toe te passen bij gegevensbewaring voor statistische doelen, die op lange termijn bewaard moeten worden.
- L Juistheid (Accuracy):** verwerkte persoonsgegevens moeten accuraat zijn, en altijd up to date gehouden worden. Inaccurate gegevens moet je onmiddellijk aanpassen of verwijderen zodra duidelijk wordt dat ze inaccuraat zijn. Beslissingen gebaseerd op inaccurate data kunnen aangevochten worden.
- L Integriteit en vertrouwelijkheid (Integrity & Confidentiality):** neem gepaste beschermingsmaatregelen om te voorkomen dat de verzamelde gegevens inaccuraat verwerkt, accidenteel verloren, verwijderd of beschadigd worden. Daar hoort ook bewustmaking van personeel en informatiebeheerders bij. Los van de wettelijke regels houd je je dus ook aan een aantal algemene principes. Maak o.a. zeker gebruik van een **SSL-certificaat** voor je website om datalekken te helpen voorkomen.
- L Verantwoordingsplicht (Accountability):** organisaties moeten zélf kunnen aantonen dat ze de nodige technische en/of organisatorische maatregelen hebben getroffen om in orde te zijn met de GDPR. Die verantwoordelijkheid ligt dus niet meer bij externe controlerende instanties. Opnieuw is informatiedeling hier cruciaal. De verantwoordingsplicht is van toepassing op alle voorgaande regels.

⁶⁶ Wat zijn persoonsgegevens? (Europese Commissie): https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_nl

⁶⁷ <https://onderzoektips.ugent.be/nl/tips/00001767/>

⁶⁸ <https://gegevensbeschermingsautoriteit.be/>

⁶⁹ <https://gegevensbeschermingsautoriteit.be/burger/thema-s/gevoelige-gegevens>

⁷⁰ <https://onderzoektips.ugent.be/nl/tips/00001772/>

Let op!

Anonimisatie en **pseudonimisatie** zijn twee verschillende begrippen. Het pseudonimiseren van persoonsgegevens (bijvoorbeeld het veranderen van een naam in een unieke code voor die persoon) leidt niet noodzakelijk tot het bekomen van strikt anonieme gegevens. In principe, en vooral bij citizen science, kunnen alle gegevens die verband houden met één persoon - al dan niet in combinatie met andere gegevens - leiden tot de uiteindelijke identificatie van die persoon. Vermijd dat je aan mensen een absolute anonimiteit garandeert indien je geen data-aggregatie (bij voorkeur nog voordat de gegevens opgeslagen worden) hebt voorzien in je dataprotocol. Pseudonimisatie en anonimisatie gebeuren bij voorkeur zo snel mogelijk na het verzamelen van de gegevens, om datalekken minder kansen te geven. Zolang je ook een lijst bewaart die de pseudoniemen/ codes koppelt met identificatiegegevens van personen, kan je niet van een absolute anonimiteit spreken - de identiteit van de personen is immers langs deze weg terug traceerbaar.

Bij de meeste citizen-scienceprojecten vind je een link naar de privacyverklaring in de 'kleine lettertjes' helemaal onderaan de webpagina. Bij sommige projecten krijg je de verklaring pas te lezen als je je effectief inschrijft als deelnemer aan het project. Zorg dat je privacyverklaring makkelijk te vinden is op je website, dat ze enerzijds volledig en juist is, en anderzijds ook bondig en duidelijk begrijpbaar voor iedereen. Handig is als je deze ook in PDF-vorm aanbiedt, zodat mensen ze kunnen downloaden.

Tip

Wil je algemene informatie delen met je citizen scientists over de GDPR, en hun rechten? Verwijs dan makkelijk door naar deze brochure van de Europese Commissie:
https://ec.europa.eu/info/sites/info/files/virtual_identity_nl.pdf

Tip

Ook universiteiten hebben vaak een goed uitgewerkte informatiepagina (bv. deze van [Universiteit Gent](https://www.ugent.be/en/research/damanager/policies/privacy.htm)⁷¹) over de GDPR of privacy in het algemeen. Vaak hebben ze, naast heldere informatie over de wettelijke regels, ook nog een eigen 'reglement' rond privacy en ethiek in het onderzoek. Bij de Universiteit Gent staan deze bijvoorbeeld helder uitgelegd op deze pagina: <https://www.ugent.be/en/research/damanager/policies/privacy.htm>. Voor uitgebreide informatie over privacy in het algemeen kan je ook terecht op de website van de [gegevensbeschermingsautoriteit](https://www.gegevensbeschermingsautoriteit.be/)⁷² (GBA).

Tip

Bij het opstellen van een datamanagementplan (DMP, zie **principe 13**) kan je gebruik maken van online tools en templates (bv. [DMPOnline](https://www.dmponline.be/)⁷³), waar ook GDPR-gerelateerde checklists in verwerkt zijn. Maak daar zeker gebruik van.

⁷¹ <https://onderzoektips.ugent.be/nl/tips/00001763>

⁷² <https://gegevensbeschermingsautoriteit.be/burger/privacy/wat-is-het>

⁷³ <https://dmponline.be/>



Praktijkvoorbeeld

Privacyverklaring

Het citizen-scienceproject Fietsbarometer⁷⁴ werkt samen met scholen om de fietsveiligheid van de routes naar scholen in kaart te brengen. Het project heeft een overzichtelijke, begrijpbare privacyverklaring⁷⁵, maar daarnaast ook een duidelijke pagina⁷⁶ met extra uitleg over wat er met alle data gebeurt en hoelang deze bewaard worden, alsook een aparte cookieverklaring⁷⁷.

De privacyverklaring duidt ook expliciet aan dat er geen **gevoelige persoonsgegevens** ('bijzondere gegevens') worden bewaard of verwerkt.

Andere voorbeelden van duidelijke privacyverklaringen, zijn deze van Curieuzeneuzen in de tuin⁷⁸, Astrosounds⁷⁹, en SOS-Antwerpen⁸⁰ die ook in PDF beschikbaar zijn.

Anonimisatie

Voor het beschermen van de privacy van de deelnemende leerlingen maakt Fietsbarometer gebruik van pseudonimisatie (elke leerling krijgt een unieke code) in combinatie met aggregatie: fietstrajecten of -segmenten worden pas op het platform weergegeven vanaf het dichtstbijzijnde kruispunt bij de woning, en enkel wanneer er minstens twee leerlingen langsheen het segment zijn gefietst. Zo is het onmogelijk om het thuis-school fietstraject van een enkele leerling te achterhalen.

In het citizen-scienceproject MamaMito⁸¹ stelden deelnemers hun moederlijke stamboom op en werd hun mitochondriaal DNA onderzocht en vergeleken met het DNA van andere deelnemers die via hun 'voormoederen' verwant waren. Deze informatie is bijzonder privacygevoelig. Vrijwilligers van Histories.vzw en Familiekunde Vlaanderen waren beschikbaar voor het helpen en het valideren van de stambomen die de deelnemers opstelden. Daarbij werden gegevens jonger dan 100 jaar (en die dus makkelijk door de deelnemers zelf te verifiëren zijn via bv. trouwboekjes en bidprentjes) niet door externen geverifieerd.

⁷⁴ <https://fietsbarometer.ugent.be/>

⁷⁵ <https://fietsbarometer.ugent.be/privacy>

⁷⁶ <https://fietsbarometer.ugent.be/jouwdata>

⁷⁷ <https://fietsbarometer.ugent.be/cookies>

⁷⁸ <https://curieuzeneuzen.be/privacy-statement/>

⁷⁹ <https://www.astrosounds.be/over-ons/privacy>

⁸⁰ <https://sosantwerpen.be/privacyverklaring/>

⁸¹ <https://historiesvzw.be/projecten/mamamito/>

9. Communiceer helder rond intellectuele eigendommen en copyrights.

Denk vóór de start van je project zorgvuldig na over de verschillende aspecten die kunnen leiden tot vraagtekens of spanningen rond intellectuele eigendommen. Dit gaat zowel over intellectuele eigendomsrechten op wat burgerwetenschappers bijdragen (bv. foto's, video's, zelfgeschreven teksten), als over de intellectuele eigendommen op of de patentering van een of meerdere outputs van je project. Schrijf correcte, volledige en duidelijke algemene deelnemingsvoorwaarden (en indien nodig gebruiks- of afnamevoorwaarden) uit, die je helder en begrijpbaar communiceert met je deelnemers vóórdat ze een bijdrage leveren.



Intellectuele eigendommen

Intellectuele eigendommen kunnen bij sommige citizen-scienceprojecten een struikelblok vormen voor verdere verspreiding of gebruik van de onderzoeksresultaten. De kans dat die problemen zich voordoen, hangt sterk af van het type bijdragen van de burgerwetenschappers in het project. Als burgers **foto's, video's of geschreven observaties** aanbrengen, is de kans op spanningen groter dan wanneer burgers bijvoorbeeld teksten transcriberen, of data in online formulieren intypen. Het aanbrengen van persoonlijke gegevens kan een bezorgdheid over privacy wekken, maar verandert weinig aan intellectuele eigendoms kwesties, met uitzondering van bv. portretrechten. Wanneer burgers echter bijdragen door **data-analyse of de ontwikkeling** van een product of protocol, komen er wel weer sneller vraagtekens rond intellectuele eigendomsrechten. Denk dus goed na, lang vóór de start van je project, over de intellectuele eigendommen en deelnemingsvoorwaarden van je project, want in functie hiervan kunnen je gekozen onderzoeksprotocollen verschillen. Zo kan het bijvoorbeeld betekenen dat je burgers gecategoriseerde waarnemingen laat doen in plaats van vrije eigen beschrijvingen van een waarneming. Ook voor het verspreiden of gebruiken van de resultaten van het project is het belangrijk om duidelijke schriftelijke gebruiksvoorwaarden te stellen.

Een derde type intellectuele eigendomsrechten waar je aandacht aan moet besteden zijn de zogenaamde **3rd party copyrights**. Dit zijn de eigendomsrechten van bijvoorbeeld bestaande platformen of websites die je gebruikt voordat je gegevens vergaart of verwerkt met of door burgerwetenschappers. Houd er rekening mee dat de Algemene Voorwaarden van dergelijke platformen bijkomende restricties (of rechten) kunnen opleggen aan zowel de bijdragers als de afnemers van de resultaten van je project.

In [Scassa & Chung \(2015\)](#) lees je alles over het managen van intellectuele eigendommen bij citizen science. Let wel: de auteurs zijn Canadees en refereren dus (bv. bij patenten) vaak naar het Amerikaanse of Canadese rechtssysteem.

Tip

Heb je een project waarbij de kans groot is dat er een patent zal aangevraagd worden op de resultaten? Wees dan voorzichtig met de (veelal intellectuele) mate waarin je inbreng van medewerkers vraagt.

Tip

Sommige projecten vragen aan deelnemers om een financiële bijdrage te leveren om te kunnen deelnemen. Ook hierover moet je juiste, volledige en heldere informatie meegeven in de Algemene Voorwaarden.

Let op!

Als een burger of organisatie grote hoeveelheden gegevens aan jouw project toevoegt, kan het zijn dat die data als een 'databank'⁸² beschouwd worden. Databanken vallen in Europa onder beschermingsrecht van de [EU Database Directive](#)⁸³, die stelt dat databanken die niet onder klassieke copyrights vallen, toch kunnen beschermd worden voor een termijn van 15 jaar, en hernieuwbaar telkens wanneer de databank een substantiële vernieuwing ondergaat. Deze bescherming wordt geschonden zodra iemand een 'substantieel deel' van de databank extraheert.

Let op!

Hoewel er op vlak van copyrights (doorgaans geldig tot 70 jaar na het overlijden van de copyright houder/auteur) universele akkoorden zijn gesloten om de regelgeving wereldwijd te harmoniseren, kunnen er toch land per land verschillen zijn in interpretatie van de regels. Omvat jouw project input uit verschillende landen, wees dan extra op je hoede ([Scassa & Chung, 2015](#)).

⁸² Een databank wordt in de [EU Database Directive](#) gedefinieerd als een verzameling onafhankelijke werken, data of andere materialen die in een systematische of op een methodologische manier geordend zijn, en die elk apart raadpleegbaar zijn.

⁸³ <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009>



Praktijkvoorbeeld

De gebruikersovereenkomst⁸⁴ van het biodiversiteitsplatform Waarnemingen.be⁸⁵ vermeldt duidelijk wat de rechten en plichten zijn van burgerwetenschappers en/of projectcoördinatoren die gegevens aanbieden of afnemen van het platform. Bovendien staat de tekst ook bondig en leesbaar op deze plaats⁸⁶. De overeenkomst omvat (a) algemene bepalingen, (b) rechten en plichten bij gebruik van multimediamateriaal van het platform, (c) kaartmateriaal, en (d) de wenselijke manier van bronvermelding voor het platform en voor gebruik van individuele multimediamaterialen (foto's, geluiden). Een nieuwe gebruiker publiceert zijn/haar data standaard onder een CC-BY-NC-licentie, maar kan die licentie wijzigen. Zo kiest elke gebruiker of zijn/haar multimediamateriaal beschikbaar is voor derden of niet. Naar bronvermelding toe vraagt Waarnemingen.be wel om het platform steeds te vermelden, ook als men bijvoorbeeld foto's publiceert die men zelf op het platform geüpload heeft. Subsets van de waarnemingen.be-databank kunnen enkel voor niet-commercieel individueel gebruik gedownload worden. Voor professioneel gebruik is een aparte toestemming of overeenkomst nodig.

Open Street Map⁸⁷ is geen citizen-scienceproject maar wordt zeer vaak gebruikt bij projecten, zoals bijvoorbeeld Fietsbarometer (Universiteit Gent) en Vespa-Watch⁸⁸ (INBO). Het valt onder de open licentie OdBL⁸⁹, waardoor het makkelijker bruikbaar wordt voor de visualisatie van citizen-scienceresultaten of gegevens. Het platform bepaalt echter wel in haar algemene voorwaarden⁹⁰ dat afgeleide werken onder dezelfde voorwaarden (licentie) moeten worden gepubliceerd als deze die gelden voor de bestaande kaarten van Open Street Map. Dit kan bij sommige onderzoekers conflicteren met hun taak om resultaten te dissemineren, en kan daarom dus een reden zijn om niet met Open Street Map te werken.

⁸⁴ <https://waarnemingen.be/pages/user-agreement/>

⁸⁵ <https://waarnemingen.be/>

⁸⁶ <https://waarnemingen.be/tos/>

⁸⁷ <https://www.openstreetmap.org/copyright>

⁸⁸ <https://vespawatch.be/>

⁸⁹ <https://opendatacommons.org/licenses/odbl/>

⁹⁰ <https://www.openstreetmap.org/copyright>

10. *Weeg je datakwaliteit en -kwantiteit voorzichtig af tegenover ethiek, haalbaarheid en projectdoelen.*

Meer dan in klassiek onderzoek moet je bij burgerwetenschap vooraf bewuste afwegingen maken tussen aspecten als

- (a) haalbaarheid,
- (b) datakwaliteit en wetenschappelijke output,
- (c) ethische aspecten (bv. privacy, intellectuele eigendommen, bescherming tegen gevaren),
- (d) directe meerwaarde voor de deelnemende burgers en
- (e) maatschappelijke impact.

Hoe dan ook moet de datakwaliteit op een wetenschappelijk verantwoorde manier voldoen aan de doelstellingen van je project.



Je kan op verschillende manieren een voldoende hoge datakwaliteit voor jouw project bewerkstelligen. Dit wordt concreter uitgelegd bij **principe 15**, waar de afweging tussen kwaliteit en kwantiteit wordt gemaakt in functie van de beoogde resultaten en de haalbaarheid van de gekozen methode. Het bekomen van wetenschappelijke data of analyses is bij citizen science echter niet het enige doel: projecten hebben veelal ook een educatieve en maatschappelijke waarde. Sommige projecten, die men (indien uitgevoerd met sensoren) eerder 'citizen sensing'⁹¹ noemt, dienen hoofdzakelijk een maatschappelijk doel, namelijk het aankaarten van een milieuprobleem bij een bepaalde autoriteit. Hoewel de gegenereerde data voldoende betrouwbaar moeten zijn om het projectdoel te dienen, leidt dit niet altijd tot voldoende data voor een robuuste studie die je kan publiceren in wetenschappelijke vaktijdschriften, en dat is dus ook niet altijd nodig. Belangrijk is wel dat je de gebruikte methoden en de robuustheid ervan goed communiceert.

Ethische restricties, zoals genoemd in **principes 8** en **11**, moeten samen met de haalbaarheid (bv. restricties door financiële belemmeringen, of door situationele belemmeringen als een Covid-lockdown) in overweging genomen worden bij het bepalen van de methoden om tot een goede datakwaliteit te komen.

Let op!

Ook de selectie van je deelnemers kan een invloed hebben op zowel je datakwaliteit als je ethische standaarden. Kies wijs in functie van je projectdoel: ligt de nadruk op maatschappelijke impact door het bereiken van moeilijk te bereiken groepen? Of heb je juist burgerwetenschappers nodig die in een specifiek domein sterk geschoold zijn (bv. beroepsverenigingen, hobbyverenigingen)?

⁹¹ Zie <https://research.tilburguniversity.edu/en/publications/sensing-the-risk-in-search-of-the-factors-influencing-the-policy/> voor meer informatie over Citizen Sensing.

11. *Let op voor ethisch ongewenste neveneffecten van het vrijgeven van je data.*

Overweeg bij elke stap in je project of het ethisch is om de verzamelde data vrij te geven. Soms kan je gegevens niet publiek beschikbaar stellen, omdat dat een gevaar zou kunnen betekenen voor bv. een zeldzame of bedreigde soort, of omdat dat de privacy van mensen zou schenden of mensen zou kunnen schaden. Schenk hier bewust voldoende aandacht aan.





Soms verzamel je 'gevoelige' data: dat betekent dat ze bv. handelen over kwetsbare groepen mensen, kinderen, ouderen, zieken of psychisch kwetsbare mensen. Gevoelige data kunnen ook slaan op bv. de locatie van een bedreigde of zeldzame dier- of plantensoort. In dergelijke gevallen is het dan vaak in het belang van die persoon, of dat organisme, om hun locatie weliswaar wel goed beveiligd intern te bewaren, maar niet vrij te geven aan derden (dus niet Open te stellen). Deze overweging moet bij burgerwetenschap vaak meer worden gemaakt dan bij klassiek wetenschappelijk onderzoek. Ook wanneer een organisatie haar burgerwetenschapsdatabank financieel of technisch niet kan onderhouden zonder inkomsten te vorderen op het vrijgeven van data, kan dat een reden zijn om bepaalde datasets niet zomaar vrij te geven.



Praktijkvoorbeeld

Waarnemingen.be neemt zich voor om de data die de burgerwetenschappers verzamelen, zo breed mogelijk beschikbaar te maken voor algemeen gebruik. Toch worden sommige observaties of multimedia niet open op het platform gepubliceerd, bijvoorbeeld wanneer het gaat om een bedreigde diersoort. Daarmee wil Waarnemingen.be vermijden dat zo'n waarneming een publieke belangstelling zou uitlokken, die een bedreiging vormt voor het waargenomen dier.

Bij Fietsbarometer (Universiteit Gent, zie ook **principe 8**) werden dan weer o.a. fietssegmenten van leerlingen weggelaten die zich bevonden tussen de woonst van de betreffende leerling(e) en het eerstvolgende kruispunt op weg naar school waar de leerling(e) niet meer als enige op de route verscheen.

12. *Neem een respectvolle, gelijkwaardige houding aan tegenover burgerwetenschappers.*

In navolging van, en aansluitend bij, ***principe 11*** ga je bij het opstellen en uitvoeren van je project na of je wel goed weet wat je (toekomstige) deelnemers motiveert. Probeer daar, weliswaar binnen de limieten van je wetenschappelijke context en doelstellingen, zo goed mogelijk aan tegemoet te komen, en respectvol met hen om te gaan. Dit omvat o.a. een volledige, tijdige en heldere communicatie met je deelnemers over het gebruik van hun data (los van de GDPR-regels). Ook leg je een grote zorgvuldigheid aan de dag bij het bedanken of vernoemen van burgerwetenschappers voor hun bijdragen aan je project, bij de publicatie van je resultaten, al dan niet met een co-auteurschap, en/of via andere media.



Citizen-scienceprojecten onderscheiden zich van ander wetenschappelijk onderzoek door de deelname van burgers in een of meerdere fasen van het onderzoek. Burgers nemen typisch deel vanuit (a) een verlangen om bij te leren over een bepaald onderwerp, (b) deel uit te maken van een 'gemeenschap' en een doel te hebben, (c) het gevoel te hebben te kunnen bijdragen aan de wetenschap, en/of (d) bezorgdheid over een bepaalde situatie, en dus de ambitie om te wegen op het beleid. Wie bijdraagt, heeft de logische verwachting achteraf de resultaten te kunnen raadplegen.

Tip

De [Scivil-communicatiegids](#)⁹² voor citizen science behandelt uitgebreid dit thema.



Praktijkvoorbeeld

Het bekendste praktijkvoorbeeld is het benoemen van 'Hanny's Voorwerp', een reflectienevel die in 2007 ontdekt werd door de Nederlandse leerkracht Hanny Van Arkel. Ze vond het, tot dan toe onbekende, astronomische object op een foto die ze analyseerde voor het [Galaxy Zoo](#)⁹³-project. Tot op de dag van vandaag gebruiken astronomen over heel de wereld de naam 'Hanny's Voorwerp' als referentie voor de nevel ([bv. Keel et al., 2012](#))⁹⁴.

Leerkracht Christof Wiedemair, en enkele van zijn leerlingen, waren medeauteur op tal van wetenschappelijke publicaties door zijn bijdragen aan astronomische waarnemingen ([bv. Wiedemair et al., 2020](#))⁹⁵. Ook enkele van zijn leerlingen werden als coauteur opgenomen in relevante publicaties. Ook paleontoloog [Mark Bosselaers](#)⁹⁶ schrijft regelmatig (mee aan) wetenschappelijke publicaties.

Ook wanneer een grotere groep mensen heeft bijgedragen tot je project, kunnen zij nog opgenomen worden, hetzij als coauteur(s), hetzij in de dankbetuiging van het artikel of rapport. [Harlin et al. \(2018\)](#)⁹⁷ schreven bijvoorbeeld een hoofdstuk 'Turning Students into Citizen Scientists' in een boek over burgerwetenschap. De leerlingen van de 'Leysin American School' in Zwitserland werden gezamenlijk als coauteur opgenomen in dit hoofdstuk.

In het project [Stiemerlab](#)⁹⁸ (Stad Genk) worden buurtbewoners van de Stiemervallei zeer nauw betrokken bij verschillende fasen van het onderzoek naar de waterkwaliteit van de Stiemerbeek. Er wordt sterk ingezet op communicatie en gemeenschapsvorming, onder meer door het organiseren van workshops, het uitzenden van regelmatige nieuwsbrieven, het creëren van een aantrekkelijke website waar men de data kan raadplegen, en de nauwe betrokkenheid van het stadsbestuur.

Bij het project [B&Bee](#)⁹⁹, dat het voorkomen van wilde bijen in onze contreien wil bestuderen, wordt materiaal (een bijenhotel en staalnamemateriaal) ter beschikking gesteld van de deelnemers, en wordt ervoor gezorgd dat de wetenschappers vlot bereikbaar zijn bij vragen van deelnemers. Het project [Astrosounds](#)¹⁰⁰ zet dan weer sterk in op boeiende inhoud en heldere uitleg op maat van de deelnemers.

⁹² https://www.scivil.be/sites/default/files/paragraph/files/2019-12/Scivil%20Communicatiegids_v2_0.pdf

⁹³ <https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/>

⁹⁴ <https://iopscience.iop.org/article/10.1088/0004-6256/144/2/66>

⁹⁵ https://www.vub.be/STER/JAD/JAD26/jad26_1/jad26_1.pdf

⁹⁶ <http://www.paleobiologischekring.org/homepage/?p=2064>

⁹⁷ https://www.jstor.org/stable/j.ctv550cf2.35?seq=1#metadata_info_tab_contents

⁹⁸ <https://stiemerlab.be/>

⁹⁹ <https://bijeninuwtuin.weebly.com/bbeeproject.html>

¹⁰⁰ <https://www.astrosounds.be/>



||| ■ Datahygiëne

13. Stel een datamanagementplan (DMP) op.

In een datamanagementplan (DMP) beschrijf je hoe je tijdens en na afloop van een project ervoor zorgt dat alle gegevens die bij een onderzoeksproject aan bod komen een betrouwbare kwaliteit hebben, veilig bewaard worden en maximaal vindbaar, toegankelijk, herbruikbaar en linkbaar gemaakt worden. Beschrijf zorgvuldig hoe je dit in jouw project zal realiseren en gebruik daarvoor gepaste templates die beschikbaar zijn. Een DMP is meer dan louter administratie: het heeft een centrale ondersteunende rol in je project. Duid ook een (data)verantwoordelijke aan die het overzicht bewaart en als data-aanspreekpunt in je project kan fungeren (zie ook [*principe 7*](#)).



Datamanagementplan (DMP)

Een goed datamanagementplan (DMP) helpt je om je project beter te plannen en de kosten van je datamanagement vooraf beter in te schatten. Bovendien komen stukken uit je DMP overeen met wat je nodig hebt om te voldoen aan de Algemene Verordening Gegevensbescherming (AVG of GDPR) en om je privacyverklaring correct op te maken.

Een algemeen eenvormig datamanagement-template maken voor citizen science als geheel is moeilijk, omdat elk plan op maat gemaakt wordt van het betreffende project en onderzoeksdomein. In grote lijnen bevatten DMP's informatie over het project en over de data(sets): welke data worden gegenereerd, welke worden hergebruikt? Wat zijn de (meta)dataformaten? Welke **datastandaarden** (zie **principe 17**) gebruik je? Hoe wordt toegang tot de data georganiseerd? Hoe bescherm je de gegevens tegen accidenteel verlies? Welk budget voorziet je voor datamanagement en -opslag? Hoe zal je de gegevens ten slotte openstellen of delen, ook na het einde van het project?

Tip

De Universiteit Gent heeft een bruikbare, begrijpelijke en volledige webpagina over [Datamanagementplanning](#)¹⁰¹, met links naar tools en lijsten.

Ook op Europees niveau worden [overzichtelijke informatie](#)¹⁰², bijkomende tools en voorbeelden van datamanagementplannen gegeven via het [OpenAIRE](#)¹⁰³ project.

Tip

De UK Environmental Observation Network schreef een toegankelijke [gids](#)¹⁰⁴ over het maken van een datamanagementplan in de context van citizen science. Daarnaast verzorgde de datawerkgroep ook een aantal [aanbevelingsdocumenten](#)¹⁰⁵ rond datamanagement, waaronder een [document](#)¹⁰⁶ met een vereenvoudigd data-stappenplan voor citizen-scienceprojecten.

¹⁰¹ <https://www.ugent.be/nl/onderzoek/datamanagement/datamanagementplan>

¹⁰² <https://www.openaire.eu/how-to-create-a-data-management-plan>

¹⁰³ <https://www.openaire.eu/>

¹⁰⁴ <http://www.ukeof.org.uk/resources/citizen-science-resources/cswq-data-guidance-booklet-web.pdf>

¹⁰⁵ <http://www.ukeof.org.uk/our-work/data-advisory-group>

¹⁰⁶ http://www.ukeof.org.uk/documents/DataAdviceNote2_single.pdf

Tip

Heb je elk principe uit dit charter afgetoetst met jouw project? Dan heb je vermoedelijk al voldoende materiaal om een DMP mee te vullen.

Tip

Maak hergebruik van **bestaande templates**.

Heel wat Vlaamse onderzoeksinstituten¹⁰⁷ gebruiken de online tool 'DMP-Online'¹⁰⁸ voor het opstellen van datamanagementplannen voor onderzoeksprojecten. De tool bevat verschillende templates voor het maken van zo'n plan, met alle relevante vragen die je dient te beantwoorden vóór de start van je project of voordat je er financiering voor aanvraagt. Je kan in DMP-Online een eigen datamanagementplan aanmaken voor intern gebruik, je kan een template van je eigen instituut gebruiken (bv. INBO, UGent) of je kan een template van een financieringsorgaan volgen (bv. FWO-Vlaanderen, Belspo, Europese Commissie) die in de tool vervat zit.

Ben je verbonden aan een instituut dat gebruik maakt van DMP-Online? Dan kan je rechtstreeks inloggen via jouw organisatie. Als externe kan je toegang krijgen via je **Orcid-ID**¹⁰⁹ en (enkel) op uitnodiging van iemand uit een aangesloten instituut meewerken aan een gezamenlijk datamanagementplan.

Is jouw onderzoeksinstituting niet verbonden aan DMP-Online? Ga dan na of het geen eigen DMP-Tool heeft ontwikkeld (bv. KU Leuven: <https://dmponline.kuleuven.be/>).

Is ook dat niet het geval? Geen nood: de meeste financieringsorganen stellen hun template ook apart beschikbaar in bv. een [word-document](#)¹¹⁰ (zie [FWO-Vlaanderen](#)¹¹¹, [Belspo](#)¹¹²) of een [webpagina](#)¹¹³ (bv. Horizon-2020 programma).

Verder kan je je ook nog laten inspireren door de lijst met voorbeelden van datamanagementplannen die openbaar beschikbaar zijn op de oorspronkelijke Britse DMP-Online website: https://dmponline.dcc.ac.uk/public_plans.

¹⁰⁷ De volgende instituten zijn verbonden aan DMP-Online: ILVO, Sciensano, UAntwerpen, UCLouvain, UGent, UHasselt, ULB, ULiège, UMon, UNamur, UZGent, VITO, VLIZ, VUB.

¹⁰⁸ <https://dmponline.be/>

¹⁰⁹ Gratis aan te vragen via <https://orcid.org/> voor (burger)onderzoekers.

¹¹⁰ <https://www.fwo.be/media/1023898/fwo-dmp-templatedocx.docx>

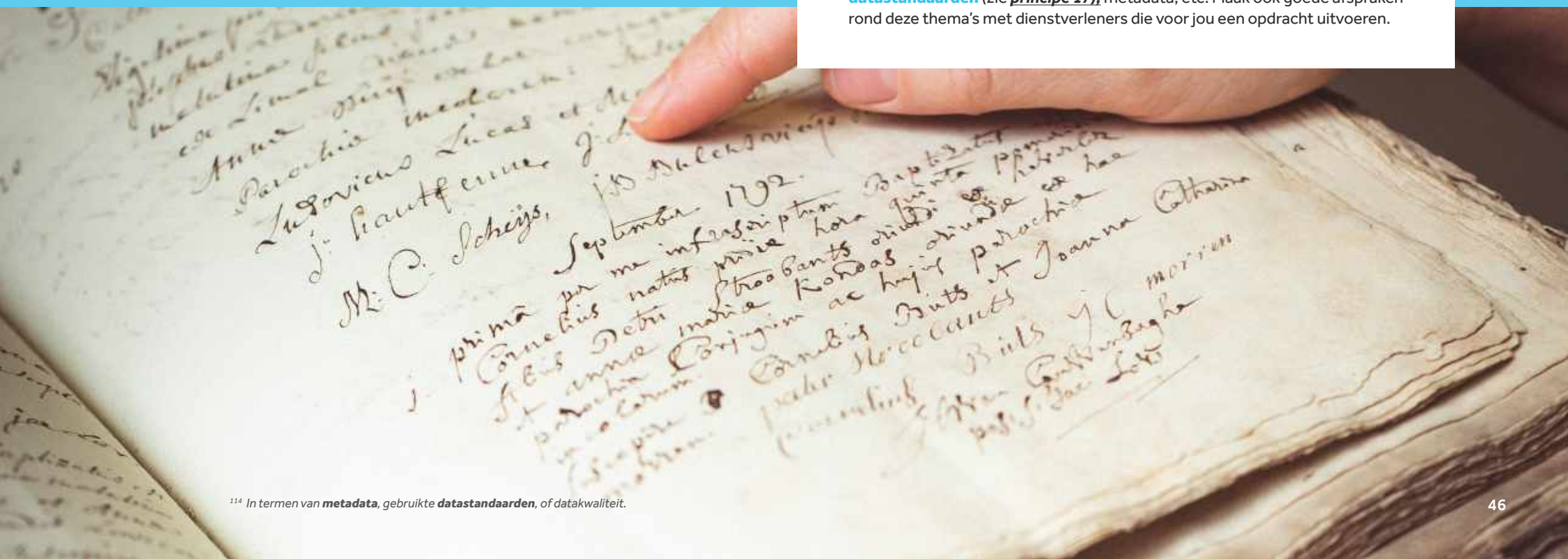
¹¹¹ <https://www.fwo.be/nl/het-fwo/onderzoeksbeleid/data-management-plan/>

¹¹² http://www.belspo.be/belspo/openscience/opendata_DMPtemplate_en.stm

¹¹³ [Data management - H2020 Online Manual \(europa.eu\)](#)

14. Hou rekening met het cumulatiefpotentieel van data: 'Behandel een kleine dataset zoals een grote'.

Bij de start van een **citizen-scienceproject** moet de organisator extra aandacht besteden aan datahygiëne. Het achterliggende idee is: liever vijf goed gedocumenteerde¹¹⁴ elementen, dan 50 slecht gedocumenteerde. Zorg hierbij voor actieve kennisdeling met je burgerwetenschappers en co-creators (bv. co-projectschrijvers) over datamanagement, **datastandaarden** (zie **principe 17**), metadata, etc. Maak ook goede afspraken rond deze thema's met dienstverleners die voor jou een opdracht uitvoeren.



¹¹⁴ In termen van **metadata**, gebruikte **datastandaarden**, of datakwaliteit.

Tip

Zorg voor **opleiding** van en bewustmaking bij je **deelnemers** rond datahygiëne.

Meer dan in eender welk klassiek onderzoek kent citizen science een zeer grote diversiteit qua grootte van datasets of kwaliteit van de metingen of observaties. Net daarom is het extra belangrijk om specifieke aandacht te schenken aan datahygiëne in functie van potentiële interoperabiliteit of integratie in datasets van andere projecten. Het is belangrijk, niet alleen voor het verhogen van de interoperabiliteit van de data zelf, maar ook voor het verhogen van de motivatie en betrokkenheid van je deelnemers, om dit aspect sterk te belichten in de communicatie naar deelnemers toe. Onthoud vooral dat wanneer je slechts enkele meetwaarden verzamelt, die op termijn toch een cruciaal deel kunnen uitmaken van een dataset waar belangrijke wetenschappelijke bevindingen aan gekoppeld kunnen worden.

Tip

Maak goede afspraken met **dienstverleners** bij het uitbesteden van een opdracht.



Praktijkvoorbeeld

Veel biodiversiteitsprojecten, zoals Teek a Break¹¹⁵ (UAntwerpen) of Bugs 2 the Rescue¹¹⁶ (VUB) sluiten aan bij bestaande crowdsourcingplatformen voor biologische waarnemingen, zoals (in dit geval) [waarnemingen.be](https://www.waarnemingen.be). Hiermee erkennen ze het belang van de interoperabiliteit van hun datasets, ook als die in vergelijking met sommige 'grotere' projecten klein lijken.



¹¹⁵ <https://www.uantwerpen.be/nl/projecten/teek-a-break/>

¹¹⁶ <https://www.bugs2therescue.be/>

15. Draag extra zorg voor de kwaliteit van data.

Datakwaliteit vormt een integraal hoofdstuk binnen een datamanagementplan, maar verdient binnen de context van citizen science extra aandacht: de kwaliteit van je data moet voldoen aan wat minimum nodig is voor eigen gebruik in jouw project, mits ze goed beschreven zijn door correcte metadata (zie **thema V**). Om tot een goede datakwaliteit te komen voor jouw project, kies je bewust het profiel van je deelnemers, en maak je hierbij zorgvuldig de onvermijdelijke afweging tegenover de eventueel aanwezige educatieve of maatschappelijke doelstellingen van je citizen-scienceproject. Hou hierbij rekening met de principes uit Thema II. Zoek ten slotte ook naar de meest geschikte balans tussen kwaliteit en kwantiteit van individuele metingen of analyses om tot een optimale kwaliteit van projectresultaten te komen (zie ook **principe 10**).



Kwaliteit versus kwantiteit

Bij burgerwetenschap is er steeds sprake van een trade-off tussen individuele datakwaliteit (resolutie, precisie en juistheid) enerzijds, en het opschalen naar grotere hoeveelheden observaties of metingen (met betere statistische representativiteit) anderzijds.

In grote lijnen kan je inzetten op verschillende aspecten:

- (a) Je zorgt voor een zeer groot aantal staalnames, metingen of waarnemingen door een **groot aantal** deelnemers bij je project te betrekken. Hierbij kan je ofwel inzetten op een maximaal aantal deelnemers, ofwel op een kleinere groep die meerdere observaties of metingen doet. Het voordeel van 'grote aantallen' is dat verschillende deelnemers bepaalde waarnemingen, metingen of analyses kunnen **dupliceren**.
- (b) Je besteedt extra aandacht aan **opleiding** van de burgerwetenschappers, zodat ze nauwkeurigere waarnemingen kunnen uitvoeren of op een meer zelfstandige manier te werk kunnen gaan. Zo kan je voorzien in extra gidsen, protocollen, een uitgebreide website of een lespakket, maar ook in meerdere contactmomenten of extra begeleiding door 'ambassadeurs' uit jouw vakgebied (bv. sterke hobbyisten, goed getrainde amateurs).
- (c) Je zet in op 'revisie': je werkt samen met onderzoekers of met de voorgenoemde 'ambassadeurs' in je studiedomein, zodat zij het werk van de burgerwetenschappers kunnen nakijken en valideren.

In sommige projecten zet men best sterk in op individuele begeleiding en opleiding van burgerwetenschappers, waardoor men nadien aan de burger een grotere zelfstandigheid kan verlenen en toch op betrouwbare resultaten kan rekenen. Andere projecten zijn meer gebaat met eenvoudige protocollen die wijd verspreid worden om op die manier een groter aantal meetpunten te bekomen. Dit hangt, zoals ook aangehaald in **principe 10**, af van je projectdoelen en de context (bv. financieel) waarin je werkt.

Ook statistische correcties achteraf kunnen helpen om de betrouwbaarheid van het geheel aan data te bewaken.

Tip



Het Amerikaanse **Environmental Protection Agency**¹¹⁷ (EPA) publiceerde een uitgebreide **gids en praktijkvoorbeelden**¹¹⁸ rond datakwaliteitsbeheersing bij citizen-scienceprojecten.

¹¹⁷ <https://www.epa.gov/>

¹¹⁸ <https://www.epa.gov/citizen-science/quality-assurance-handbook-and-guidance-documents-citizen-science-projects>

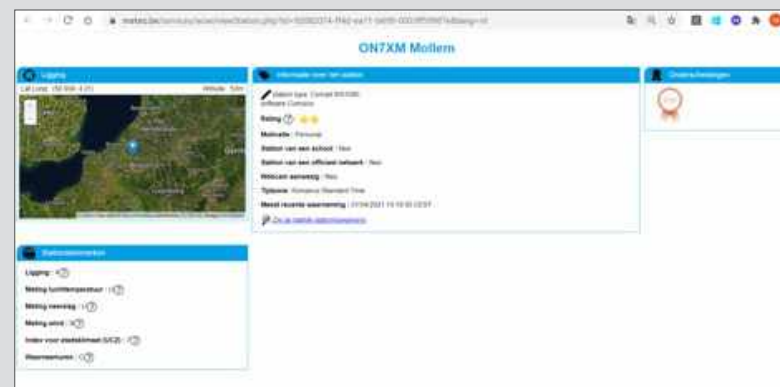


Praktijkvoorbeeld

Bij een project als CurieuzeNeuzen¹¹⁹, waarbij het NOx-gehalte in Vlaanderen in kaart gebracht werd voor de maand mei 2019, mikte men op de kracht van de aantallen: 20 000 deelnemers zorgden voor een bijzonder groot aantal meetpunten. Daartegenover stond dan wel dat er geen gedetailleerde metingen met sensoren mogelijk waren bij al die deelnemers, omwille van de haalbaarheid. Bij het project MamaMito¹²⁰, waarin deelnemers een stamboom langs moeders lijn reconstrueerden, zette men dan weer sterk in op begeleiding van de burgerwetenschappers. Hiervoor werd samengewerkt met Familiekunde Vlaanderen, het Davidsfonds, en enkele heemkundige kringen. In het VLINDER¹²¹-project (Vlaanderen in de Weer) plaatste men dan weer 60 hoge kwaliteit weerstations op weloverwogen plaatsen in Vlaanderen. Validatie van data die worden ingegeven op het platform DoeDat¹²² van de Plantentuin Meise, gebeurt door de wetenschappers zelf. Zij ervaren deze manier van werken nog steeds als efficiënter dan wanneer zij zonder de hulp van burgers hun waarnemingen doen.

Het internationale weerplatform WOW¹²³ gebruikt **metadata** van weerstations om een kwaliteitslabel aan elk station te koppelen.

Burgerwetenschappers die gegevens van hun weerstation naar het platform doorsturen, kunnen daar ook de specificaties en omgevingsfactoren van hun station ingeven, waarna het webplatform automatisch een ster-rating toekent.



¹¹⁹ <https://2018.curieuzeneuzen.be/>

¹²⁰ <https://historiesvzw.be/projecten/mamamito/>

¹²¹ <https://vlinder.ugent.be/>

¹²² <https://www.doedat.be/>

¹²³ <https://wow.meteo.be/nl/>

16. Stel een conceptueel datamodel op voor jouw project.

Een conceptueel **datamodel** (ook informatiemodel of domeinmodel genoemd) is een manier om de informatie of gegevens die verzameld of hergebruikt worden in een project, eenduidig en logisch te beschrijven. Dit is nodig om zowel binnen het project als daarbuiten over de betekenis van de data te kunnen communiceren, en helpt de fysieke opslag en publicatie van gegevens voor te bereiden. Het laat ook toe om verder te bouwen op bestaande **datastandaarden** (zie **principe 17**), wat sterk aanbevolen wordt¹²⁴. Merk op dat een conceptueel datamodel niets zegt over de technische implementatie in databanken, applicaties en systemen. Je hoeft dus geen IT-kennis te hebben om zo'n model op te stellen.

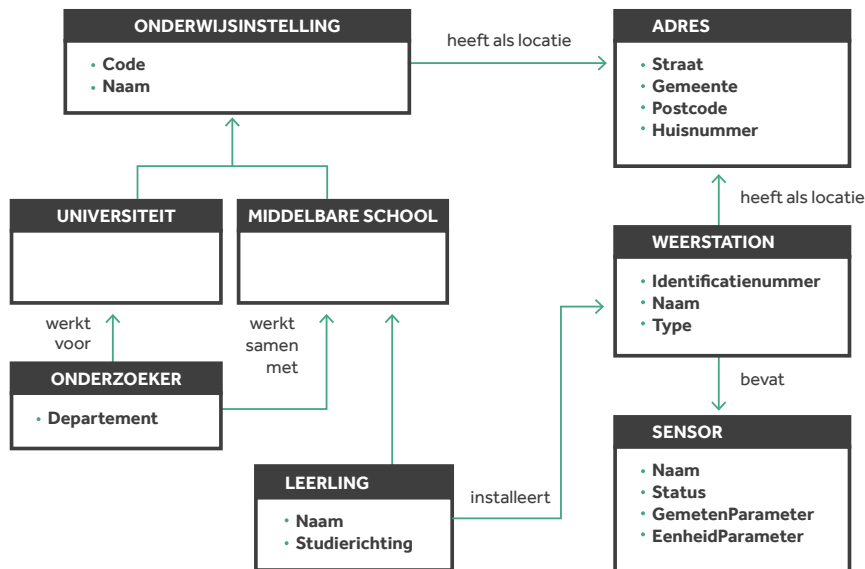
¹²⁴ Merk op: je kan een datamodel opstellen zonder enige datastandaard te volgen, enkel met de bedoeling om je eigen gegevens te beschrijven. Daarom koppelen we [principe 16](#) los van [principe 17](#).

Beschrijf minstens de entiteiten, attributen en relaties

Eender welke verzameling informatie (of dus: data) in de wereld kan je beschrijven aan de hand van drie elementen:

- (a) **entiteiten** [= reële objecten of concepten, dingen met betekenis],
- (b) **attributen** [= de eigenschappen van die entiteiten], en
- (c) **relaties** [= wat het verband is tussen die entiteiten].

Met deze drie elementen kan je eenvoudig een 'conceptueel datamodel' uittekenen dat een bepaalde set aan gegevens beschrijft (zie Figuur 4). Nadien kan je nog extra nuances toevoegen aan zo'n schema.



Figuur 4

Datamodelering toegepast op het domein van het VLINDER-burgerwetenschapsproject: elk rechthoekje beschrijft een entiteit (grijs) met daarbinnen eventueel attributen. De verbindingslijnen duiden relaties tussen entiteiten aan. De holle pijlen wijzen op een 'is een'-relatie: "Een Universiteit is een type Onderwijsinstelling, en erft daar dus de attributen (eigenschappen) van over".

Zorg voor heldere definities

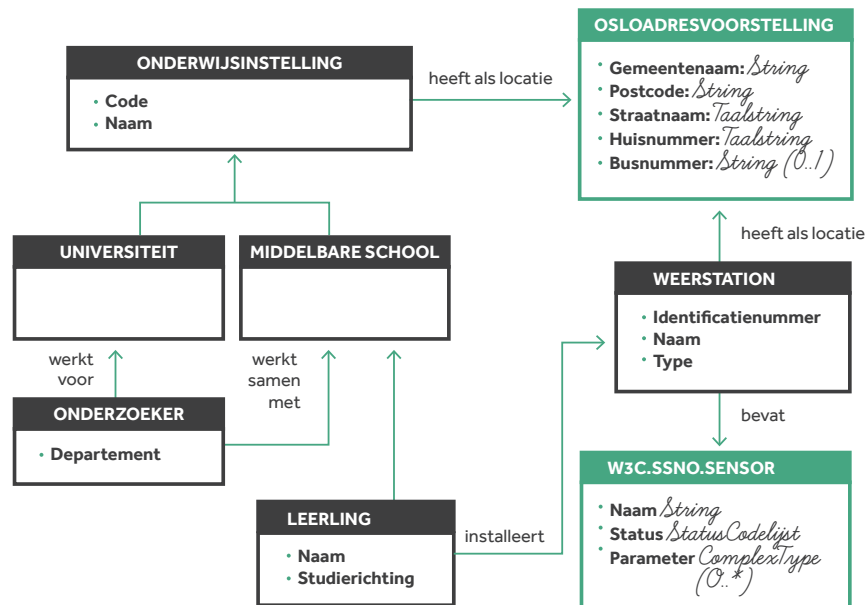
Entiteiten, attributen en relaties definieer je op een ondubbelzinnige manier, zodat er geen misverstanden kunnen ontstaan over de betekenis van de verzamelde gegevens. Wanneer dergelijke definities worden vastgelegd en erkend door een brede groep domeinexperten, worden ze samen een **vocabularium** genoemd, en kunnen ze deel uitmaken van een **datastandaard**.

Datastandaarden

De eerste reflex moet zijn om in je model bestaande datastandaarden (zie ook [principe 17](#)) te gebruiken, en waar nodig deze standaarden uit te breiden.

Het heeft geen zin om veel tijd en energie te investeren in het beschrijven van entiteiten, attributen en relaties waarvoor reeds **datastandaarden** bestaan. Of je nu locaties wilt uitmodelleren of de waarneming van diersoorten: de kans is zeer groot vdat anderen dit al eerder gedaan hebben. Door bestaande standaarden te volgen win je niet alleen tijd, maar kan je je data ook meteen gaan uitwisselen volgens die standaard, wat samenwerking en **interoperabiliteit** ten goede komt.

Tracht ook zoveel mogelijk reeds bestaande datamodellen te hergebruiken en te combineren, om daar dan nadien enkel nog die entiteiten, attributen en relaties aan toe te voegen die nodig zijn om de data uit jouw project te beschrijven.



Figuur 4

Zelfde datamodel als in de vorige figuur, maar nu gebruiken we voor de twee blauwe entiteiten bestaande datastandaarden. Uit het OSLO-applicatieprofiel 'adres' werd 'adresvoorstelling' hergebruikt, en er werd ook gelinkt naar de beschrijving van de entiteit 'sensor' uit de W3C-SSNO-standaard.

UML-klassendiagram

Maak gebruik van een UML-klassendiagram

Hoewel er verschillende modelleertalen bestaan die geschikt zijn voor conceptuele datamodellering (bv. **ERD** en **ORM**), zijn 'UML Class diagrams' de meest aangewezen manier om conceptuele datamodellen op te stellen, en wel om drie redenen:

1. 'UML Class' is ruimer dan ERD en ORM en is bovendien een ISO-standaard.
2. De instapdrempel voor een beginner is laag (gebruik van entiteiten, attributen en relaties volstaat), terwijl gevorderden en experts over alle mogelijkheden beschikken om nuances toe te voegen of om de datamodellen automatisch om te zetten in code of documentatie.
3. Elk zichzelf respecterend softwarepakket voor informatiemodellering ondersteunt de UML-modelleertaal. UML is dan ook zeer gangbaar in de wereld van de informatiemodellering en datastandaarden.

Tip



Overzicht van de bouwstenen van een UML-klassendiagram: <https://www.uml-diagrams.org/class-diagrams-overview.html>.



Videotutorial over UML-klassendiagram (10 minuten): <https://youtu.be/UI6lqHOVHic>.



Gratis, online tekenomgeving waarin je direct kunt gaan modelleren: <https://www.draw.io/>.



IV ■ Datastandaarden en -formaten

17. *Bouw verder op bestaande en voor jouw project relevante datastandaarden.*

Datastandaarden leggen de betekenis en structuur van uitgewisselde informatie vast in **vocabularia** (met definities), datamodellen (met entiteiten, attributen en relaties) en **codelijsten** (met uniform hanteerbare codes). Ze zijn vaak domeinspecifiek, en worden door domeinexperts onderling afgesproken. Het gebruik van datastandaarden is cruciaal om te komen tot '**semantische interoperabiliteit**'. Wanneer verschillende (citizen-science)projecten gebruik maken van dezelfde datastandaarden, vereenvoudigt dat de informatie-uitwisseling en kennisdeling. Gebruik zoveel mogelijk de bestaande datastandaarden uit jouw onderzoeksdomein.

Een oplistings maken van alle relevante datastandaarden is onbegonnen werk, in de eerste plaats omdat citizen-scienceprojecten zich in verschillende domeinen bewegen (luchtkwaliteit, biodiversiteit, mobiliteit, gezondheid, taalkunde,...) en in de tweede plaats omdat in bepaalde specifieke onderzoeksdomeinen dergelijke standaarden nog in ontwikkeling zijn. We spreken hier over **domeinspecifieke standaarden**, waarvan de meest bekende de **TDWG**-standaarden en de **INSPIRE** dataspecificaties zijn (zie onder). Specifiek voor citizen science, worden er per onderzoeksdomein meer passende afgeleide standaarden ontwikkeld door werkgroepen van experts uit elk domein (zie '**praktijkvoorbeelden**').

TDWG-standaarden

<https://www.tdwg.org/>

Biodiversiteitsstandaarden, waarvan de 'Darwin Core'¹²⁵ (met entiteiten 'organisme', 'specimen', 'taxon' en hun attributen) allicht de bekendste is. Onder meer gehanteerd door het GBIF-platform. Het taxonomisch register *Aphia*¹²⁶ bevat een oplistings van alle gekende mariene soorten in de wereld en biedt een overzicht van de aanvaarde taxonomische namen (incl. synoniemen). Dit register wordt wereldwijd gebruikt als standaardlijst.

INSPIRE data specificaties

<https://inspire.ec.europa.eu/data-specifications/>

Deze standaarden werden uitgewerkt met het oog op het delen van ruimtelijke gegevens tussen Europese lidstaten, ter ondersteuning van milieubeleid. Deze Engelstalige dataspecificaties (met UML-modellen, definities en codelijsten) strekken zich uit over uiteenlopende, milieugerelateerde domeinen (bv. 'atmospheric conditions', 'habitats and biotopes', 'species distribution', 'human health and safety'). De toepassing ervan wordt ondersteund door uitvoerige documentatie, toolkits en een internationale community. Deze standaarden worden o.a. gebruikt in de geoviewer van het *kustportaal*¹²⁷ (Vlaams Instituut voor de Zee).

¹²⁵ <https://www.tdwg.org/standards/dwc/>

¹²⁶ <http://www.vliz.be/nl/taxonomic-register-marine-species>

¹²⁷ <https://www.kustportaal.be/nl>

¹²⁸ Zie <https://www.iso.org/standard/36760.html>. Deze wordt o.m. gehanteerd door Digitaal Vlaanderen bij het publiceren van geografische raster- en vectordata.

¹²⁹ Zie <https://www.iso.org/iso-639-language-codes.html>. Worden bijvoorbeeld gebruikt in websiteadressen, maar kunnen ook voor andere toepassingen een meerwaarde betekenen.

¹³⁰ <https://www.w3.org/TR/vocab-ssn/>

Daarnaast bestaan er ook standaarden die relevant zijn ongeacht het domein, bijvoorbeeld omdat ze **basis- of locatiedata** beschrijven waaraan domeinspecifieke gegevens kunnen gekoppeld worden, of omdat ze **generieker (domeinoverschrijdend)** zijn opgevat. Hierbij een niet-exhaustief overzicht van enkele interessante datastandaardportalen:

ISO-standaarden

<https://www.iso.org/standards.html>

Deze internationale standaarden kunnen interessant zijn voor een citizen-scienceproject. Zo is de ISO 19131-standaard¹²⁸ uitermate geschikt voor het beschrijven van geografische data, en kan er gebruikt gemaakt worden van de taalcodes in ISO 639¹²⁹ voor het beschrijven van talen of taalfamilies.

W3C-standaarden

<https://www.w3.org/TR/?tag=data>

Het W3C standaardiseert het internet wereldwijd. Van daaruit ontwikkelde men ook aanbevelingen en vocabularia die interessant kunnen zijn voor citizen-scienceprojecten. Voorbeelden zijn de 'Semantic Sensor Network Ontology'¹³⁰ (beschrijft sensoren en hun observaties), de 'Data Catalog Vocabulary'¹³¹ (DCAT, geschikt voor metadata) en 'Best Practices for Publishing Linked Data'.¹³²

OSLO-datastandaarden

<https://data.vlaanderen.be/>

De OSLO-standaarden zijn compatibel met internationale standaarden (ISO, INSPIRE, ISA, ¹³³...) maar zijn eenvoudiger toe te passen omdat ze door Vlaamse domeinexperts zijn opgesteld in het Nederlands. Momenteel zijn er 94 erkende standaarden¹³⁴, en 42 kandidaat-standaarden. De OSLO-standaarden bevatten vocabularia, maar ook 'applicatieprofielen'. Standaarden die betrekking hebben op basisdata, zoals 'adressen', 'gebouwen', 'wegen', 'openbaar domein', 'personen' en 'contactvoorkeuren' kunnen bruikbaar zijn voor veel citizen-scienceprojecten bij ons, net als standaarden die zich op een specifiek (wetenschappelijk/levens/beleids)domein focussen, zoals 'bodem en ondergrond' of 'lucht en water'.

¹³¹ <https://w3c.github.io/dxwg/dcat/>

¹³² <https://www.w3.org/TR/ld-bp/>

¹³³ https://ec.europa.eu/isa2/solutions/core-vocabularies_en

¹³⁴ <https://data.vlaanderen.be/standaarden/#erkendeStandaarden>

PPSR-core-datastandaarden

<https://core.citizenscience.org/docs/observation>

De **Werkgroep Data & Metadata**¹³⁵ van de Citizen Science Association (CSA) werkt sinds geruime tijd aan standaarden voor data uit projecten met publieke participatie. De **PPSR**-core bestaat uit een 'Observation Data Model' (ODM) en een 'Common Data Model' (CDM). Daarnaast werden ook twee PPSR-metadatamodellen opgesteld (resp. voor datasets, en voor projecten; zie **principe 24**).

In het **Observation Data Model** (ODM) is de entiteit 'een observatie', en zijn de attributen allerlei zaken die gezien, gehoord, opgemerkt of aangegeven worden door een burgerwetenschapper die deze 'observatie' uitvoert. Het ODM beschrijft verschillende soorten 'observaties'. Een reeks van die 'observaties' samen vormt een dataset. Soms kan je attributen uit een bepaald domein ook gebruiken voor een ander domein, en dat is goed, want dan kunnen datasets uit verschillende domeinen makkelijker met elkaar gecombineerd worden. Op de website van

PPSR-ODM¹³⁶ vind je een lijst met domeinen waarin reeds standaarden zijn afgesproken; de meest bekende zijn de **Darwin core** (voor biodiversiteit) en **Ecological Metadata Language**¹³⁷ (EML) voor ecologisch onderzoek.

Het **Common Data Model** (CDM) is een **PPSR**-overkoepelend model dat de andere data- en metadatamodellen van PPSR aan elkaar rijgt. Dit model geeft een context aan groepen projecten, die bijvoorbeeld binnen een bepaald kaderprogramma of een bepaalde campagne aan elkaar gelinkt kunnen worden.

Tip

Als je beleidsimpact wil creëren met je citizen-scienceproject, pols dan eerst bij de relevante beleidsinstanties aan welke voorwaarden/standaarden je data moeten voldoen om makkelijk met beleidsdata geïntegreerd te kunnen worden. Op die manier vergroot je de mogelijke impact van je data en van je project.

¹³⁵ <https://www.citizenscience.org/get-involved/working-groups/data-and-metadata-working-group/>

¹³⁶ <https://core.citizenscience.org/docs/observation/>

¹³⁷ <https://eml.ecoinformatics.org/>

¹³⁸ <https://github.com/tdwg/citizen-science>

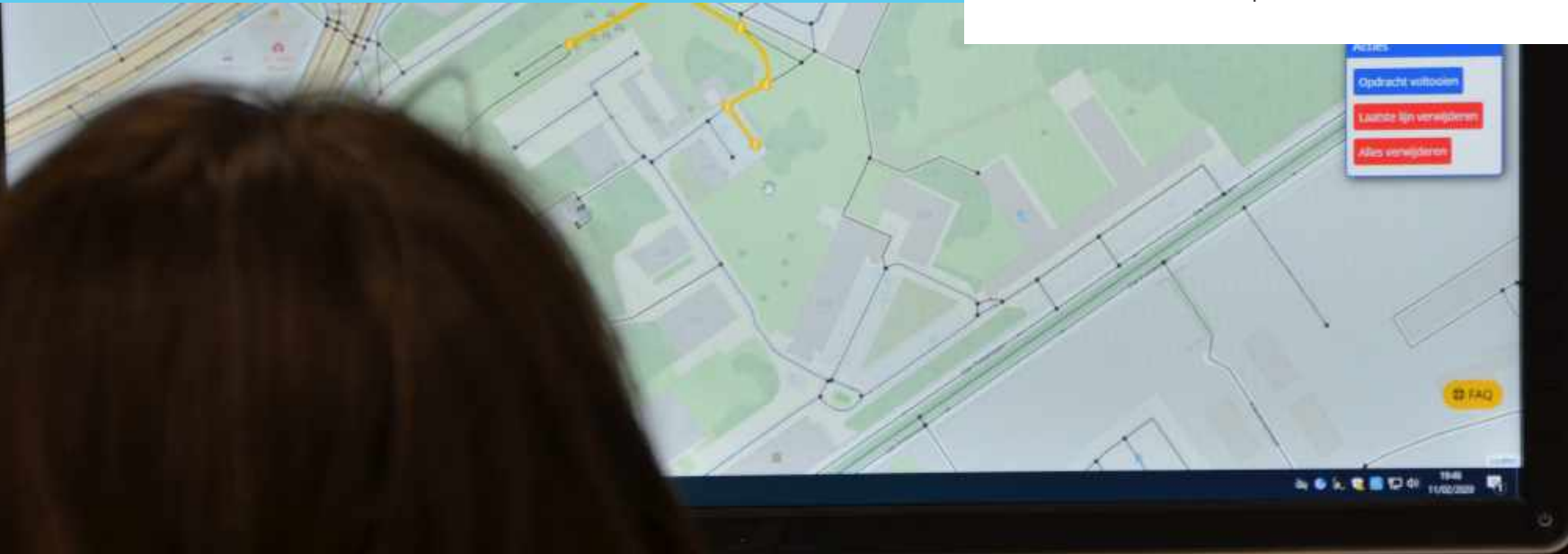
**Praktijkvoorbeeld**

Binnen het domein van Biodiversiteit (TDWG) buigt de Werkgroep Citizen Science zich over het toepassen van de Darwin core standaard in citizen-scienceprojecten. Meer informatie hierover vind je op de **Github**¹³⁸-pagina van deze werkgroep, waar onder andere vertegenwoordigers van het Instituut voor Natuur en Bosonderzoek (INBO) in zetelen.



18. Maak gebruik van machineleesbare, open formaten.

Door gegevens aan te bieden in open (d.w.z. specificaties niet geheim gehouden of in licentie gegeven door één organisatie¹³⁹, maar vrijgegeven en/of gestandaardiseerd) en machineleesbare (d.w.z. gestructureerde) formaten kom je een stap dichterbij 'technische interoperabiliteit'. Je gegevens zullen door méér softwareapplicaties ingelezen en verwerkt kunnen worden. Je vermijdt hiermee ook het risico op 'vendor lock-in'.



¹³⁹ Merk op dat een formaat open gedocumenteerd kan zijn, maar nog steeds propriëitair wegens onderhevig aan licenties.

Tip

Besteed, voordat je je data publiceert, voldoende aandacht aan het kiezen van het meest geschikte bestandsformaat. Dit formaat moet niet alleen de beoogde toepassingen ondersteunen, maar dus ook hergebruik stimuleren door machineleesbaar en open te zijn.

Hieronder enkele voorbeelden van bestandsformaten¹⁴⁰ die je beter vervangt door een 'open' alternatief:

Propriëtair formaat	Voorbeeld van 'open' alternatief
.doc (tekst met opmaak)	.odt / .docx
.xls (data in tabelvorm)	.ods / .csv / .xlsx
.dwg (CAD-tekeningen)	.dxf
.shp (geografische vectordata)	.gml
.bmp (afbeeldingen)	.tiff
.wma (geluid)	.mp3 / .flac
PHP of Java dataserialisatie	.json / .xml



Praktijkvoorbeeld

De projecten VLINDER¹⁴¹ & Leuven.Cool¹⁴², die hun data aanbieden via het WOW-be¹⁴³ platform, doen dat niet alleen in grafieken (figuren) en xls-formaat, maar ook in .csv. De API die toegang geeft tot de data van VLINDER¹⁴⁴ wordt ook open en beschikbaar gesteld in json-formaat, via Github.

¹⁴⁰ Meer info over deze bestandsformaten vind je via een zoekopdracht op <https://docs.fileformat.com>

¹⁴¹ <https://vlinder.ugent.be/Metingen.html>

¹⁴² <https://leuven.cool/>

¹⁴³ <https://wow.meteo.be/nl/>

¹⁴⁴ https://app.swaggerhub.com/apis-docs/bmesuere/VLINDER/1.0#/default/get_



19. Ken globaal unieke en persistente identificatoren toe aan je data.

Door het toekennen van een unieke identificatiecode aan elk concept of elke meting in je dataset, maak je communicatie over (de betekenis van) data, zowel tussen mensen als tussen machines, een pak eenduidiger en makkelijker. Kies voor **Uniform Resource Identifiers (URI's)** als globaal (letterlijk: wereldwijd) unieke identificatoren. Deze 'webadressen' dienen vooral om dataconcepten of -objecten uniek aan te duiden, en in de tweede plaats om er informatie over te raadplegen¹⁴⁵.

Een gepubliceerde **URI** mag niet veranderen en moet steeds naar het object in kwestie blijven verwijzen. Enkel op die manier kunnen stabiele verwijzingen naar of tussen objecten worden gecreëerd, en kan men komen tot een web van 'Linked Open Data', waarbij de data verspreid zijn (decentraal) maar toch gecombineerd kunnen worden.

¹⁴⁵ Enkel wanneer een URI 'resolvable' is, d.w.z. niet enkel een object identificeert maar ook een voorstelling daarvan kan teruggeven via het internet, kan je spreken over een URL (Uniform Resource Locator). Elke URL is een URI, omgekeerd niet.

URI's

Hoe kan je URI's aanbieden om naar jouw projectdata te verwijzen?

Laat ons waarnemingen.be als (fictief) voorbeeld nemen om stap voor stap een URI aan te bieden. Op deze website kunnen vrijwilligers in België en Nederland waarnemingen van plant- en diersoorten registreren.

1. Bepaal welke gegevens (beelden, metingen, locaties, geannoteerde documenten, alfanumerieke gegevens, enz) volledig door jouw project beheerd worden. Of dus: welke data zoeken gebruikers bij jou, die ze nergens anders kunnen vinden? Raadpleeg hiervoor je domeinmodel (zie **principe 16**).

→ **Toegepast:** soortbeschrijvingen, observaties, waarnemers, foto's, geluiden ...

2. Denk na over het **domein waaronder je je data wil aanbieden**. Aangezien de domeinnaam een onderdeel zal vormen van de URI's van je projectdata en gebruikte dataconcepten, moet je absolute garanties hebben dat je dit domein kunt 'opeisen', **stabiel** kunt houden en zelf beheren.

Je kan er ook voor opteren om je gegevens op een online platform aan te bieden mits dit persistente URI's aanbiedt voor de opgeladen data. In dat geval hoef je geen eigenaar te zijn van de domeinnaam, en kan je het publiceren van jouw projectdata met URI's (of bijvoorbeeld als **RDF triples**) overlaten aan dit (internationale) platform. Ook door de Vlaamse overheid worden initiatieven¹⁴⁶ opgezet om het ontsluiten van (sensor)data te faciliteren.

Een derde mogelijkheid is gebruik te maken van online services¹⁴⁷ die persistente URL's (PURL) aanbieden, en deze achterliggend laten doorverwijzen naar je eigen service-URL's.

→ **Toegepast:** het domein 'waarnemingen.be' is reeds toegewezen aan **Natuurpunt**¹⁴⁸ en bijgevolg in eigen beheer, dus is een mogelijke kandidaat voor gebruik in URI's.

3. Bepaal het **patroon** waaraan jouw URI's moeten voldoen. Dit is met name zinvol wanneer je aan de hand van de URI-structuur wil zien of het gaat om ...

- een eigen definitie van een dataconcept (object, attribuut of relatie) – bv. /ns;
- de identificatie van een dataobject – bv. /id;
- een voorstelling van het dataobject – bv. /doc.

Ter inspiratie voor je eigen patroon kan de 'Vlaamse URI-standaard voor data'¹⁴⁹ gebruikt worden, die bindende regels op dat vlak oplegt aan Vlaamse overheden.

→ **Toegepast:** men kiest ervoor om <https://waarnemingen.be/id/{soort,observatie,foto...}/{numeriekveld}> als patroon te gebruiken voor de identificatie van soorten, observaties, foto's ... Men zal ervoor zorgen dat deze URI's resolvable zijn, door bij het aanroepen van deze URI's in een browser een bijhorende webpagina met meer informatie te tonen (onder dezelfde URI).

4. Voorzie niet enkel URI's voor je data(punten), maar ook voor de concepten achter je data (de entiteiten, attributen en relaties uit je domeinmodel), zodat gebruikers en systemen via die URI een (machine)leesbare definitie kunnen terugvinden.

Merk op: deze stap is iets minder evident dan het aanbieden van URI's ter identificatie van datapunten, maar is een voorwaarde om te komen tot 4-sterren open data en de ontsluiting van je data via het **RDF**-formaat.

Je kan deze definities zelf opmaken en onder een URI op je eigen domein aanbieden, maar je kan ook hergebruik maken van bestaande **LOD**-databanken en **ontologieën** online (bv. DBPedia, Wikidata, Geonames, Darwin Core¹⁵⁰ ...) om URI's te verkrijgen voor definities van dataconcepten.

→ **Toegepast:** men kiest ervoor om geen eigen definities aan te maken en te ontsluiten via URI's, maar hergebruik te maken van de Extensible Observation Ontology¹⁵¹. Daarin wordt o.m. een observatie met al zijn eigenschappen reeds gedefinieerd. De URI voor de definitie van 'observatie' is bijvoorbeeld <http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#Observation>.

¹⁴⁶ Context broker/data pipeline, sensordataplatform e.a.

¹⁴⁷ Bijvoorbeeld <http://identifiers.org>, <http://www.purlz.org/> en <https://archive.org/services/purl/>.

¹⁴⁸ Zie <https://www.dnsbelgium.be/nl/whois/info/waarnemingen.be/details>.

¹⁴⁹ https://data.vlaanderen.be/cms/VlaamseURI-StandaardVoorData_V1.0.pdf

¹⁵⁰ Zie <https://dwc.tdwg.org/text/> voor een toepassing van Darwin Core in tekstgebaseerde bestanden zoals CSV.

¹⁵¹ <https://biportal.bioontology.org/ontologies/OBOE>

5. Gebruik de persistente URI's in de plaats van de klassieke, (alfa)numerieke identificatoren in al je ontsluitingsvormen (zowel in datasets, webservices als op websites). Zo worden gebruikers gestimuleerd om van de URI's gebruik te maken om naar jouw data te verwijzen. Achter veldnamen kan je op dezelfde manier de URI plaatsen die naar de definitie van het dataconcept verwijst.

→ **Toegepast:** op de website 'waarnemingen.be' worden observaties onder de volgende vorm aangeboden:

Observatie	Vogelsoort	Locatie
https://waarnemingen.be/id/observatie/1	Kraanvogel – <i>Grus grus</i>	Westkapelle
https://waarnemingen.be/id/observatie/2	Parelduiker – <i>Gavia arctica</i>	Oostende
https://waarnemingen.be/id/observatie/3	IJsduiker – <i>Gavia immer</i>	Willebroek

6. Voor gevorderden: bied je projectdata in **RDF**-, **Turtle**- of **JSON-LD**¹⁵²-formaat aan. RDF is de native taal van het semantische web. Net zoals een databank met SQL bevroegd kan worden, kunnen RDF-bronnen met SPARQL doorzocht worden.

¹⁵² Zie <https://www.w3.org/RDF/> resp. <https://json-ld.org/>.



Praktijkvoorbeeld

Gebruik van URI's om naar dataconcepten en -objecten te verwijzen

<https://orcid.org/0000-0001-8888-635X>

De 'ORCID iD' is een uniek nummer dat door **onderzoekers** kan aangevraagd worden. Deze URI kan bijvoorbeeld vermeld worden bij alle wetenschappelijke publicaties van die persoon. Zo kan je nooit twee researchers met dezelfde voor- en achternaam met elkaar verwarren wanneer je hun lijst van publicaties wil opvragen.

<https://www.uniprot.org/uniprot/P98161>

Deze URI identificeert een menselijk **proteïne** in de UniProt-databank.

http://dbpedia.org/resource/Julius_Caesar

Deze URI identificeert het **onderwerp** 'Julius Caesar' in de DBpedia Knowledge Base.

<https://data.vlaanderen.be/id/adres/983512>

Deze URI identificeert het **adres** van Scivil in het Vlaamse adressenregister.

<https://data.vlaanderen.be/ns/weg#EuropeseWeg>

Deze URI identificeert de **definitie** van een Europese weg in de context van het Vlaamse wegenregister.



20. *Incorporeer je projectdata in het netwerk van Linked Open Data.*

Heb je je projectdata van globaal unieke, persistente identificatoren voorzien? Dan kan je je data extra waardevol maken door ze te koppelen met andere informatiebronnen op het web. Zo kan je, vaak zonder veel extra moeite, rijkere datasets aanbieden, waarin de relatie gelegd wordt tussen de door jou verzamelde projectdata en andere (wetenschappelijke) bronnen of naslagwerken. Het zorgt er uiteindelijk voor dat gebruikers jouw data binnen een bredere context kunnen gaan verkennen.

Linked Open Data

Hoe kan je jouw projectdata deel laten uitmaken van het netwerk van Linked Open Data?

Verder bouwend op het voorbeeld van waarnemingen.be:

- Onderzoek welke bijkomende informatie een verrijking zou vormen voor je eigen dataset.
 - **Toegepast:** men zou graag meer context voorzien bij de locaties waarop waarnemingen werden gedaan, zodat later gerichte, geografische bevragingen kunnen gedaan worden. Daarnaast wil men vogelsoorten linken met de informatie daarover op Wikipedia.
- Maak opnieuw hergebruik van bestaande **LOD**-databanken (zie <https://lod-cloud.net/>) om op zoek te gaan naar bronnen die de gezochte informatie aanbieden met inbegrip van persistente URI's. Deze eigen data worden vervolgens gekoppeld met deze URI's.
 - **Toegepast:** voor de locaties zal men gebruik maken van de officiële gemeenten in het Vlaamse adressenregister en de Nederlandse **BAG** (GeoNames werd ook in overweging genomen); voor de Wikipedia-inhoud kan men terecht op Wikidata. De URI's die door deze bronnen worden aangeboden worden gebruikt om de eigen data op de website te verrijken.

Observatie	Vogelsoort	Locatie
1	Kraanvogel – <i>Grus grus</i>	Westkapelle
2	Parelduiker – <i>Gavia arctica</i>	Oostende
3	IJsduiker – <i>Gavia immer</i>	Willebroek

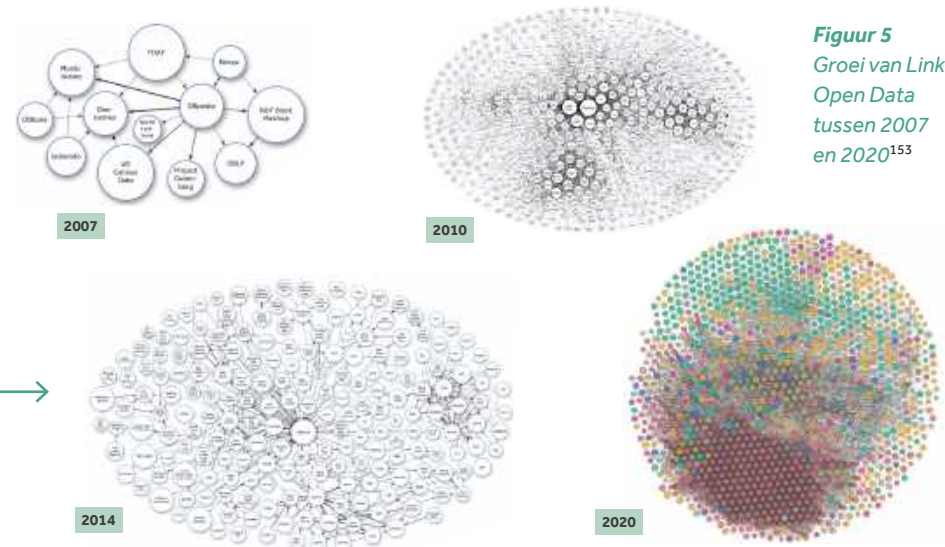
- Aangezien je nu naar externe bronnen verwijst, die niet onder eigen beheer staan, is het belangrijk een **beleid** op te stellen voor het **omgaan met deze externe informatie**, en ook met gebroken of foutieve hyperlinks. Zo kan je bekijken of foutieve informatie aan de bron gemeld kan worden, of een tool gebruiken die op zoek gaat naar 'broken links'. Wees ook zorgvuldig in de bronnen waarmee je linkt: welke kwaliteit moeten deze minimaal hebben?



Praktijkvoorbeeld

Omzetten van dataset naar Turtle en koppeling aan DBpedia-elementen

In het kader van het project 'Nieuws van de Grote Oorlog' van Het Archief werd een script gemaakt (<https://github.com/viaacode/hetar-chief2lod>) om de metadata van de gearchiveerde kranten geschikt te maken voor bevraging op het semantische web, namelijk door deze metadata als Turtle-**triples** aan te bieden. Via automatische karakterherkenning in combinatie met DBpedia **Spotlight**¹⁵⁴ werd tekst uit foto's gekoppeld aan elementen uit DBpedia. Hiermee werd de eigen LOD-dataset gelinkt aan een andere LOD-bron, waardoor bevragingen over deze beide bronnen heen mogelijk worden. Lees meer op <https://meemoo.be/nl/publicaties/historisch-krantenarchief-gepubliceerd-als-linked-open-data>. Hoewel dit geen citizen-scienceproject is, leent dit project zich, zeker in combinatie met mogelijke andere gekoppelde data, perfect voor citizen science.



Figuur 5
Groei van Linked Open Data tussen 2007 en 2020¹⁵³

¹⁵³ Bron: <https://lod-cloud.net/#diagram> Licentie: CC-BY. Check ook Frank & Zander (2019) voor meer informatie.

¹⁵⁴ Zie <https://www.dbpedia-spotlight.org/>



V. Metadata

21. Voorzie je data van zo rijk en accuraat mogelijke metadata.

Metadata betekent 'data over data'. Ze vormen de 'bijsluiter' bij (gepubliceerde) onderzoeksdata en -resultaten en zijn een cruciaal gegeven om je data vindbaar te maken. Hoe rijker de data beschreven worden d.m.v. metadata, hoe makkelijker een potentiële gebruiker kan nagaan wat de herkomst, kwaliteit en 'scope' is van de dataset. Het laat de gebruiker ook toe te beoordelen of de data geschikt zijn voor zijn/haar beoogde doel, en om contact op te nemen met de uitgever ervan. Aangezien datasets aan verandering onderhevig zijn, moeten ook de bijhorende metadata accuraat en up-to-date worden gehouden.



Gemeenschappelijke kern van basisinformatie

- De meeste metadatastandaarden delen een **gemeenschappelijke kern** van basisinformatie die, ongeacht het domein waarin het citizen-scienceproject zich bevindt, moeten meegenomen worden bij het opstellen van metadata voor je dataset:
- Identificator van de beschreven dataset**
Gewoonlijk zijn metadata, en de dataset die ze beschrijven, aparte bestanden. De link tussen die twee moet expliciet gemaakt worden door in de metadata de (globaal) unieke identificatiecode te vermelden van de dataset. Dit kan bijvoorbeeld een identificator zijn die automatisch gegenereerd wordt bij het beschrijven van je dataset in een (meta)dataportaal. Merk op: daar waar we in **principe 19** spraken over persistente identificatoren voor individuele concepten of metingen in een dataset, wordt hier de identificator voor de gehele dataset bedoeld. Dit kan een **GUID** zijn of een **URI**.
- Titel, beschrijving, thema's en trefwoorden**
Deze velden laten toe jouw dataset terug te vinden op portalen waarop honderden of duizenden datasets beschreven staan. Het is dus raadzaam hier de nodige aandacht aan te besteden.

Tip

Zorg ervoor dat de titel identificerend is over alle dataset(serie)s en hun versies heen, dat deze leesbaar Nederlands bevat (afkortingen vermijden) en niet te lang wordt. Maak deze ook 'future proof' door een soort versionering in de titel op te nemen wanneer je verwacht dat er nog versies, observaties of opnames zullen volgen.

Tip

Je projectnaam hoeft niet in de titel te vermelden: daarvoor kan je beter andere velden (zoals 'auteur', 'uitgever' of 'beschrijving') gebruiken.

Veld	Slecht voorbeeld	Goed voorbeeld
Titel	Leuvenair PM2.5/PM10 '19	Westkapelle
Beschrijving	Sensordata	Sensordata (PM2.5/PM10) uit burgerinitiatief 'Leuvenair' t.b.v. het in kaart brengen van de luchtkwaliteit in Leuven
Thema's	Milieu	Milieu, Gezondheid
Trefwoorden	(niet ingevuld)	Open data, luchtkwaliteit, burgerwetenschap

- Auteur en/of uitgever met contactinformatie**
Zorg voor up-to-date contactgegevens (projectcoördinator, datamanager en/of organisatie) en werkende e-mailadressen, telefoonnummers en websites zodat geïnteresseerde gebruikers of dataleveranciers met jou in contact kunnen treden voor vragen, suggesties en ideeën.
- Informatie over (gebruiks)rechten**
Onder deze rubriek kan je je open licentie (zie **principe 2**) vermelden, met eventuele restricties op het gebruik van de dataset. Ook de gewenste bron-/naamsvermelding kan hier opgenomen worden.
- Taal, formaat, actualiteit en (ruimtelijke/temporele) begrenzing van de dataset**
Dit zijn belangrijke metagegevens om zich een goed beeld te kunnen vormen van de data en de bruikbaarheid daarvan. Zo kan één gebruiker specifiek op zoek zijn naar recente data (voorbij twee jaar) over het Vlaamse grondgebied, terwijl een andere gebruiker vooral geïnteresseerd is in historische data over één stad. Ook het formaat (zie **principe 18**) kan hier meegegeven worden, waardoor men zicht krijgt op de softwaretoepassingen waarmee men de data kan bekijken of verwerken. De actualiteit kan zich vertalen in een 'laatst bijgewerkt'-datum, of de vermelding van de updatefrequentie van de dataset.

- Het wordt sterk aanbevolen om **nog een stap verder te gaan** in het beschrijven van metadata, door deze te verrijken met volgende informatie:
- **Bronnen, inwinningsmethodes en processtappen**
Hoe werden de data aangemaakt of ingewonnen? Zeker voor citizen-scienceprojecten is het een meerwaarde de methodologie (waarbij burgers een rol opnamen) summier maar volledig te beschrijven. Werd er gebruik gemaakt van andere databronnen, en zo ja, welke transformatie- of verrijkingsstappen werden daarop doorgevoerd? In de metadata kunnen verwijzingen naar andere, gerelateerde of complementaire bronnen worden opgenomen.
- **Datakwaliteit**
Met welke accuraatheid en precisie werden metingen uitgevoerd? Gaat het om ruwe of naverwerkte data? (Hoe) werden er calibraties gedaan? Hoe groot was de steekproef? Welke sensoren werden gebruikt? Ook conformiteit met datastandaarden (d.i. de dataset volgt de semantiek die beschreven staat in een erkend informatiemodel) is zeker het vermelden waard.
- **URL's van webservices**
Indien de projectdata niet (enkel) als dataset, maar (ook) via een webservice worden aangeboden, dan kan vanuit de metadata van de dataset verwezen worden naar de URL's van deze services. Zo worden geografische databronnen typisch niet alleen als dataset ontsloten, maar ook via OGC-conforme kaartdiensten, zoals een **WM(T)S**.
- **Applicatieprofiel of objectencatalogus**
Last but not least is het zeer zinvol om werk te maken van een gedocumenteerd **applicatieprofiel**¹⁵⁵ of een nette objectencatalogus¹⁵⁶. Dit neemt de vorm aan van een website of een los document. Daarin worden de entiteiten, attributen, relaties en codelijsten beschreven die deel uitmaken van je conceptueel datamodel (zie **principe 16**) en/of dataset, vergezeld van hun definities en datatypes. Op die manier kan een gebruiker bijvoorbeeld de exacte betekenis van een attribuut of code opzoeken.

¹⁵⁵ Voorbeeld applicatieprofiel voor Gebouwenregister: <https://data.vlaanderen.be/doc/applicatieprofiel/gebouwenregister/>

¹⁵⁶ Voorbeeld objectencatalogus voor Basiskaart Vlaanderen (GRB): https://metadata.vlaanderen.be/srv/dut/cataog_search#/metadata/9650dfb1-9e7f-4ce2-8ee9-98a1b7733ee4

22. Denk eraan de metadata zo dicht mogelijk bij de bron te capteren.

Als projectleider van een citizen-scienceproject zorg je ervoor dat alle nodige metadata gecapteerd worden bij de gegevens die in jouw project ingewonnen of geanalyseerd worden. Anders dan bij regulier onderzoek heb je niet altijd controle over welk meettoestel of welke analysetool gebruikt wordt bij bepaalde citizen-sciencetaken. Wanneer die informatie relevant is (en niet in tegenstrijd met de GDPR), zorg je technisch voor mogelijkheden om je deelnemers deze metadata te laten toevoegen aan de verzamelde gegevens.



Zo zou je bijvoorbeeld, wanneer deelnemers foto's toevoegen aan hun waarnemingen, ervoor kunnen zorgen dat de deelnemer bij inschrijving in een formulier aangeeft welk type fototoestel hij/zij gebruikt. Een andere manier om deze metadata te bemachtigen is te vragen de instellingen van de camera zo aan te passen dat parameters als (f-stops, sluitertijd, ISO, GPS-locatie) automatisch mee opgeslagen worden bij de foto's zelf.

Denk dus vooraf niet alleen na over de manier waarop data zullen verzameld worden, maar ook over hoe deze meteen, en zo dicht mogelijk bij de bron, van metadata kunnen voorzien worden.



Praktijkvoorbeeld

Zoals bij **Principe 15** aangegeven, verzamelt het internationale weerplatform **WOW**¹⁵⁷ de **metadata** van weerstations om een algemeen kwaliteitslabel aan elk station te koppelen, en de eigenschappen van het station weer te geven voor de gebruiker van de website. Burgerwetenschappers die gegevens van hun weerstation naar het platform doorsturen, kunnen daar ook de specificaties en omgevingsfactoren van hun station ingeven. Informatie die meegegeven wordt omvat o.a.: de motivatie van de burgerwetenschapper, of het station behoort tot een school of een officieel netwerk, of er een webcam aanwezig is, in welke tijdzone het station zich bevindt, en wanneer de meest recente waarneming werd opgemeten. Verder geeft het dashboard ook o.a. de ligging (open/ gesloten), de waarneemuren en een index die weergeeft of het station zich in een stadszone bevindt of niet.

¹⁵⁷ <https://wow.meteo.be/nl/>

23. Ken aan je metadataset een globaal unieke en persistente identicator toe.

Net zoals bij de dataset het geval is, dient ook naar de bijhorende set van metadata verwezen te worden met een globaal unieke en persistente identicator. Ook hier kan voor een **GUID** of **URI** gekozen worden die automatisch door het metadataportaal wordt gegenereerd. Via deze identificatoren kunnen data en metadata onlosmakelijk met elkaar verbonden worden, en kan er op ondubbelzinnige wijze naar dataset en metadataset worden verwezen.





Praktijkvoorbeeld

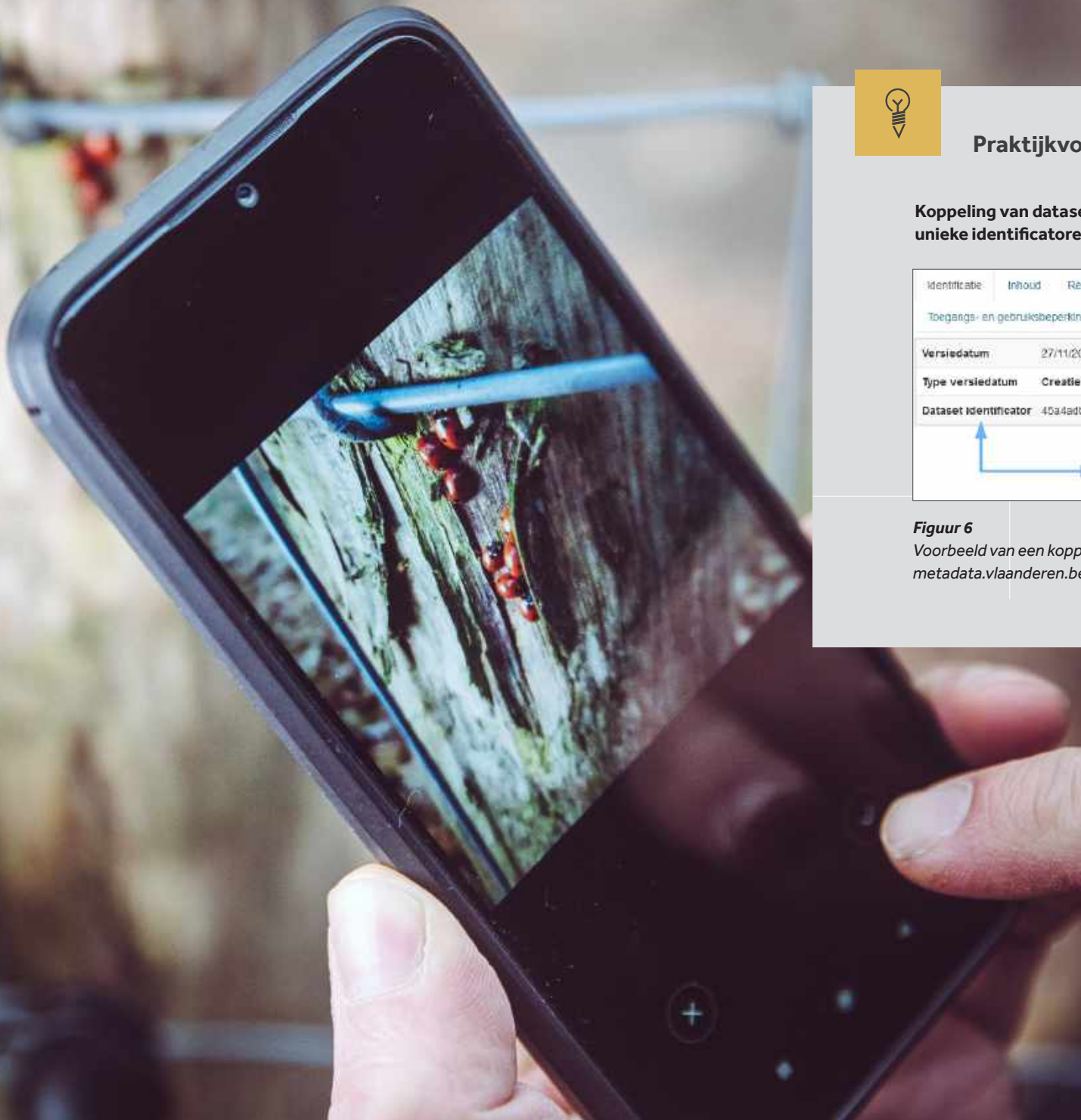
Koppeling van dataset met bijhorende metadataset via globaal unieke identificatoren

Identificatie	Inhoud	Referentiesysteem	Kwaliteit
Toegangs- en gebruiksbeperkingen: --			
Versiedatum	27/11/2017		
Type versiedatum	Creatie: Aanmaakdatum of opnamedatum van de dataset		
Dataset Identifier	45a4adbd-349c-4cb7-b83b-4d56ab50abbfd		
	Metadata Identifier (fileIdentifier)	7bd48678-7c0b-4230-b734-cb5638c56096	

Note: A blue arrow points from the Dataset Identifier to the Metadata Identifier.

Figuur 6

Voorbeeld van een koppeling van data en metadata uit metadata.vlaanderen.be



24. Gebruik de juiste standaard voor je metadata.

Ook voor metadatabeschrijvingen bestaat een overvloed aan standaarden. Velen vormen een uitbreiding of specialisatie van de Dublin Core¹⁵⁸, een set van 15 elementen (later gestandaardiseerd als **ISO 15836**) die toelaat bronnen over domeinen heen te metadateren. De velden in de Dublin Core vormen zowat de ondergrens. Je kiest echter beter voor een andere (domeinspecifieke) standaard die rijkere metadatering toelaat. Door een standaard te volgen kunnen metadata vlot tussen systemen uitgewisseld worden (bv. doorgave van regionaal/nationaal naar internationaal platform), waardoor je metadata ook in andere portalen (dan hetgeen waarin je je metadata de eerste keer beschreef) teruggevonden kunnen worden. Een standaard volgen heeft dus positieve gevolgen voor de vindbaarheid van je dataset.



¹⁵⁸ Opgesteld tijdens een workshop in 1995 in <https://www.wikidata.org/wiki/Q988937>.

Gemeenschappelijke kern van basisinformatie

Ook metadata worden opgesteld volgens (**domeinspecifieke**) standaarden. Afhankelijk van het onderzoeksdomein waartoe je project behoort, kunnen er wetenschappelijke datastandaarden ontwikkeld zijn.

Tip

Probeer zoveel mogelijk jouw metadata te laten voldoen aan de gevestigde standaarden in jouw onderzoeksdomein. Conformiteit bereiken met zo'n metadatastandaard is doorgaans kwestie van het kiezen van een portaal dat de standaard in kwestie ondersteunt en begeleiding aanbiedt bij het invullen van de noodzakelijke velden.

Hierbij een niet-exhaustief overzicht van metadatastandaarden:

- **Dublin Core (Metadata Element Set)**
<https://www.dublincore.org/specifications/dublin-core/dc-mi-terms/>
Zeer generiek opgevat. Zie de 15 'elements'. Geschikt voor het beschrijven van fysieke en digitale 'resources' (zoals boeken of afbeeldingen).
- **DCMI Metadata Terms**
<https://www.dublincore.org/specifications/dublin-core/dc-mi-terms/>
Uitbreiding van de 'core' met bijkomende 'termen'. Evenzeer toepassings- en domeinneutral.
- **DataCite**
<http://schema.datacite.org/>
Bedoeld voor het gestandaardiseerd citeren en ophalen van bronnen (bv. publicaties). Hanteert DOI-nummers.
- **SDMX**
<https://sdmx.org/>
Uitbreiding van de 'core' met bijkomende 'termen'. Evenzeer toepassings- en domeinneutral.
- **DCAT**
<https://www.w3.org/TR/vo-cab-dcat-2>
Uitgebreide W3C-standaard bedoeld voor het beschrijven van datasets en data services in (overheids)catalogi. Bouwt verder op DCMI.
- **Metadata DCAT**
<https://data.vlaanderen.be/doc/applicatieprofiel/metadata-dcat/>
Applicatieprofiel gebaseerd op DCAT dat zich richt op het verzamelen van informatie over generieke datasets, distributies en services die door een overheid beschikbaar gesteld worden.
- **DCAT-AP-VL**¹⁵⁹
<https://data.vlaanderen.be/doc/applicatieprofiel/DCAT-AP-VL>
Applicatieprofiel afgeleid van Metadata DCAT voor de publicatie van generieke open data in het Vlaams Open Data Portaal (VODAP).
- **DDI**
<https://ddialliance.org/>
Bedoeld voor het beschrijven en metadateren van data uit enquêtes en andere observatiemethoden gebruikt in de sociale, economische, gedrags- en gezondheidswetenschappen.
- **EML**
<https://eml.ecoinformatics.org/>
Standaard ontwikkeld voor aard- en milieuwetenschappen en ecologie. Geschikt voor taxonomische en geografische data en het citeren van bronnen. Gebruikt door GBIF-platform.
- **GDI-Vlaanderen metadataprofiel**¹⁶⁰
<http://www.geopunt.be/nl/voor-experts/metadata>
Profiel gebaseerd op relevante ISO-standaarden en de INSPIRE-kaderrichtlijn, specifiek bedoeld voor de publicatie van geografische datasets (coördinaten, zones, afstanden, hoogtemetingen enz.). Mapbaar op de GeoDCAT-standaard.

¹⁵⁹ Citizen-scienceprojecten kunnen hiervan gebruik maken mits de publicatie van de metadata door een Vlaamse gemeente of overheidsinstantie gebeurt.

¹⁶⁰ Combinatie van ISO 19115 (metadata voor geografische datasets), ISO 19119 (services), ISO 19110 (objectencatalogus) en 'INSPIRE Implementing Rules for Metadata'. Ook hier publicatie door overheid vereist.

Let op!

In de citizen-science wereld zijn een aantal metadatastandaarden voor zowel projecten als datasets ontwikkeld. Deze baseren zich zo veel mogelijk op bestaande metadatastandaarden. Het loont dus meer dan de moeite om deze te consulteren.

- **PPSR-CORE DMM**
<https://core.citizenscience.org/docs/dataset/>
 Geschikt voor metadateren van datasets in citizen-scienceprojecten. Een dataset wordt hier gedefinieerd als een collectie van data, verzameld in een project door burgerwetenschappers aan de hand van een enkel protocol. Je kan een project dus associëren met meerdere datasets.
- **PPSR-CORE PMM**
<https://core.citizenscience.org/docs/project>
 Geschikt voor metadateren van citizen-scienceprojecten. Belangrijke elementen die in dit model opgenomen zijn, zijn:
 - *De titel en beschrijving van het project*
 - *Eigenaarschap en contactgegevens*
 - *De duurtijd van het project (start/einde)*
 - *Projectpartners*
 - *Financieringsprogramma of -bron*
 - *Informatie over de publieke participatie (op welke manier nemen burgerwetenschappers deel? In welke rol?)*
 - *Links naar andere (projectgerelateerde) sites en bronnen*
 - *Projectgerelateerde grafische elementen*

**Praktijkvoorbeeld**

De **PPSR-Core PMM** wordt reeds gebruikt in de volgende citizen science webplatformen om projectgegevens met elkaar uit te wisselen: [Scistarter.org](https://scistarter.org/)¹⁶¹, [CitSci.org](https://www.citisci.org/)¹⁶², [Atlas of Living Australia – Biocollect](https://atlasoflivingaustralia.org.au/)¹⁶³, [CitizenScience.gov](https://citizenscience.gov/)¹⁶⁴. Een meer gedetailleerd voorbeeld van zo'n data-uitwisseling staat beschreven op de [website](#) van de **PPSR-Core PMM**.

¹⁶¹ <https://scistarter.org/>

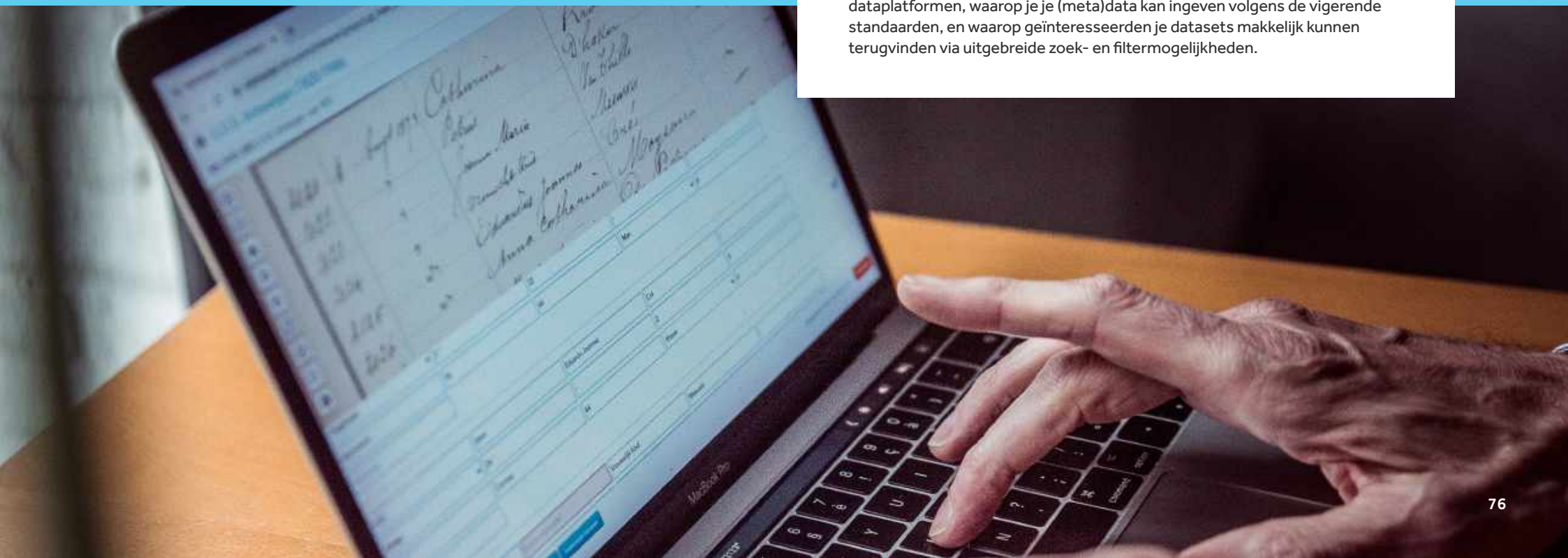
¹⁶² <https://www.citisci.org/CWIS438/Websites/CitSci/Home.php?WebSiteID=7>

¹⁶³ <https://ala.org.au/biocollect/>

¹⁶⁴ <https://www.citizenscience.gov/#>

25. Maak je metadata, en dus je datasets en onderzoeksresultaten, vindbaar door ze te registreren op een doorzoekbaar portaal.

Persistente identificatoren en rijke metadatabeschrijvingen alleen maken je datasets nog niet vindbaar op het internet. Wanneer je metadata nergens online geregistreerd en (door zoekmachines) geïndexeerd worden, dan blijven je datasets onzichtbaar voor de buitenwereld. Het is daarom aangewezen op zoek te gaan naar (de binnen jouw domein) gerenommeerde (meta) dataplatformen, waarop je je (meta)data kan ingeven volgens de vigerende standaarden, en waarop geïnteresseerden je datasets makkelijk kunnen terugvinden via uitgebreide zoek- en filtermogelijkheden.



Metadataplatformen

Een volledige oplistings maken is ook hier niet mogelijk, maar volgende **metadataplatformen** zijn alvast het vermelden waard:

- **FRIS-onderzoeksporaal van het departement EWI**
<https://researchportal.be/nl>
In FRIS worden jaarlijks ruim 35 000 nieuwe publicaties geregistreerd¹⁶⁵. De metadata van deze wetenschappelijke artikelen en boek(bijdragen) zijn doorzoekbaar op kennisinstelling, type, jaar en discipline. Naar de feitelijke data wordt gelinkt, o.m. via **DOI**-nummers.
- **Metadata Vlaanderen**
<https://metadata.vlaanderen.be>
Dit portaal kan door citizen-scienceprojecten gebruikt worden om interessante (overheids)data op het spoor te komen, maar ook om de eigen projectdata te metadateren en vindbaar te maken. Je kan er geografische (volgens GDI-Vlaanderen metadataprofiel) en open data (volgens DCAT-AP-VL) beschrijven, en binnenkort ook open API's (volgens metadata DCAT). Ook hier worden dan links voorzien naar de data op de door jou gekozen locatie. De ingevoerde metadata worden automatisch gedeeld met federale en Europese metadatanodes.

¹⁶⁵ Bron: <https://researchportal.be/nl/cijfers>

¹⁶⁶ <https://waarnemingen.be/>

¹⁶⁷ https://www.gbif.org/occurrence/search?country=BE&publishing_org=4d3cee8-5699-439d-a899-decac9cbbdac

¹⁶⁸ <https://www.plantentuinmeise.be/nl/>

¹⁶⁹ <https://www.gbif.org/publisher/a344ee9f-f1b7-4761-be2c-58ee6d741395>



Praktijkvoorbeeld

Metadata van de website 'waarnemingen.be'¹⁶⁶ (Natuurpunt) stromen (semantisch operabel dankzij het gebruik van standaarden, en technisch interoperabel via gedocumenteerde **API's**) automatisch door naar het GBIF-portaal, waardoor de door vrijwilligers ingezamelde data vindbaar worden voor een zeer breed publiek van onderzoeksinstituten en onderzoekers.

De indrukwekkende lijst¹⁶⁷ van meer dan 18 miljoen waarnemingen op het Belgisch grondgebied.

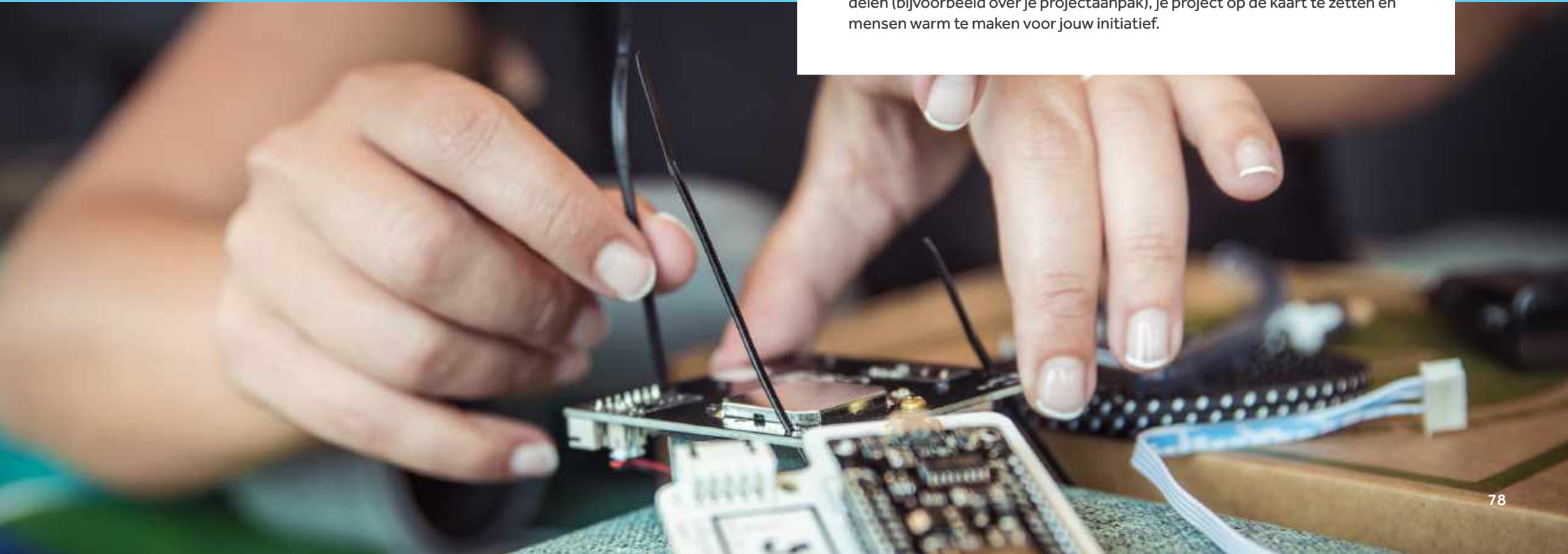
Plantentuin Meise¹⁶⁸ is een van de onderzoeksinstituten die datasets levert aan het GBIF-portaal¹⁶⁹. De bij benadering 2 miljoen observaties (15 datasets) kunnen door anderen gebruikt worden voor analyses. Wetenschappers die data uit het portaal extraheren voor onderzoek, en nadien hun resultaten publiceren, refereren daarin naar de bron van hun data. Dat leverde de Plantentuin reeds 423 citaties in publicaties op.

Tip

ook wanneer je dataset verouderd is, niet langer onderhouden wordt of zelfs compleet verdwenen is van zijn aangegeven locatie (bv. offline gehaald), dan nog laat je de bijhorende metadata best staan op het metadataplatform. Via de metadata kunnen researchers en andere geïnteresseerden nog steeds de mensen, organisaties en publicaties terugvinden achter de data, en dat op zich is zeer waardevol. Dat de dataset niet langer beschikbaar is, kan doorgaans aangegeven worden via een statusveld in het platform. Bij archivering van een metadatarcord moeten mogelijks ook andere velden worden aangepast (bv. weghalen van het luik 'distributie').

26. *Maak je citizen-scienceproject vindbaar voor collega's en vrijwilligers.*

Er bestaan metadatastandaarden en portalen voor het beschrijven van datasets en publicaties, maar ook voor het vindbaar maken van citizen-scienceprojecten en de daarmee verbonden personen en organisaties. Het belang daarvan mag niet onderschat worden. Het laat je immers toe om in contact te treden met collega-projectcoördinatoren, inzichten en kennis te delen (bijvoorbeeld over je projectaanpak), je project op de kaart te zetten en mensen warm te maken voor jouw initiatief.



Citizen-scienceproject vindbaar

Enkele websites waarop je je **citizen-scienceproject vindbaar** kan maken, of je netwerk kan uitbreiden:

- **Iedereen Wetenschapper (Metadata Element Set)**
<https://www.iedereenweten-schapper.be/>
Op Vlaams niveau is deze door Eos Wetenschap en de Jonge Academie¹⁷⁰ opgerichte website de plaats bij uitstek om citizen-scienceprojecten zichtbaar te maken en geëngageerde vrijwilligers te betrekken.
- **EU-Citizen.Science**
<https://eu-citizen.science/>
Voor citizen-scienceprojecten is recent op Europees niveau een databank opgericht, waarin metadata van projecten kunnen opgevoerd worden. Daarnaast zijn op de website ook hulpbronnen, tools en trainingen te vinden voor je project.
- **SciStarter**
<https://scistarter.org/>
Een andere CS-website die, net zoals Zooniverse (zie verderop), tot doel heeft de barrières tussen de academische wereld en niet-wetenschappers te slopen. Zie ook de website van partnerorganisatie CitSci.org (<https://www.citsci.org/>).
- **Zooniverse**
<https://www.zooniverse.org/>
Met meer dan 1,6 miljoen geregistreerde vrijwilligers is Zooniverse uitgegroeid tot het grootste citizen-scienceplatform op het internet. Het werd opgericht met het doel de inspanningen van vrijwilligers om te zetten in meetbare resultaten en nieuwe inzichten.
- **FRIS-onderzoeksportal van het departement EWI**
<https://researchportal.be/nl>
Naast publicaties zijn in FRIS ook organisaties, onderzoekers en wetenschappelijke projecten terug te vinden. De teller staat ondertussen op ruim 40 000 geregistreerde projecten die o.m. doorzocht kunnen worden op kennisinstelling, financierder, discipline en startdatum.

De PPSR-Core metadatastandaarden voorzien een apart PPM-model voor het metadateren van citizen-scienceprojecten (zie principe 24 voor meer uitleg).

¹⁷⁰ Meer info: <https://jongeacademie.be/>



Buyle R., De Vocht L., Van Compernelle M., De Paepe D., Verborgh R., Vanlshout Z., De Vidts B., Mechant P., Mannens E., 2016. Open Standards for Linked Organisations. EGOSE'16, nov 22-23, 2016. St-Petersburg, Russian Federation. DOI: <https://doi.org/10.1145/3014087.3014096>

Frank M.T., Zander S. (2019) The Linked Data Wiki: Leveraging Organizational Knowledge Bases with Linked Open Data. In: Fred A. et al. (eds) Knowledge Discovery, Knowledge Engineering and Knowledge Management. IC3K 2017. Communications in Computer and Information Science, vol 976. Springer, Cham. https://doi.org/10.1007/978-3-030-15640-4_15

Harlin J., Kloetzer L., Patton D., Leonhard C., Leysin American School high school students. 2018. Turning Students into Citizen Scientists. In: Hecker S., Hacklay M., Bowser A., Makuch Z., Vogel J., Bonn A., 2018. Citizen Science: Innovation in Open Science, Society and Policy. UCL Press, London. <https://doi.org/10.14324/111.9781787352339>

Keel W.C., Lintott C.J., Schawinski K., Bennert V.N., Thomas D., Manning A., Drew Chojnowski S., van Arkel H., Lynn S. 2012. The history and environment of a faded Quasar: Hubble Space Telescope Observations of Hanny's Voorwerp and IC 2497. The Astronomical Journal, 144 (2), 16pp. DOI: <https://doi.org/10.1088/0004-6256/144/2/66>

Porter W.T., Barrand Z.A., Wachara J., DaVall K., Mihajevic J.R., Pearson T., Salkeld D.J., Nieto N.C. 2021. Predicting the current and future distribution of the western black-legged tick, *Ixodes pacificus*, across the Western US using citizen science collections'. PLOS ONE, 16 (1), e0244754. DOI: <https://doi.org/10.1371/journal.pone.0244754>

Resnik D.B., Elliott K.C., Miller A.K., 2015: A framework for addressing ethical issues in citizen science. Environmental Science & Policy 54, pp. 475-481. <https://doi.org/10.1016/j.envsci.2015.05.008>

Reyserhove L., Desmet P., Oldoni D., Adriaens T., Strubbe D., Davis A.J.S., Vanderhoeven S., Verloove F., Groom Q. 2020. A checklist recipe: making species data open and FAIR. Database 1-12, article ID baaa084; DOI: <https://doi.org/10.1093/database/baaa084>

Scassa T., Chung H., 2015: Managing Intellectual Property Rights in Citizen Science: A Guide for Researchers and Citizen scientist. Washington DC, Woodrow Wilson International Center for Scholars. <https://www.wilsoncenter.org/publication/managing-intellectual-property-rights-citizen-science-guide-for-researchers-and-citizen>

Schoonvaere K., Laget D., Adriaens T., Desmet P., Villers V., de Graaf D. 2020. Vespa-Watch: Invasiemonitoring van de Aziatische hoornaar met hobbyimkers en het publiek. Eindrapport in

het kader van de oproep Citizen Science van het Departement Economie, Wetenschap en Innovatie (EWI) van december 2017. Honeybee Valley en Instituut voor Natuur- en Bosonderzoek, Brussel. DOI: <https://doi.org/10.21436/inbor.19019045>

Wiedemair C., Sterken C., Eenmae T., Neito R., Oberhauser V., Moser D., Mair M., Lercher G. 2020. CCD photometry of CY Aquarii VI. The 2019-2020 seasons. The Journal of Astronomical Data 26, (1), 1-8.

Wilkinson MD., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., 't Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. 2016: The FAIR Guiding Principles for scientific data management and stewardship. Nature Scientific Data 3, 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>

Afkortingenlijst

Afkorting	Voluit
API	<i>Application Programming Interface</i>
AVG	<i>Algemene Verordening Gegevensbescherming</i>
BAG	<i>Basisregistratie Adressen en Gebouwen</i>
CAD	<i>Computer-aided design</i>
CC	<i>Creative Commons</i>
CS	<i>Citizen Science</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DDI	<i>Data Documentation Initiative</i>
DMM	<i>Dataset Metadata Model</i>
DMP	<i>Datamanagementplan</i>
DPO	<i>Data Protection Officer</i>
DOI	<i>Digital Object Identifier</i>
ECSA	<i>European Citizen Science Association</i>
EML	<i>Ecological Metadata Language</i>
ERD	<i>Entity-Relationship Diagram</i>
EWI	<i>Economie, Wetenschap & Innovatie</i>
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FOSB	<i>Flemish Open Science Board</i>
FOSS	<i>Free and Open-Source Software</i>
FRIS	<i>Flanders Research Information Space</i>
GBIF	<i>Global Biodiversity Information Facility</i>
GDPR	<i>General Data Protection Regulation</i>
GPL	<i>(GNU) General Public License</i>
GUID	<i>Globally Unique Identifier</i>
JSON-LD	<i>JavaScript Object Notation for Linked Data</i>
ICOS	<i>Integrated Carbon Observation System</i>

Afkorting	Voluit
INSPIRE	<i>Infrastructure for Spatial Information in Europe</i>
ISO	<i>International Organization for Standardization</i>
LOD	<i>Linked Open Data</i>
NetCDF	<i>Network Common Data Form</i>
OAJ	<i>Open Access Journals</i>
ODC	<i>Open Data Charter of Open Data Commons</i>
OGC	<i>Open Geospatial Consortium</i>
ORM	<i>Object-role modeling</i>
OSLO	<i>Open Standaarden voor Linkende Organisaties</i>
OWL	<i>Web Ontology Language</i>
PMM	<i>Project Metadata Model</i>
PPSR	<i>Public Participation in Scientific Research</i>
PSI	<i>Public Sector Information</i>
PURL	<i>Persistent Uniform Resource Locator</i>
RDF	<i>Resource Description Framework</i>
SDMX	<i>Statistical Data and Metadata eXchange</i>
SPARQL	<i>SPARQL Protocol And RDF Query Language</i>
SQL	<i>Structured Query Language</i>
SSNO	<i>Semantic Sensor Network Ontology</i>
TDWG	<i>Taxonomic Databases Working Group</i>
Turtle	<i>Terse RDF Triple Language</i>
UML	<i>Unified Modeling Language</i>
URI	<i>Uniform Resource Identifier</i>
VODAP	<i>Vlaams Open Data Portaal</i>
WM(T)S	<i>Web Map (Tile) Service</i>
W3C	<i>World Wide Web Consortium</i>

Begrip	Definitie	Bronnen & referenties
Anonimiseren	Anonimiseren, ook wel data masking genoemd, is een methode waarbij persoonsgegevens zodanig worden bewerkt dat deze niet meer gebruikt kunnen worden om een persoon te identificeren. Deze bewerking is onomkeerbaar. Geanonimiseerde data zijn volgens de Algemene Verordening Gegevensbescherming (AVG) dan ook geen persoonsgegevens meer. Veelal omvat het anonimisatieproces een combinatie van pseudonimisatie en data-aggregatie.	https://www.fgsupport.nl/wat-is-anonimisering
Applicatieprofiel	Een applicatieprofiel is een specificatie voor gegevensuitwisseling voor applicaties (softwaretoepassingen) die een bepaalde use case vervullen. Het laat toe om naast een gedeelde semantiek (zie 'datamodel' en 'vocabularium') ook bijkomende restricties op te leggen, zoals het vastleggen van kardinaliteiten of het gebruik van bepaalde codelijsten. Een applicatieprofiel kan dienen als documentatie voor analisten en ontwikkelaars.	https://data.vlaanderen.be/ns/
Authentieke gegevensbron	Een authentieke gegevensbron is de meest volledige, kwalitatief hoogstaande verzameling van gegevens die op elektronische wijze worden bijgehouden, en die voor de overheidsinstanties nuttig of noodzakelijk zijn bij de uitvoering van de taken van algemeen belang waarmee ze zijn belast of bij de uitvoering van verplichtingen die op hen rusten.	Art. III.66 §1 van het Bestuursdecreet: https://codex.vlaanderen.be/PrintDocument.ashx?id=1030009&datum=&geannoteerd=false&print=false
Burgerwetenschap	Zie 'Citizen Science'	
Codelijst	Letterlijk: een lijst met codes. Wordt gebruikt in datamodellen wanneer een eigenschap (attribuut) geen enkelvoudige waarde aanneemt (zoals een stuk tekst of een getal), maar verschillende mogelijkheden (codes) uit een vooraf gedefinieerde (maar uitbreidbare) lijst kan krijgen. Voorbeeld: <ul style="list-style-type: none"> • Een sensor heeft een naam. Die naam kan als waarde een stukje tekst ('string') aannemen. • Een sensor heeft een status. De status moet gekozen worden uit een codelijst met als mogelijkheden {online, offline, onbekend}. 	(eigen definitie)

Begrip	Definitie	Bronnen & referenties
Citizen Science	<p>'Citizen Science' of burgerwetenschap is wetenschappelijk onderzoek dat in zijn geheel of gedeeltelijk door niet-wetenschappers (burgers) wordt uitgevoerd, vaak in samenwerking met of onder begeleiding van professionele wetenschappers. De burgers die deelnemen aan een citizen science project worden burgerwetenschappers of citizen scientists genoemd.</p> <p>Citizen science is een brede term, die vele ladingen dekt. Burgerwetenschappers kunnen, afhankelijk van het project waaraan ze bijdragen, een breed gamma van taken uitvoeren; ze voeren tellingen en observaties uit, bouwen en installeren sensoren, annoteren of transcriberen afbeeldingen en teksten, voeren data-analyses uit, bedenken onderzoeksvragen en onderzoeksmethodes en rapporteren of dissemineren onderzoeksresultaten.</p>	<p>https://www.scivil.be/aan-de-slag/voor/citizen-science-iets-voor-mij</p>
Datamanagementplan	<p>Een datamanagementplan (DMP) is een formeel document dat beschrijft hoe onderzoekers voor, tijdens en na het onderzoek omgaan met de verzamelde data:</p> <ul style="list-style-type: none"> hoe data zal verzameld worden; de omvang en het formaat van de data; de manier waarop de data zal opgeslagen worden; wie zowel tijdens de uitvoering van het onderzoek als daarna toegang zal hebben tot de data. <p>Dergelijk plan bevat eveneens informatie over de juridische en ethische aspecten van data.</p>	<p>https://www.ugent.be/nl/onderzoek/datamanagement/datamanagementplan</p>
Datamodel	<p>Met een datamodel, of gegevensmodel, wordt beschreven hoe de gegevens in een informatiesysteem gestructureerd zijn.</p> <p>Men onderscheidt</p> <ul style="list-style-type: none"> • het conceptueel datamodel; • het logisch datamodel; • het fysiek datamodel. 	<p>https://nl.wikipedia.org/wiki/Datamodel</p>

Begrip	Definitie	Bronnen & referenties
Datastandaard	<p>Datastandaarden zijn noodzakelijk wanneer verschillende partijen met elkaar willen samenwerken en gegevens uitwisselen, maar er begripsverwarring ontstaat over de begrippen (het 'jargon') binnen het domein in kwestie. In dat geval kunnen domeinexperten beslissen samen concepten uit de realiteit te gaan definiëren, en de structuur van (of relaties tussen) gegevens vast te leggen.</p> <p>Een datastandaard is het gedocumenteerd resultaat van deze oefening. Wanneer de verschillende partijen dezelfde concepten en definities uit de datastandaard hanteren, dan is gegevensuitwisseling mogelijk en wordt (semantische) interoperabiliteit bereikt.</p> <p>Een datastandaard kan de vorm aannemen van een vocabularium, ontologie, conceptueel datamodel (= domeinmodel) of applicatieprofiel, of een combinatie daarvan. In datamodellen kunnen via de gekozen modelleertaal veel nuances toegevoegd worden om het domein gedetailleerd én eenduidig te beschrijven.</p>	(eigen definitie)
Data-aggregatie	Het samenvoegen van data (bijvoorbeeld door een gemiddelde waarde te nemen, of de minimum-en maximumwaarde).	(eigen definitie)
Dataserialisatie	Het omzetten van een dataobject naar een formaat dat (sequentieel) opgeslagen of over een computernetwerk verzonden kan worden. Naast de gegevens zelf wordt ook (een deel van) de datastructuur meegeleverd in het geserialiseerde formaat. Deze metagegevens worden gebruikt om de gegevens bij deserialisatie correct te interpreteren.	https://nl.wikipedia.org/wiki/Serialisatie
DBpedia	Een machineleesbare, 'linked data'-versie van de artikelen op Wikipedia.	https://nl.wikipedia.org/wiki/DBpedia
DOI-nummer	Digital Object Identifier: een uniek nummer dat aan een tijdschriftartikel wordt toegekend.	https://library.uic.edu/help/article/1966/what-is-a-doi-and-how-do-i-use-them-in-citations#:~:text=A%20DOI%2C%20or%20Digital%20Object,to%

Begrip	Definitie	Bronnen & referenties
FAIR	<p>Om het potentieel van onderzoeksdata volop te benutten, is het noodzakelijk om ze zo vindbaar (Findable), toegankelijk (Accessible), interoperabel (Interoperable) en herbruikbaar (Reusable) als mogelijk in het onderzoeksecocysteem op te nemen (Force 11, 2014; Wilkinson, et al., 2016; GoFAIR, n.d.).</p> <p>De FAIR-principes bestaan uit 15 facetten. Wat voorop staat is dat onderzoeksdata niet alleen FAIR moeten zijn voor mensen, maar ook voor machines. De FAIR-principes zijn niet meer weg te denken uit het datamanagementlandschap en staan aan de basis van de bouwtekening van de 'European Open Science Cloud'.</p>	<p>Naar:</p> <p>https://datasupport.researchdata.nl/start-de-cursus/i-in-vogelvlucht/begrippen/rdm-open-en-fair</p> <p>https://datasupport.researchdata.nl/start-de-cursus/iv-oogstfase/european-open-science-cloud</p>
FAIR-datamanagement	<p>FAIR-datamanagement is het geheel van beslissingen en maatregelen die tijdens de levenscyclus van onderzoeksdata worden getroffen om onderzoeksdata zo FAIR als mogelijk af te leveren. Wat daarbij belangrijk is, is de keuze voor een datalicentie, voor het dataformaat, het vastleggen van metadata en datadocumentatie, naamgevingsconventies, etc.</p>	<p>https://datasupport.researchdata.nl/start-de-cursus/i-in-vogelvlucht/begrippen/rdm-open-en-fair</p>
Gevoelige persoonsgegevens	<p>Gevoelige of bijzondere persoonsgegevens zijn gegevens die volgens de AVG of GDPR-wetgeving een verhoogde bescherming moeten krijgen. Het gaat om gegevens over de etnische afkomst of het ras van personen, de politieke, religieuze of levensbeschouwelijke overtuigingen of het al dan niet lid zijn van een vakbond, genetische informatie, biometrische informatie (bv. de iris van een oog, of een vingerafdruk), gezondheidsgegevens of gegevens met betrekking tot de seksuele geaardheid van een persoon. Ook gegevens met betrekking tot strafrechterlijke vervolging horen tot deze categorie.</p>	<p>Naar:</p> <p>https://www.gegevensbeschermingsautoriteit.be/professioneel/themas/gevoelige-gegevens</p> <p>https://onderzoektips.ugent.be/nl/tips/00001772/</p>
GUID	<p>Globally Unique Identifier, een door software automatisch gegenereerde cijfer- en lettercombinatie die verondersteld wordt wereldwijd uniek te zijn. Dit maakt GUID's uitermate geschikt om te gebruiken als identificatoren, aangezien identificatoren uniek naar een object moeten verwijzen.</p> <p>De typische vorm van een GUID is als volgt: {3F2504E0-4F89-11D3-9A0C-0305E82C3301}</p>	<p>Naar:</p> <p>https://nl.wikipedia.org/wiki/Globally_unique_identifier</p>

Begrip	Definitie	Bronnen & referenties
Interoperabiliteit	Producten, systemen of organisaties zijn interoperabel als ze zonder beperkingen samen kunnen werken. Om dit te bereiken zijn standaarden, protocollen en procedures nodig. Men spreekt o.m. over 'semantische' (definities van dataconcepten), 'technische' (technische richtlijnen), 'organisatorische' en 'juridische' interoperabiliteit.	Naar: https://nl.wikipedia.org/wiki/Interoperabiliteit https://vloca-kennishub.vlaan
Linked Data	Gestructureerde gegevens (data) die gelinkt zijn aan andere gegevens en daardoor beter bruikbaar zijn in semantische queries. De methode is gebaseerd op de technologie van HTTP-URI's en RDF. 'Semantische queries' zijn bevestigingen op het semantische web, wat een doorontwikkeling is van het wereldwijde web waarbij de betekenis van data wordt ingesloten in webpagina's, zodat ook automatische systemen intelligente zoekopdrachten kunnen uitvoeren.	Naar: https://nl.wikipedia.org/wiki/Linked_data https://nl.wikipedia.org/wiki/Semantisch_web
Metadata	Metadata zijn gegevens die de karakteristieken van bepaalde gegevens beschrijven. Het zijn dus eigenlijk data over data. Bijvoorbeeld: wanneer je temperatuursmetingen doet, kan je in de metadata opnemen dat de grootte waarin je meet de graad Celsius is, en niet de graad Fahrenheit.	https://nl.wikipedia.org/wiki/Metadata
Ontologie	Een ontologie staat in de informatiewetenschap voor het weergeven van een set van concepten in een (kennis)domein. In een ontologie worden definities van concepten en relaties tussen concepten vastgelegd. De Web Ontology Language (OWL) is een W3C-standaard voor het definiëren van webontologieën.	https://www.ensie.nl/bibliotheecarium/ontologie
Open data	<ul style="list-style-type: none"> 'Open' betekent dat data vrij geconsulteerd, gebruikt, aangepast en gedeeld kunnen worden door eender wie, waarbij hoogstens voorwaarden worden opgelegd om de herkomst en openheid ervan te vrijwaren. 'Open data' is een term die wordt gebruikt om vrij beschikbare informatie aan te duiden. De voorwaarden waaronder deze informatie beschikbaar is wordt beschreven in licenties en gebruiksvoorwaarden. Bij open data wordt er wel naar gestreefd om de beperkingen in hergebruik tot een minimum te beperken. 	Naar: http://opendefinition.org https://okfn.org/opendata/ https://opendatahandbook.org/guide/nl_BE/what-is-open-data/ https://nl.wikipedia.org/wiki/Open_data

Begrip	Definitie	Bronnen & referenties
Open Licentie	Een licentie is een soort belofte die je maakt voor anderen dat ze jouw werk (data, databank, software, of eender welk ander werk dat onder de licentie valt) mogen consulteren, gebruiken, distribueren zonder dat je hen zal aanklagen. Tenminste, als ze de voorwaarden naleven die in jouw licentie-overeenkomst staan. Jij kiest dus wat met jouw informatie mag gebeuren, en de licentie expliciteert dat.	(eigen definitie)
ORCID iD	De 'ORCID iD' is een uniek nummer dat door onderzoekers kan aangevraagd worden. Deze URI kan bijvoorbeeld vermeld worden bij alle wetenschappelijke publicaties van die persoon. Zo kan je nooit twee researchers met dezelfde voor- en achternaam met elkaar verwarren wanneer je hun lijst van publicaties wil opvragen.	
Preprint-Service	Een pre-print service is een webportaal waar wetenschappelijke artikels beschikbaar zijn nog voordat ze door het peer-review-proces gaan. Door hun voorlopige tekst online te delen stellen wetenschappers hun bevindingen al aan elkaar voor voordat het volledige proces van publicatie voltooid is. Let op: preprint-services doen geen review of kwaliteitscontroles op de inhoud van de artikels, weer je daarvan bewust. Voorbeelden zijn ArXiv, BiorXiv en MedrXiv.	(eigen definitie)
Pseudonimiseren	Pseudonimiseren is een methode waarbij identificerende gegevens met een bepaald algoritme worden vervangen door versleutelde gegevens (het pseudoniem). Het algoritme kan aan een persoon steeds hetzelfde pseudoniem toekennen, waardoor informatie uit verschillende bronnen kan worden gecombineerd. Na de encryptie is het dus nog steeds mogelijk om de betrokkene te identificeren: met behulp van het algoritme kan de versleuteling namelijk weer worden teruggedraaid. Omdat het versleutelingsproces omkeerbaar is, zijn gepseudonimiseerde data over een identificeerbare persoon volgens de Algemene Verordening Gegevensbescherming (AVG) nog steeds persoonsgegevens. Na pseudonimisering zijn de gegevens namelijk tot een persoon te herleiden.	https://www.fgsupport.nl/wat-is-anonimisering

Begrip	Definitie	Bronnen & referenties
PSI-richtlijn	Public Sector Information Richtlijn van de Europese Commissie, ook wel bekend onder de naam Open Data Richtlijn. Deze richtlijn bepaalt de voorwaarden waaronder gegevens uit de publieke sector moeten worden vrijgegeven (o.a. naar licenties toe).	Naar: https://overheid.vlaanderen.be/organisatie/informatiemanagement/omzetting-psi-richtlijn https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN
Semantische interoperabiliteit	Het bereik van samenwerking op het vlak van gebruikte termen (benaming) en betekenis van de uitgewisselde data. Zie ook 'Interoperabiliteit'.	(eigen definitie)
SSL-certificaat	SSL staat voor Secure Sockets Layer, dit is een extra beveiliging die je kan toepassen om je website veiliger te maken tegen datalekken. Websites met een SSL-certificaat hebben een URL die start met 'https://'.	Naar: https://nl.wikipedia.org/wiki/Transport_Layer_Security
Triple	Een semantische 'triple' is een set bestaande uit drie elementen waarin betekenis over data zit vervat. De vorm is (subject, predicaat, object). Het subject is in essentie het object dat beschreven wordt. Het predicaat beschrijft een eigenschap, kenmerk of relatie. Het object ten slotte is de waarde van dat kenmerk. Voorbeeld: ("Italië", "heeftAlsHoofdstad", "Rome")	Naar: https://en.wikipedia.org/wiki/Semantic_triple https://nl.wikipedia.org/wiki/Resource_Description_Framework
URI	Een uniform resource identifier (URI) is een internet-protocolelement. Het is een eenmalige benaming van een 'bron', een stuk informatie, data of dergelijke meer. URI's volgen het formaat protocol: //[gebruiker:wachtwoord@]host(naam)[:poort]/path	https://nl.wikipedia.org/wiki/Uniform_resource_identifier
Vendor lock-in	'Vendor lock-in' maakt een klant afhankelijk van een leverancier voor producten en diensten, omdat hij niet in staat is om van leverancier te veranderen zonder substantiële omschakelingskosten of ongemak. Toegepast op software betekent het bijvoorbeeld dat bestandsformaten of -protocollen incompatibel zijn met die van concurrerende software.	https://nl.wikipedia.org/wiki/Vendor_lock-in
Vocabulary	Een woordenschat, lexicon, vocabularium of vocabulaire is de verzameling woorden die door personen (binnen een domein) gebruikt worden.	Naar: https://nl.wikipedia.org/wiki/Woordenschat



Kapeldreef 75, 3001 Leuven

info@scivil.be

www.scivil.be



DEPARTEMENT
ECONOMIE
WETENSCHAP &
INNOVATIE