

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

ISBN 9789463965637

Depot n° D/2021/3241/104

Voorgeschreven 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	Thema
1. Principe	Principe
Afkortingen- of begrippenlijst	Term toegelicht in afkortingen- of begrippenlijst
Principe	Kruisverwijzing naar principe of thema

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	9
2. Publiceer je data onder een open licentie die je kiest uit een beknopte, aanbevolen lijst	9
3. Publiceer je onderzoeksresultaten en -bevindingen waar mogelijk in 'Gold Open Access Journals'	9
4. Publiceer, waar mogelijk, ook de software die je ontwikkelt onder open licenties	10
5. Ga actief op zoek naar bestaande open data	10
6. Vraag raad aan ondersteunende diensten	10

II. PRIVACY & ETHIEK

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

III. DATAHYGIËNE

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

IV. DATASTANDAARDEN EN -FORMATEN

17. Bouw verder op bestaande en voor jouw project relevante datastandaarden	14
18. Maak gebruik van machineleesbare, open formaten	14
19. Ken globaal unieke en persistente identificatoren toe aan je data	14
20. Incorporeer je projectdata in het netwerk van Linked Open Data	14

V. METADATA

21. Voorzie je data van zo rijk en accuraat mogelijke metadata	15
22. Denk eraan de metadata zo dicht mogelijk bij de bron te capteren	15
23. Ken aan je metadataset een globaal unieke en persistente identicator toe	15
24. Gebruik de juiste standaard voor je metadata	15
25. Maak je metadata, en dus je datasets en onderzoeksresultaten, vindbaar door ze te registreren op een doorzoekbaar portaal	16
26. Maak je citizen-scienceproject vindbaar voor collega's en vrijwilligers	16

Referenties	17
Afkortingenlijst	17
Begrippenlijst	18

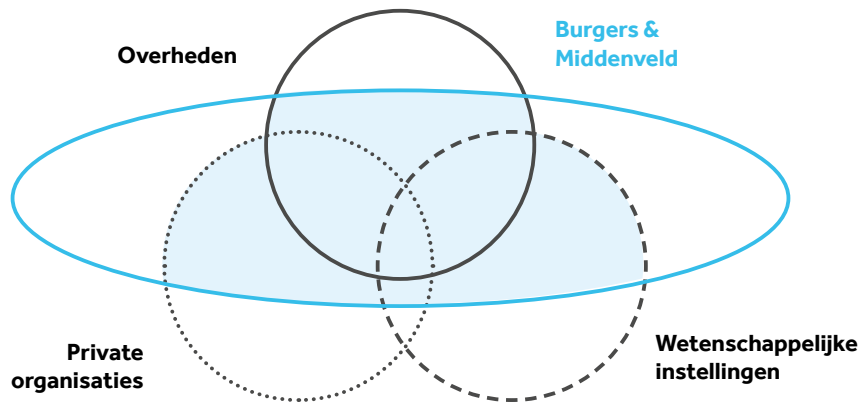


Inleiding

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. Zo zijn er 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 mogelijks niet welke richtlijnen te volgen, of weten überhaupt niet dát 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, overheiden, 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/>

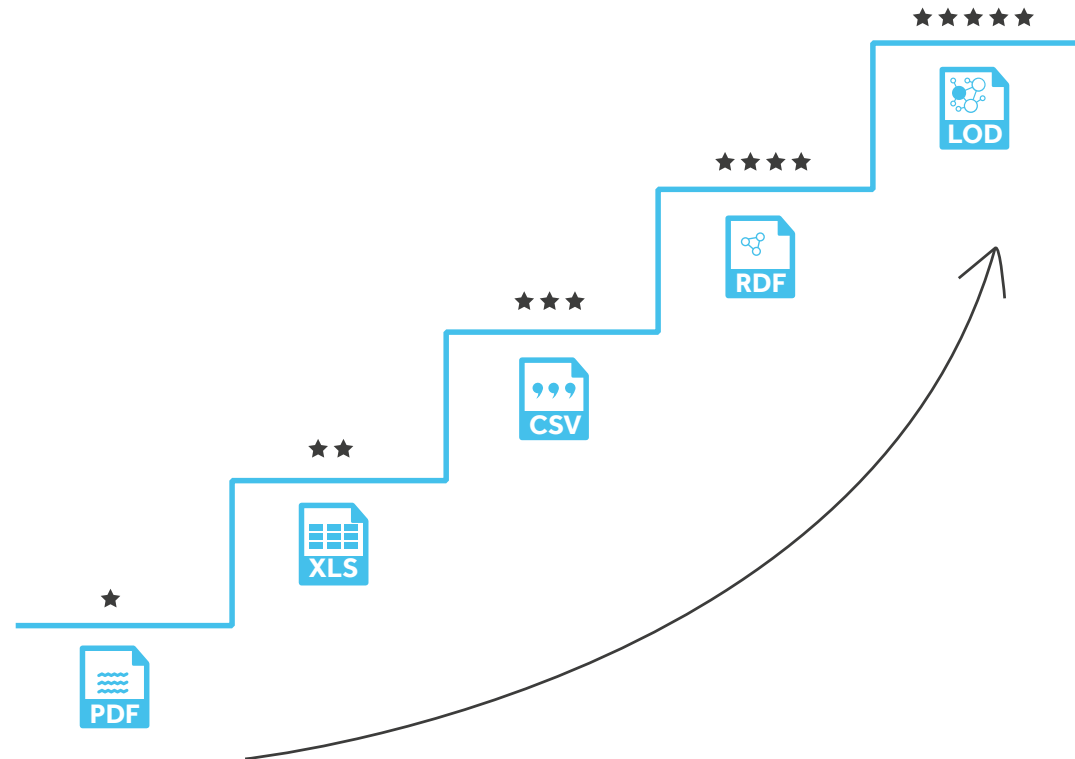
³ Wanneer burgers in de plaats van beleidsmakers initiatieven ontwikkelen, spreekt men over 'grassroots'-initiatieven.

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 oplist 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 klinkt 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.⁵ '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'.

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 vermenigvuldiging, 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.

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 je gegevens of kennis kunt gebruiken, hergebruiken of verspreiden. Ze vormen het legale kader voor de omgang met gegevens en voorkomen verwarring rond bv. intellectuele eigendomsrechten die er al dan niet op rusten. Gebruik licenties die open, internationaal bruikbaar en machineleesbaar zijn.

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.

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.



Open attitude

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.⁷ 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.

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**.

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 infrastructures, 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 Stewards of Research Data Officers van grote bedrijven en organisaties zijn hiervan op de hoogte: consulteer ook hen!

⁷ Lees meer over de voordelen van FOSS op <https://opensource.com/article/17/8/enterprise-open-source-advantages>.



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.

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

Respecteer de **GDPR**⁸-regels voor omgang met persoonsgegevens van zowel je deelnemende burgerwetenschappers, als van eventuele andere personen waarvan gegevens (on)opzettelijk gecapteerd worden tijdens je project.

⁸ General Data Protection Regulation of GDPR, in het Nederlands ook de Algemene Verordening Gegevensbescherming (AVG).

II. Privacy & ethiek

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.

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.

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.

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.



III . 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**).

14. Hou rekening met het cumulatiepotentieel 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 gedocumenteerde⁹ 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.

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**).

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 wordt.¹⁰ 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.

⁹ In termen van **metadata**, gebruikte **datastandaarden**, of datakwaliteit.

¹⁰ 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**.

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.

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 organisatie¹¹, 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**'.

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 raadplegen.¹²

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.

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.

¹¹ Merk op dat een formaat open gedocumenteerd kan zijn, maar nog steeds *propriëtair* wegens onderhevig aan licenties.

¹² 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.

V. Metadata



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

Metadata betekent 'data over data'. Ze vormen de 'bijsluiters' 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.

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

Als projectleider van een citizen-scienceproject zorg je ervoor dat alle nodige metadata gecaptureerd 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.

23. Ken aan je metadataset een globaal unieke en persistente identificator 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 identificator. 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.

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 Core¹³, 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.

¹³ Opgesteld tijdens een workshop in 1995 in <https://www.wikidata.org/wiki/Q988937>.

V. Metadata

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.

26. *Maak ook 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.

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>

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
CS	<i>Citizen Science</i>
DMP	<i>Datamanagementplan</i>
ECSA	<i>European Citizen Science Association</i>
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
GDPR	<i>General Data Protection Regulation</i>
ISO	<i>International Organization for Standardization</i>
LOD	<i>Linked Open Data</i>
OSLO	<i>Open Standaarden voor Linkende Organisaties</i>
PSI	<i>Public Sector Information</i>
PURL	<i>Persistent Uniform Resource Locator</i>
URI	<i>Uniform Resource Identifier</i>

Begrip	Definitie	Bronnen & referenties
Codelijst	<p>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.</p> <p>Voorbeeld:</p> <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)
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>	https://www.scivil.be/aan-de-slag/voor/citizen-science-iets-voor-mij
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>	https://www.ugent.be/nl/onderzoek/datamanagement/datamanagementplan

Begrip	Definitie	Bronnen & referenties
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>
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>	<p>(eigen definitie)</p>
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 onderzoeksecosysteem 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>

Begrip	Definitie	Bronnen & referenties
FAIR-datamanagement	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.	https://datasupport.researchdata.nl/start-de-cursus-/i-in-vogelvlucht/begrippen/rdm-open-en-fair
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
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)

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 bereiken van samenwerking op het vlak van gebruikte termen (benaming) en betekenis van de uitgewisselde data. Zie ook 'Interoperabiliteit'.	(eigen definitie)
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](https://www.kapeldreef75.be) info@scivil.be www.scivil.be

DIGITAAL
VLAANDEREN



Vlaamse
overheid en



DEPARTEMENT
ECONOMIE
WETENSCHAP &
INNOVATIE