Guide to Data Charter for Citizen Science

v1.0

A basic set of principles to support open and interoperable citizen-science data

Flanders
State of
the Art

SCivil
Citizen Science
Vlaanderen

Contributors

Name	Organisation	Involved in project	Role
Mieke Sterken	*Citizen science advisor,* Scivil & KULeuven	BRITEC, MoMster, CS Roadmap for Local Authorities	Author
Ruben Capiau	*Business analyst,* Digitaal Vlaanderen	Flemish basic registers, OSLO, INSPIRE, digitalisation projects	Author, chair of WG
Elien Dewitte	*Science officer,* VLIZ	SeaWatch, Grote Schelpenteldag	Acknowledged contributor, WG participant
Sien Benoit	*PhD student,* UGent	Cycle barometer	Acknowledged contributor, WG participant
Tim Adriaens	*Science officer,* INBO	Vespa-Watch	Contributor, WG participant
Christophe Stroobants	*Team manager,* VMM	CurieuzeNeuzen, Clean Air Project	Contributor, WG participant
Kristijn Swinnen	*Senior researcher,* Natuurpunt	waarnemingen.be, Mijn Tuinlab	Contributor, WG participant
Maarten Trekels	*Data scientist,* Meise Botanic Garden	Groene Pioniers	Contributor, WG participant
Nils Walravens	*Senior researcher,* IMEC-SMIT-VUB	Open Data Charter 'Smart Flanders'	Contributor, WG participant
Jef Van Laer	*Citizen science advisor,* Scivil	Support for all EWI citizen-science projects	Contributor, WG participant
Annelies Duerinckx	*Scivil coordinator,* Scivil	CS Roadmap for Local Authorities	Contributor, WG participant
Mathias De Schrijver	*Information policy advisor,* Digitaal Vlaanderen	Flanders Open Data Portal, Administration Decree	Reviewer
Tom Callens	*Information policy advisor,* Digitaal Vlaanderen	Flanders Geographical Data Infrastructure, INSPIRE	Reviewer
Geraldine Nolf	*Product owner,* Digitaal Vlaanderen	Metadata Vlaanderen, Data Publication process	Reviewer
Delfien Vandyck	*Associate professor,* UGent	Cycle barometer	Reviewer
Ziggy Vanlishout	*Programme manager,* Digitaal Vlaanderen	Authentic data sources, Flemish basic registers, OSLO	Acknowledged
Pieter Colpaert	*Postdoctoral researcher,* UGent-IDLab-imec	Public web APIs	Acknowledged
Philippe Michiels	*Technical & data architect,* imec City of Things	Flanders Open City Architecture (VLOCA)	Acknowledged
Data Management Working Group	*Controlling body,* Vlaams Informatie- en ICT-beleid	N/A	Acknowledged

Licence

This document, the 'Data Charter for Citizen Science', prepared by the data management working group of Scivil (*https://www.scivil.be/*), is published under a **CC BY- 4.0** licence. Visit *https://creativecommons.org/licenses/by/4.0/* to read the licence text.

Layout: Femke Vanhellemont (*lafem.*)

File No D/2021/3241/206

Prescribed manner of naming (*attribution*):

Legend

| Theme	Theme
1. *Principle*	Principle
	Guide to application, tips & tricks, tools, etc.
	Practical example
List of abbreviations or terms	Term explained in list of abbreviations or terms
Principle	Cross-reference to principle or theme

Contents

Introduction

With the Data Charter for Citizen Science, Scivil, the Flemish knowledge centre for citizen science (CS), provides a guide to achieving well-documented data in a citizen-science project.

The charter is for everyone who wants to start or participate in a citizen-science project, i.e. both project initiators and participants (citizen scientists) from all sectors of the so-called 'quadruple helix' (citizens, public authorities, scientific institutions, private organisations).

A citizen-science project is any project that attempts to gather knowledge or insights in a scientifically responsible manner, together with or by citizens.

A data-guide made for Flanders, translated to English

This roadmap was originally written in Dutch for a Flemish audience in a Flemish context. It has now been translated to English, for European and wider dissemination. This is why you will come across a lot of Flemish examples of citizen science, as well as referrals to Flemish and Belgian political, social and educational structures. Dutch names of publications, projects and institutions are, by default, not translated.
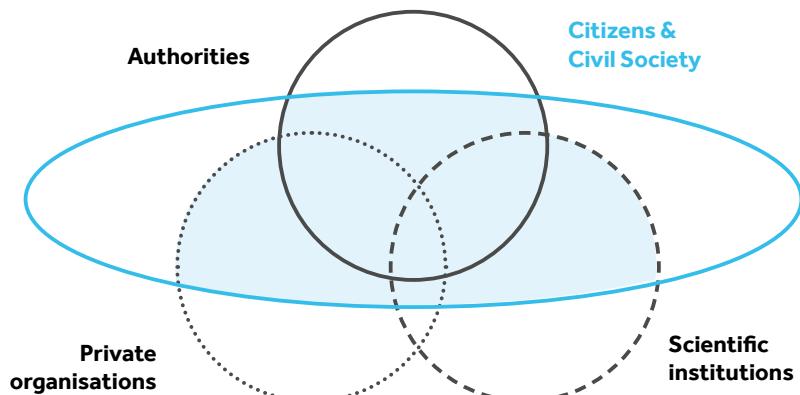
Among other things, keeping citizens involved, protecting their privacy, and the many possible research methods and data types make it difficult to formulate a set of universal rules for dealing with data.

With this Data Charter for Citizen Science, Scivil wants to offer an overview of the most important data-related recommendations that you should also aim for in a citizen-science project. We base this on the FAIR data principles, the Open Data Charter (Smart Flanders) and the 10 principles of the European Citizen Science Association **(ECSA).** The charter aims to bring overview and relief to all aspects related to data in citizen science.

A basic principle here is to strive for the publication of data that is 'as open and **interoperable** as possible'. More than anywhere else, the added value of citizen science can be increased exponentially by systematically adhering to the principles in this Charter. Through open and clearly defined data, both large and small citizen-science projects can eventually lead to important scientific insights, as their results can be used and combined with datasets from other citizen-science projects, governments, research institutions or even the 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]**

 In both academic and governmental contexts, there are initiatives around the world to make general data traffic more streamlined, efficient and secure. The international **FAIR** principles[1] for scientific data (Wilkinson et al. 2016) in the academic world. At Flemish level, there are the Open Standards for Linking Organisations **(OSLO)** (Buyle et al. 2016) and the Open Data Charter[2] of Smart Flanders; they are designed to ensure that organisations make their data more open and linked, in accordance with European guidelines (for example the INSPIRE guidelines for geographical data, and the - of the European Commission).

Citizens or initiators of citizen-science projects may not know which guidelines to follow, or that there are regulations at all. After all, both academics and government bodies, civil-society organisations, private companies or citizens[3] can start a citizen-science project, whether or not in partnership and whether or not sponsored by a private partner, academic partner or government agency. In addition to the tangle of existing advice and regulations from various societal regulators, citizen science also involves other challenges that may affect the handling of data.

Note

obtaining '5-star open data'

Globally, there is a growing trend towards publishing '5-star open data', also known as 'Linked Open Data' (LOD).
This involves structuring and defining data in such a way that you can easily link it digitally (within the boundaries of privacy and security regulations) and in which data exchange becomes easier. Sir Tim Berners-Lee, the inventor of the world wide web, translated this so-called interoperability into a graduated scale (see Figure 2).
The higher you are on the scale, the easier it is for others to reuse your project data.
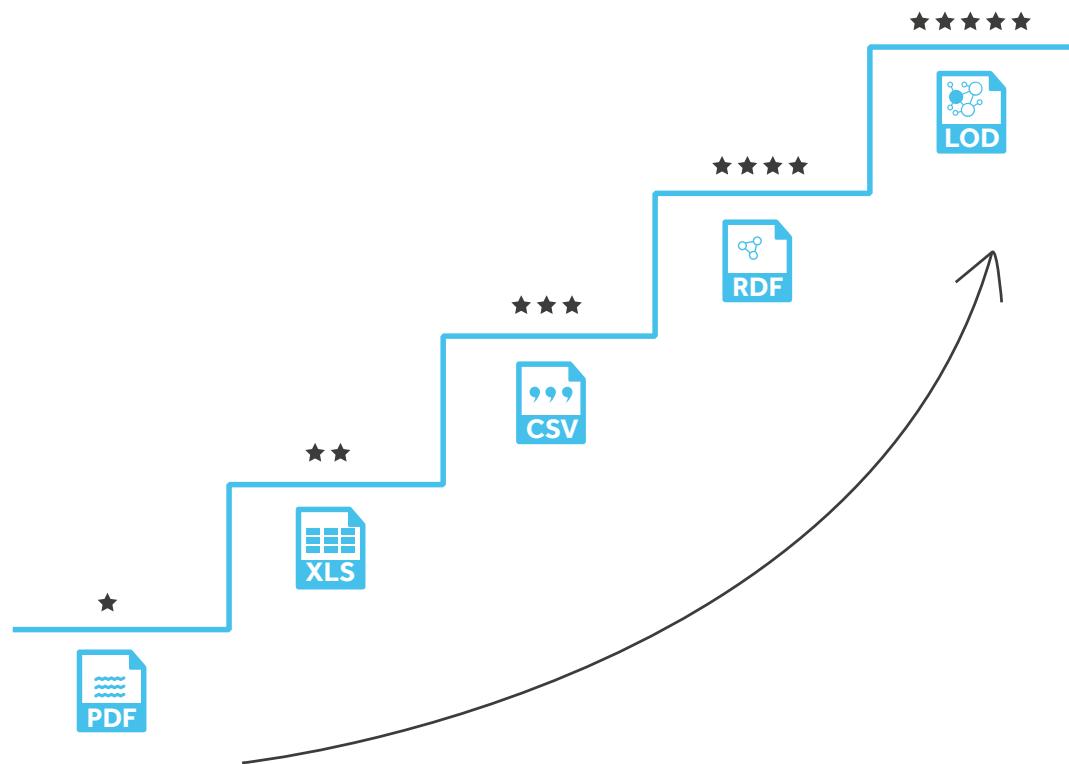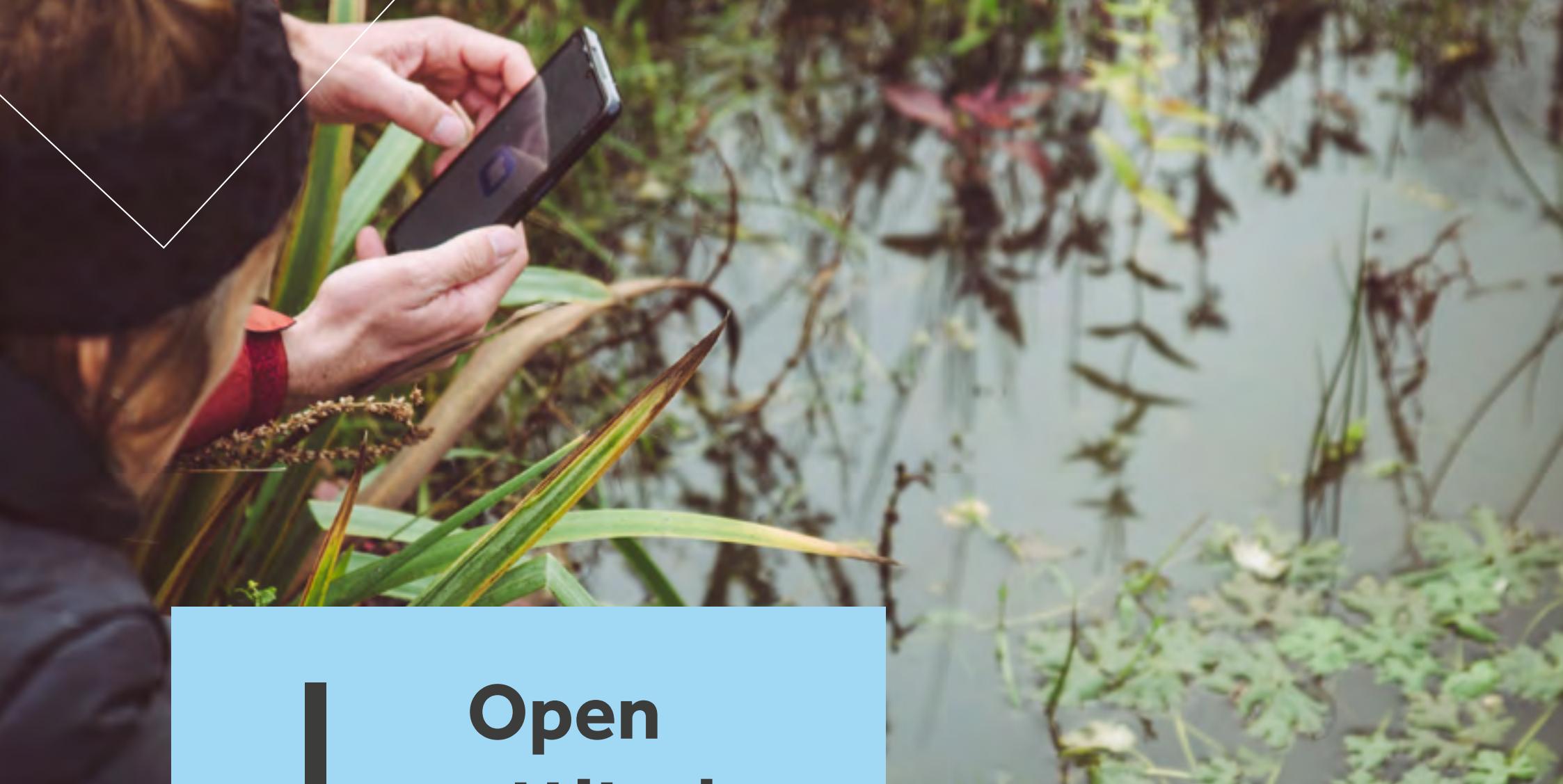

Figure 2
'5-Star Open Data'.

If you follow all the principles of the Data Charter for Citizen Science, you will get 5-star open data. The table below illustrates which principle you have to satisfy as a minimum in order to earn an extra star, and what this means in concrete terms or for the users of the project data concerned.

Stars	Achieved after ...	Meaning
★ offered as open data	*principle 2*	You understand that you cannot lock data away in silos and that data only really pays off when others can access and reuse it. So you choose to share your project data with the outside world under an **open licence**. It is possible that you are still offering data in a suboptimal, non-machine-readable form (e.g. a scan of a table with results or a PDF with a list of measurement points): for this reason, you (only) receive one star for the time being.
★★ structured and machine-readable	*principle 18*	You offer your project data not only under an open licence, but also in a **structured (and therefore machine-readable) way.** You put text in a Word file, figures in an Excel file and your SLR camera delivers high-quality RAW images. In doing so, you take the step towards '2-star open data'. Users can now visualise the data or make calculations with it. But do they have the right software?
★★★ stored in an open data format	*principle 18*	By choosing an **open data format**, users no longer need to use or buy specific software to work with your data. Interested parties can now open your text and spreadsheet documents in a (free) software package of their choice without any problems and without any data loss. Your photos no longer give the error message 'unknown file format' when a colleague opens them on his or her computer. This turns your data into highly reusable '3-star open data'. Now suppose that a biology student wanted to link his own research data to observation '100' from your dataset (sighting of a wild swan in the Blaarmeersen on 2021-01-10T15:23:22). However, he confused this observation with observation '100' from another, European dataset (sighting of a griffon vulture in the Pyrenees on 2017-07-05T06:10:05) and thus drew the wrong conclusions. How can we prevent this in the future?
★★★★ unique identifiable	*principle 19*	With your project data you are now leading the pack. From now on, anyone can unambiguously link to an individual data point in your dataset. Confusion between *https://waarnemingen.be/id/observatie/100* and *https://nature-observatory.org/id/observation/100* is no longer possible thanks to the **globally unique, persistent identifiers**. The meaning of the collected data can also be found out by other systems, even fully automatically, via the URIs of the data concepts. Your dataset can now be included in the Linked Open Data (LOD) network.
★★★★★ linked to external sources	*principle 20*	Your project data is now open, structured, readable by various applications, uniquely identified and **linked to external sources**. Users can explore other LOD sources from your enriched dataset and thus get more context to your data. It is possible to use query languages such as SPARQL to perform intelligent queries on your dataset in combination with other sources. Your project data is now fully-fledged '5-star open data'!
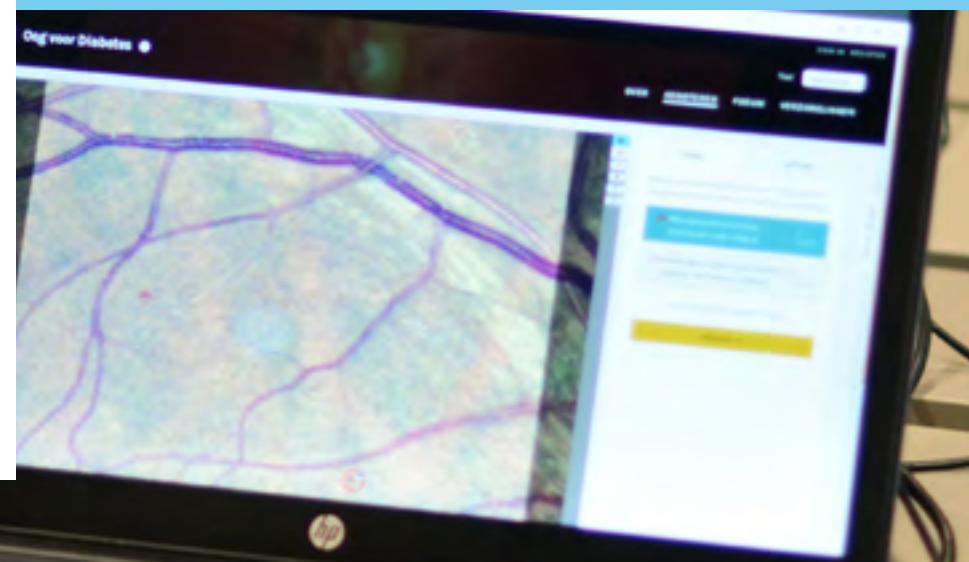
I. Open attitude

1. *Aim to publish your data openly on the web, or give a clear and well-founded reason if this is not possible.*

'Open' means that anyone can freely access, use, modify and share your data. At most, there are conditions to safeguard the origin and openness of your data. 'On the web' implies that you can access the data through a web browser.

In practice, this means releasing your data under an **open licence** (*see principle 2)*. If this is not possible, explicitly state the reason for not opening up (some) data. In such cases, try to make at least some of the data (for example **metadata**, *i.e. information about the data)*open. The starting point is therefore 'as open as possible, as closed as necessary'.

NB: 'open data' does not necessarily mean making the data available free of charge or immediately online. Except for research data and high-quality datasets that fall under the European Open Data Directive[6] (see Articles 10 and 14 of the **PSI Directive**), in principle, marginal costs can be charged for duplication, distribution and dissemination. Furthermore, it may be necessary to request data explicitly. However, these practices are discouraged because they create an additional barrier to the reusability of data.

[6] *Also known as the Public Sector Information Directive or PSI Directive: https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN. Article 10 of this directive defines the conditions under which research data should be available for reuse, Article 14 of the directive sets out the arrangements for publication and reuse of high-quality datasets.*

What data are you sharing?

∟ If possible, the raw, unprocessed data will be made available as open data. Preferably, the post-processed data, research results and insights are also shared 'openly'.

∟ In some cases, data cannot be released under an open licence because there is a good reason for it. For example, the data maybe sensitive (e.g. location of endangered species, entered through Natuurpunt's www.waarnemingen.be), used to guarantee income for the maintenance of the dataset in which it is published or not disclosed because of privacy protection, confidentiality or predefined copyrights. Aggregating data before publishing can be a solution when the unprocessed data is privacy-sensitive (see also ***principle 7)***.

∟ Research data that falls under the European Public Sector Information Directive (**PSI Directive, Article 10**) (if publicly funded and made public via an institutional or thematic database) should be made available openly and free of charge from 17 July 2021. As soon as the recently amended directive is transposed into the Flemish regulations, more information will be available on the central website of the Flemish government[7].

Why this principle?

A *Ethics*: citizen science is a form of research par excellence that relies on the voluntary cooperation of citizens. It is therefore right to give them access at least to the data that they have generated or analysed themselves (see ***principle 12***).

B *Linked data*: in addition, in citizen science it is at least as important as in regular scientific research to make generated datasets open and linked, because 'many small bits' (i.e. a number of measured values or a small data set from which conclusions can only be drawn within the scope of the project) will over time together produce a unique, larger dataset from which additional conclusions can be drawn (see ***principle 14***).

C *Legislation* imposes it: imposes it according to the **PSI Directive**

How will it be applied?

1 For each raw or post-processed dataset, determine whether it (a) can represent added value for other uses, and (b) can/must be published openly.

2 Provide the open data with a licence from the shortlist in ***principle 2***.

3 Make the open data accessible through a web browser, i.e. Place it behind an http(s) or (s)ftp link.

4 Describe briefly which data will not be opened up and what the reasons are for this.

---

[5] freely translated to http://opendefinition.org, https://okfn.org/opendata/ and https://opendatahandbook.org/guide/nl_BE/what-is-open-data/
[6] Also known as the Public Sector Information Directive or PSI Directive: https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN. Article 10 of this directive defines the conditions under which research data should be available for reuse, Article 14 of the directive sets out the arrangements for publication and reuse of high-quality datasets.
[7] https://overheid.vlaanderen.be/organisatie/informatiemanagement/omzetting-psi-richtlijn

Practical example

The VLINDER[8] (VLaanderen IN DE weeR) project of Ghent University consists of a network of 60 weather stations that monitor various weather parameters (temperature, precipitation, air pressure, air humidity, wind speed) at research institutions, local partners and schools all over Flanders. The Leuven.cool[9] project (KU Leuven, KMI, klimaatvzw Leuven2030) is also researching urban microclimates using sensors in the city and in citizens' gardens. Both projects link their data to the crowdsourcing platform WOW-BE[10] and make it openly available there as well. For each weather station, anyone can directly download the data in different formats (e.g. SVG, PDF, XLSX, CSV). The terms of use and information on licences are not (yet) explicitly or clearly stated, but can be consulted on the website of the UK Met Office[11].

The Nieuws van de Groote Oorlog[12] project of Het Archief[13] uses newspaper images from the First World War. However, some of the texts or pictures in those newspapers may still be protected by copyright, so they cannot simply be published as 'open'. To be able to publish them as '**Linked**', only the metadata of the newspapers is available under an open licence.

The actual texts were then published as linked data[14] to which the conditions of use apply.
For example, the newspaper articles can be linked to the online List of Names[15] of war victims (military and civilian) of the 'In Flanders Fields' museum[16]. Although these databases are not the result of citizen-science projects, they can serve as a source for citizen-science projects in the future.

The citizen-science project Vespa-Watch[17], financed by the Department of Economy, Science and Innovation of the Flemish Government, also systematically releases its data as 'open'. The modalities of that openness are described in the open data policy[18] of the Research Institute for Nature and Forest[19] (INBO), which is leading this project.

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 The website of Meemo, the Flemish Institute for Archives, contains extensive information on Linked and 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/
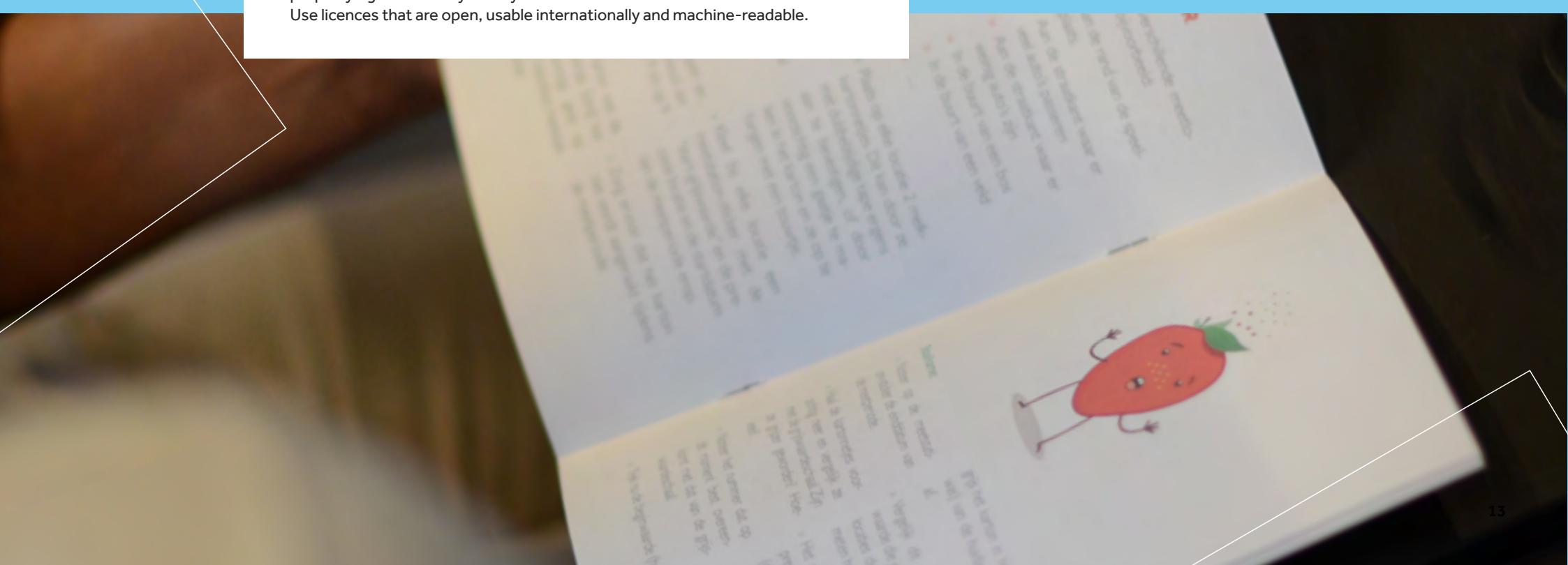
2. *Publish your data under an open licence that you choose from a short, recommended list.*

Licences are essentially a set of rules that describe the conditions under which data or knowledge can be used. They provide the legal framework for handling data (of any kind) necessary to avoid confusion about e.g. intellectual property rights that may or may not be attached to data or information. Use licences that are open, usable internationally and machine-readable.

At http://opendefinition.org/licenses/ you will find the complete list of licences that fit the definition of 'open' as laid down by the Open Knowledge Foundation.[20] According to Groom et al. (2017), the most commonly used licences in citizen science are those of Creative Commons CC (mainly CC0, CC-BY, CC-BY-NC) and/or Open Data Commons ODC (mainly PDDL, ODC-BY, ODbL). The CC licences CC-BY-SA and CC-BY-NC-SA can also be used.

The principle of the CC licences is that they consist of a simple combination of four sharing restrictions:

∟ **'BY'** means that reference to the source must be made by means of an acknowledgement;

∟ '**NC'** (non-commercial) means that data may not be used for commercial purposes;

∟ **'SA'** (share-alike) indicates the permission to create derivative works on condition that they are released under the same licence as the original material;

∟ **'ND'** (no derivatives) implies that the original material may not be adapted or used as a basis for derivative works. However, this restriction applies primarily to works of art and (written) media, and is therefore not considered here.
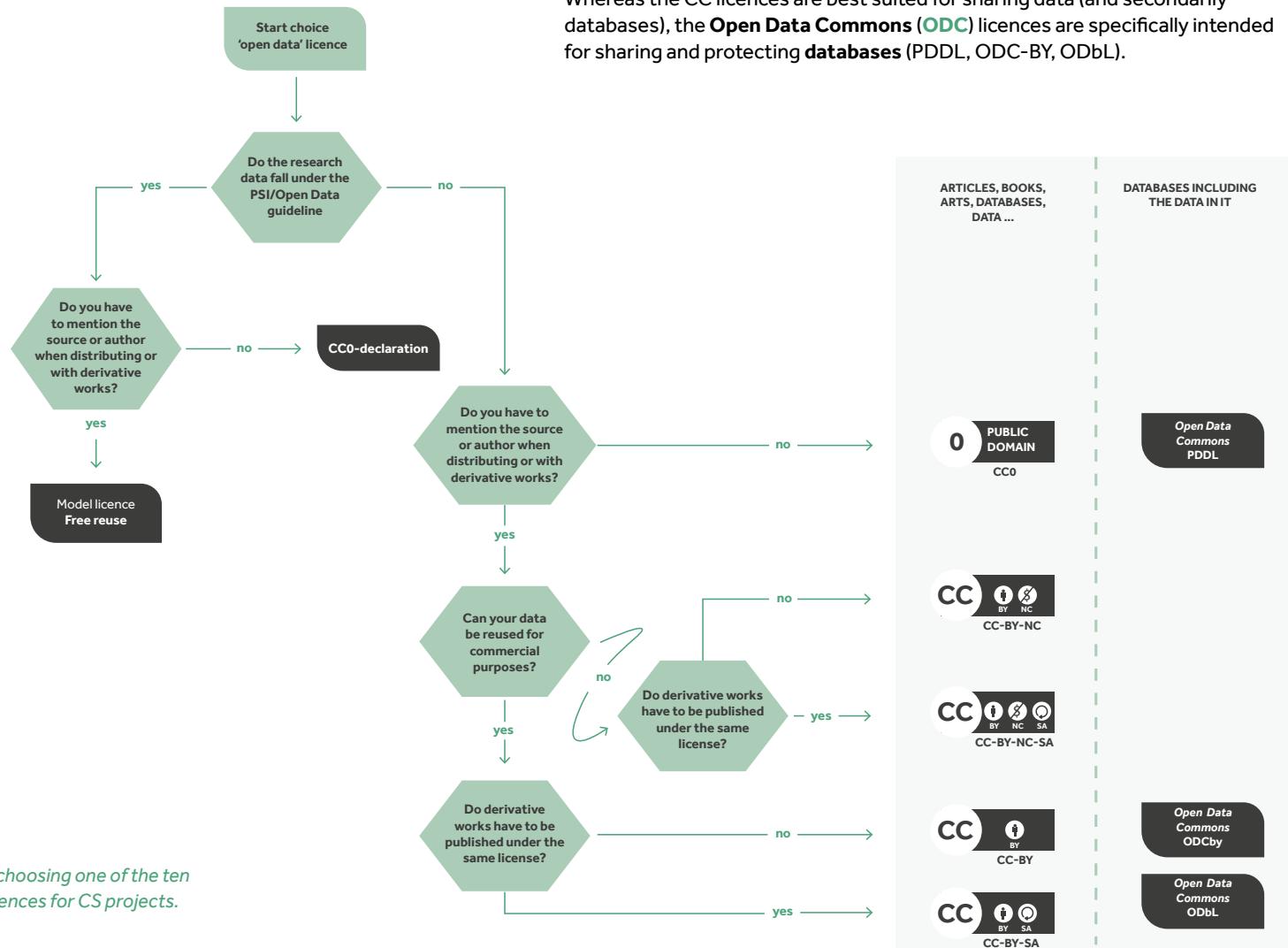
There is also a public domain statement (CC0) that does not impose any restrictions.

> **Tip**
>
> Step-by-step help with 'putting together' the most suitable CC licence can be found at https://creativecommons.org/share-your-work/.

---

[20] https://okfn.org/

Whereas the CC licences are best suited for sharing data (and secondarily databases), the **Open Data Commons** (**ODC**) licences are specifically intended for sharing and protecting **databases** (PDDL, ODC-BY, ODbL).

Figure 3
Decision tree for choosing one of the ten recommended licences for CS projects.

Does your research data fall under the **PSI/Open Data Directive**[21]?
Then use:

⌐ CC0 statement
https://overheid.vlaanderen.be/cc0-verklaring, of

⌐ 'Free Reuse' model licence
https://overheid.vlaanderen.be/modellicentie-gratis-hergebruik

Below are the other eight licences:

⌐ CC0 "Public domain"
https://creativecommons.org/share-your-work/public-domain/cc0/

NB: conventions regarding scientific source citations are
retained.

⌐ CC-BY "Acknowledgement"
https://creativecommons.org/licenses/by/4.0

⌐ CC-BY-NC "Acknowledgement-NonCommercial"
https://creativecommons.org/licenses/by-nc/4.0

⌐ CC-BY-SA "Acknowledgement-ShareAlike"
https://creativecommons.org/licenses/by-sa/4.0

⌐ CC-BY-NC-SA "Acknowledgement-NonCommercial-ShareAlike"
https://creativecommons.org/licenses/by-nc-sa/4.0

⌐ ODC-PDDL (Public Domain Dedication and License)
https://opendatacommons.org/licenses/pddl/summary/

⌐ ODC-By (Attribution License)
https://opendatacommons.org/licenses/by/summary/

⌐ ODC-ODbL (Open Database License)
https://opendatacommons.org/licenses/odbl/summary/

Practical example

CC licences within EWI citizen-science projects

⌐ The 'Vespa-Watch' project (EWI call 2017) publishes its data
under the CC0 licence, as it uses the platform iNaturalist. The data from
iNaturalist is linked to the worldwide open data platform GBIF[22] (Global
Biodiversity Information Facility; see also *principle 5*). With GBIF the inter
national agreement applies that linked data is published under a CC0,
CC-BY or CC-BY-NC licence, or released under an equivalent licence,
preferably CC-BY (see ''GBIF Terms of Use'[23]).

⌐ The 'AstroSounds' project (EWI call 2019) uses a
'CC-BY-NC-SA' licence.

Tip

A large number of existing platforms for uploading observations use
specific licences. As a participant in a citizen-science project, always
check under which licence the data you enter will fall. Conversely, it is
important for coordinators of citizen-science projects to give this
information clearly to all contributors at the start of the project.

---

3. *Publish your research results and findings where possible in Gold Open Access Journals.*

If you publish data, you must also provide the context and meaning of that data: it is important to make available reliable information about the research objective, the methods used, the results and their possible significance.
As a scientist, you should preferably publish your findings in Gold Open Access Journals, which are peer-reviewed and open to everyone. Is an (open) peer-reviewed article not a viable option? Then describe your research in a thorough report, which you make available online under an open licence (see *principle 2*).

Academics mainly use the system of '**peer-reviewed**' publication. When a research article is sent to a journal, (usually anonymous) external scientists thoroughly review the article, after which they either approve it, reject it or approve it with minor or major modifications. It is important to ensure that articles claiming to publish 'research' are published in peer-reviewed journals. These are considered in scientific circles to be the most reliable articles, especially if they are published in journals included in the Web of Science[24] database (so-called 'A1 articles'). They also all have a **DOI number**, and, for each statement that is not directly the result of the own research, contain a reference to another, previously published article or reliable source.

More and more universities and government institutions are publishing their scientific articles in '**Open Access**'. These articles are not hidden behind paywalls or copyright restrictions. 'Open Access' for publications means freely available on the internet, so that any user can read, download, copy, share, print, search or link the publications without financial, legal or technical restrictions. The only conditions are related to the correct citation of authors and the preservation of research integrity (Budapest Open Access Initiative, 2002[25]).

Publications can be made Open Access in two ways:

⌐ by archiving documents in public archives
 (= 'Green Open Access')

⌐ by publishing the article directly in Open Access journals
 (= '**Gold Open Access**').

Public archives can be institute-bound (e.g. by research institute such as UGent[26] or INBO[27], which keep the published articles of their own scientists online), or topic-related (e.g. via **preprint services** such as ArXiv[28]).
An international list of public archives can be found on 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/opendoar/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/*

The second way (that of 'Gold Open Access'), which is preferable, is via **Open Access Journals** (OAJ). These are journals that make the published articles freely available and free of charge to everyone on the internet. Well-known Open Access publishers are Public Library of Science[30] and BioMed Central[31]. The group around the renowned scientific journal Nature also has an Open Access journal under its wings: Scientific Reports.[32]

A list of Open Access Journals can be found in the Directory of Open Access Journals.[33] This website provides access to more than 15,000 Open Access scientific papers worldwide, with more than 5 million peer-reviewed articles.
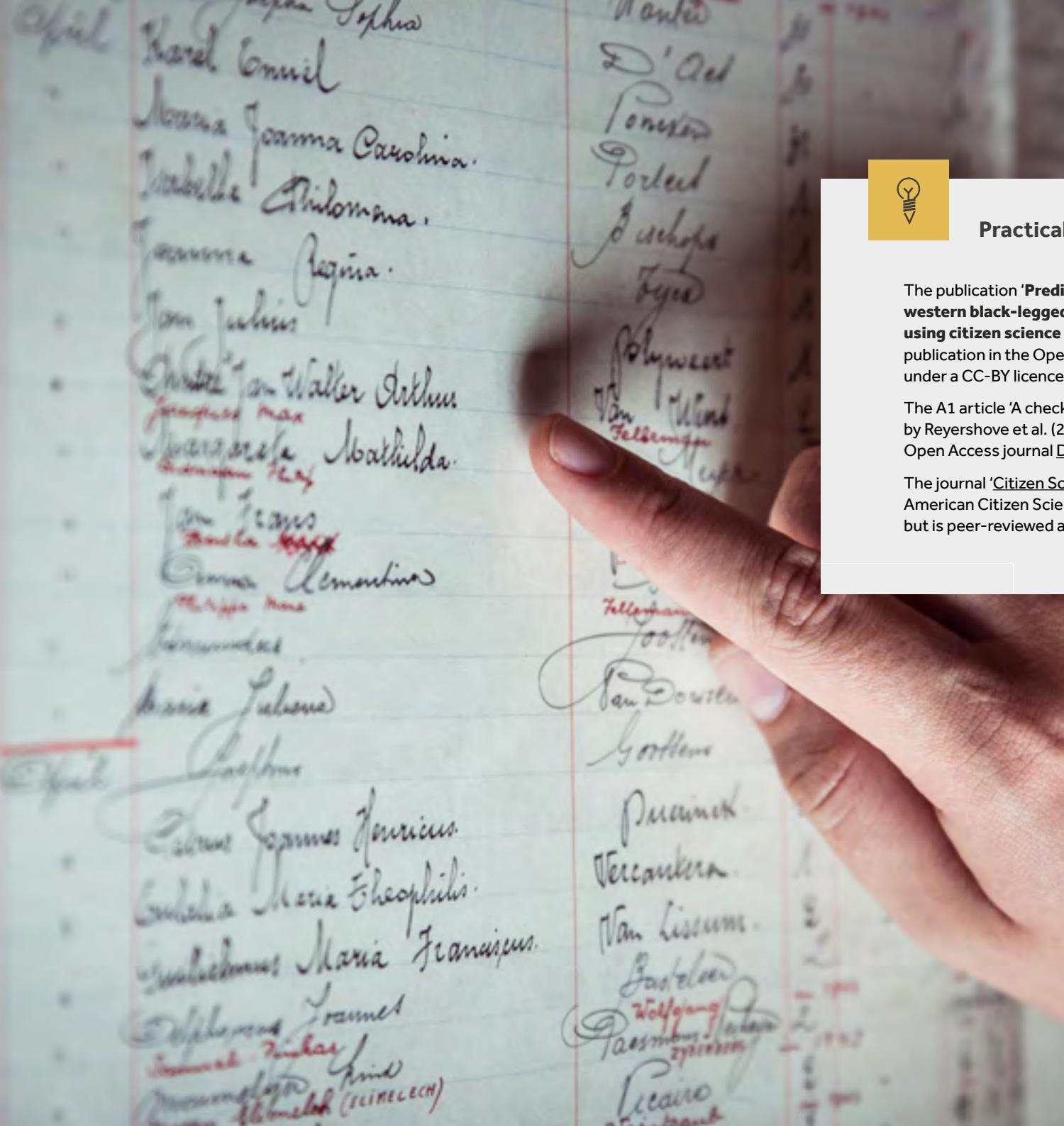
> **Tip**
>
> For schools, the very first peer-reviewed scientific journal in Europe was recently launched: The Open Schools Journal for Open Science. The journal accepts scientific articles written by pupils from both primary and secondary schools, who produce scientific data under the guidance of their teachers, e.g. through a citizen-science initiative. The journal is published in several languages and is also free of charge.

> **Tip**
>
> Is your project being funded by the European Horizon Europe programme? Then you can publish your research in the hew Open Research Europe publication portal.[34]

Is your research not strong enough to get through the review process of a scientific journal? In that case, make a clear, comprehensive report with (a) the research objective, (b) an introduction that frames the research in a broader context, (c) the methods used to collect the data, including the procedures followed to safeguard good data quality, (d) the results themselves, possibly in an Annex, (e) the statistical analyses of the results and/or the interpretation of the results [this is called 'the discussion'], and then (f) your conclusions from the research. Such a report helps to demonstrate that the data collected from your project is reliable and allows others to make their own assessment of the reliability of your data. This increases the likelihood that others will also use your data.

Practical example

The publication '**Predicting the current and future distribution of the western black-legged tick, Ixodes pacificus, across the Western US using citizen science collections**' by Porter et al. (2021) is an A1 publication in the Open Access journal PLOS ONE[35], which publishes under a CC-BY licence.

The A1 article 'A checklist recipe: making species data open and FAIR' by Reyershove et al. (2020) is also published under CC-BY licence in the Open Access journal Database.[36]

The journal 'Citizen Science, Theory and Practice'[37] of the American Citizen Science Association (CSA[38]) is not an A1 journal, but is peer-reviewed and Open Access.

4. *Where possible, also publish the software you develop under open licences.*

Sometimes software is developed for or by a citizen-science project.
It is strongly recommended that this source code also be published openly.
There are a number of good reasons for making the underlying code open
source.[39] It allows others, for example, to contribute to the development and
testing of the application. The open source code also offers full transparency
on how the data is handled. Open-source applications also attract a
community of interested parties who want to contribute to the solution.
You can also, in turn, use open-source libraries that others offer.

[39] *Read more about the benefits of FOSS at https://opensource.com/article/17/8/enterprise-open-source-advantages.*

There are a number of well-known websites[40] **on which you can publish source code openly:**

└ GitHub

e.g. https://github.com/opendata-stuttgart/sensors-software (scripts and applications), https://github.com/opendata-stuttgart/luftdaten-local-labs (website with JavaScript) and https://github.com/dylanrees/citizen-science (sources)

└ GitLab

e.g. https://gitlab.com/micro-iot-platform/projects/temp-humidity-monitor

└ BitBucket

There is an Open Source Project License offered by parent company Atlassian: https://www.atlassian.com/software/views/open-source-license-request.

└ SourceForge

A popular website for open-source projects: https://sourceforge.net/.

└ Google Code

Not so much for publication, but worth mentioning because Google offers interesting code libraries openly via https://opensource.google/projects(e.g. code for data analysis or visualisation).

The most common licences[41] **for open software (code) are:**

└ MIT License

https://choosealicense.com/licenses/mit/
A short and simple licence that allows others to do just about with your software (code), including creating and distributing closed-source versions, under a different licence, not necessarily releasing the source code.

└ Apache License 2.0

https://choosealicense.com/licenses/apache-2.0/
As with the MIT License, derivative works may be distributed under a different licence, and if necessary without the source code. Compared with MIT, this licence builds in a few more protection clauses (e.g. concerning patents) and is therefore better suited for larger software projects.

└ GNU General Public License v3.0 (GPLv3)

https://choosealicense.com/licenses/gpl-3.0/
This licence allows just about everything, except for the creation and distribution of closed-source versions. The source code must always be disclosed. Derivative works must be provided under the same licence, and it must be clear what changes were made.

More details and explanations on the choice of the most appropriate licence can be found at https://opensource.guide/legal/.

---

[40] based on *https://itsfoss.com/github-alternatives/*
[41] based on *https://opensource.guide/legal/*

Practical example

In the NoiseTube project[42] (Vrije Universiteit Brussel), people are offered an app with which they can monitor noise (dB) with their smartphone. The project published the source code for the software used openly on the server of the VUB[43], under a GNU-LGPL 2.1. licence. The project is also listed in Google Code[44], which makes the source code easier to find (by analogy with *principle 25*, but for the software used in the project).

Also, in the VLINDER project (Ghent University), the code[45] behind the dashboard[46] is made available on Github.

The crowdsourcing platform DoeDat[47] (Meise Botanic Garden) extends the interface of the open-source platform DigiVol[48] (The Australian Museum & Atlas of Living Australia); the DoeDat interface offers multiple languages, for example. DoeDat has made its own code[49] available through Github so that other institutes can also set up a multilingual version of DigiVol.

In the citizen-science project Vespa-Watch[50] (INBO & Honeybee Valley), observations of Asian hornets were registered and synchronised with the platform iNaturalist.[51] The final report of the project (Schoonvaere et al. 2020) clearly describes that the application uses open-source software and frameworks, and was also developed as open-source software itself. The source code is available at https://github.com/inbo/vespa-watch. A functionality to synchronise data with the iNaturalist platform was also developed, which was developed as a standalone open-source module (Python library) and is available at https://pyinaturalist.readthedocs.io/.

The European Sensor.community[52] network also offers its software[53] for reading air quality sensors openly on Github, as does the Zooniverse platform[54], which is under the open Apache 2.0 licence.[55]

[42] NoiseTube - http://www.noisetube.net/help (website) and https://code.google.com/archive/p/noisetube/ (source code)
[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. *Actively search for existing open data.*

More and more data is available as open data on the web. Actively search for existing open data relevant to your project. You may save yourself a lot of time and money, and it can greatly enrich the content of your project. Do use reliable sources, though, preferably ones compatible with current data standards (see *principle 17*). The government also pursues an active open data policy: so make use of it. Researchers have the right to access and reuse pubic-sector information in application of the **PSI/Open Data Directive**

Tip

For more information on open data in the Flemish government (charter, action plan, model licences, metadata and URI standards), please visit https://overheid.vlaanderen.be/open-data-bij-de-vlaamse-overheid.

Tip

Take a look at the following data portals.

	🟡	🇧🇪	🇪🇺	🌍
General data	Flanders Open Data Portal (VODAP): https://opendata.vlaanderen.be/	https://data.gov.be/nl	https://data.europa.eu/euodp/en/home	https://dataverse.harvard.edu/
			https://www.euro-peandataportal.eu/	
Geodata	Geopunt: https://www.geopunt.be/	Geo.be: https://www.geo.be/	INSPIRE data portal: https://inspire-geo-portal.ec.europa.eu/	
Biodiversity				GBIF: https://www.gbif.org/
Genetics				GenBase: https://www.ncbi.nlm.nih.gov/genbank/
Marine biology				WoRMS: http://marinespecies.org/
Atmosphere: greenhouse gases			ICOS: https://www.icos-cp.eu/data-services/about-data-portal	
Mobility		Transportdata: https://www.trans-portdata.be/nl/		
Satellite images	Terrascope (VITO): https://terrascope.be/nl			

Tip

A basic list of international open data portals can be found on the website of the Open Access publisher PLOS: https://journals.plos.org/plosone/s/recommend-ed-repositories.

Tip

To find government-related **authentic data sources** [56] at federal, Flemish and Walloon level, please visit https://dt.bosa.be/nl/gegevensuitwisseling/authentieke_bronnen/overzicht_authentieke_bronnen.

[56] *The government speaks of authentic data as being 'unique and original data' that is correct and up to date. The aim is to register this data only once in an authentic source, so that the data only has to be submitted once to a public authority, while other authorities can (under conditions) consult these sources, and so they do not have to request the same data separately from citizens or legal persons.*

6. *Seek advice from support services.*

Scivil facilitates and supports citizen science in Flanders. As a result, Scivil is aware of the latest developments in citizen science, and can point you towards infrastructures, specialised institutes or partners. The Flemish government is also working on several initiatives that are all aimed at opening up research and government data as interoperably and openly as possible. The Data Stewards or Research Data Officers of large companies and organisations are aware of this: consult them too!

Tip

Make use of the knowledge hub that Scivil represents, so that you can be guided immediately to the services and expertise of our many partners. This way, you do not need to reinvent the wheel and your project can get off to a flying start.

For questions concerning the communication of your data, you can take a look at the Scivil communication guide.[57] For specific questions on privacy and ethics, we also refer to *principle 7*.

Tip

Take a look at the websites of initiatives such as VODAP[58], OSLO[59], VLOCA[60] and FOSB.[61]

---

[57] https://www.scivil.be/publications
[58] Flanders Open Data Portal - https://opendata.vlaanderen.be/
[59] Open Standards for Linking Organisations - https://data.vlaanderen.be/
[60] Flanders Open City Architecture - https://vloca.vlaanderen.be/
[61] Flemish Open Science Board - https://www.ewi-vlaanderen.be/nieuws/flemish-open-science-board-fosb-opgericht

Compared with classic scientific research, citizen science brings significant new challenges in terms of privacy and ethics. This may concern (a) data quality and integrity, (b) sharing of results and intellectual property, (c) conflicts of interest and (d) use of volunteers (see Resnik et al. 2015). Questions of ethics also differ greatly from project to project and from domain to domain. In the medical sciences, for example, proper guidelines have already been established for the collection of personal information.

In other fields, such as geology or biology, personal data is often not so much the subject of research as a side effect, as citizen science is used as a method. Consider, for example, the risk of exposing the whereabouts of people taking location-based measurements, or consider the pictures that participants take of, say, trees, which inadvertently depict children out for a walk. If you initiate a citizen-science project, whether or not in co-creation with citizens, it is important to deal with this in a well-considered way in your research design and to communicate clearly about it.

II. Privacy & Ethics

7. *Pay active attention to privacy and ensure knowledge sharing.*
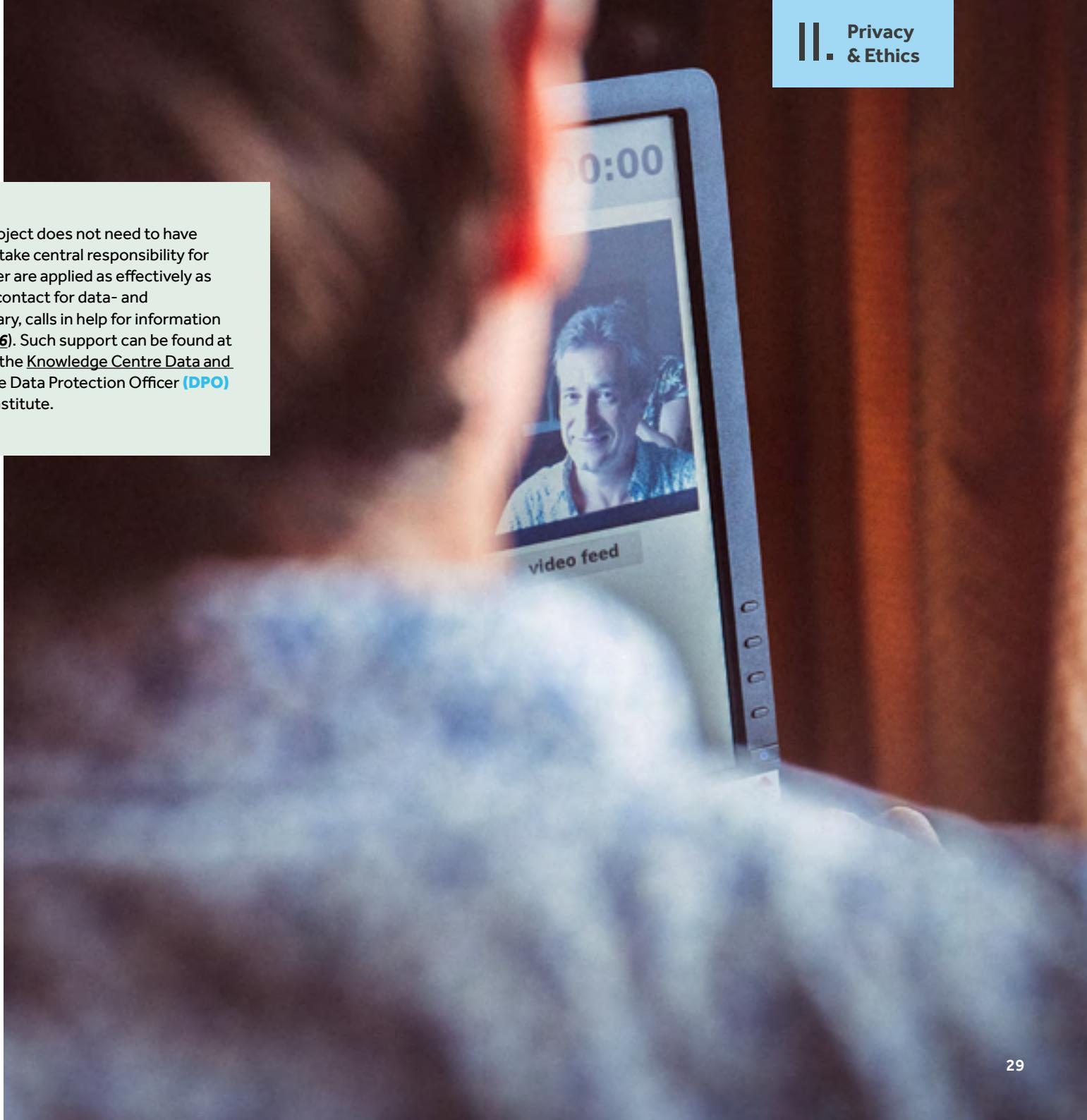
In order to achieve correct data processing, as a project leader you actively work on informing and training your staff and your citizen scientists.
This includes (a) clear communication about data management, privacy and ethics in citizen science, (b) the appointment of a data manager within the project, and (c) explicit attention to this topic in tenders and agreements with third parties.

Citizen scientists should receive a basic information package before participating in the project, covering privacy, rights and obligations, as well as the integrity of data handling.

Tip

A data officer within a citizen-science project does not need to have an extensive data background, but does take central responsibility for ensuring that the principles in this charter are applied as effectively as possible. This person is also the central contact for data- and privacy-related questions and, if necessary, calls in help for information and support from outside (see *principle 6*). Such support can be found at Scivil[62], the Data Protection Authority[63], the Knowledge Centre Data and Society[64] and other services - such as the Data Protection Officer **(DPO)** present in every university or research institute.

[62]   https://www.scivil.be/
[63]   https://www.gegevensbeschermingsautoriteit.be/professioneel
[64]   https://data-en-maatschappij.ai/

8. *Meet GDPR guidelines with your project.*

Respect the **GDPR**[65] rules for handling personal data of both your participating citizen scientists and any other people whose data is (un)intentionally captured during your project.

Personal data

Personal data includes all data on physical persons that, whether or not in combination with other data, can lead to the identification of an individual. A good, detailed description of what personal data is can be found on the website of the European Commission.[66] This may therefore include things such as a person's location, certain medical characteristics or digital data (e.g. usernames or IP addresses).

GDPR

The **GDPR** includes a number of basic principles, and an overarching additional principle ('accountability'):

⌐ **Lawfulness**: the collection, processing and disclosure of personal data must be covered by a national or international law that permits the processing of such personal data. A good overview of legal grounds can be found on the website of Ghent University[67] (consent, general interest, justified interest, legal obligation, execution of an agreement, vital interest). In some cases, such as **sensitive data**, this is not enough and the person whose data is being processed must give additional explicit consent.
The Data Protection Authority[68] (DPA) describes on its website[69] what the increased protection for such data consists of, among other things. Which information you should include in an informed consent can be found on this page[70] from Ghent University.

⌐ **Transparency**: inform the individuals whose data is collected by your project in advance in a clear, comprehensive and transparent (i.e. understandable) manner of how their data will be processed, stored and/or released or transferred. That too must be done lawfully.

⌐ **Purpose limitation**: de the reasons for which you collect personal data must serve a specific, explicit and legitimate purpose. You should also communicate these reasons to the person about whom you are collecting data. The way in which the data is processed must also be compatible with the intended purpose. This also means that you may not use data previously collected for a specified purpose for any other purpose.

⌐ **Data minimization**: the amount of data and the data quality should be kept as low as possible in line with the objectives of the research. This means that you only collect data that is strictly necessary for and relevant to the intended purpose. Specifically, the GDPR also recommends that all data be at least **pseudonymised** where possible.

⌐ **Storage Limitation**: specify for how long you will keep the data and in what form. If you do not pseudonymise data immediately at the time of capturing, you must pseudonymise this data internally as soon as possible (i.e. pseudonymisation as close to the source as possible), as soon as identification of the persons is no longer necessary. Exceptions to this principle can be applied to, among other things, the retention of data for statistical purposes, which must be kept for the long term.

⌐ **Accuracy**: processed personal data must be accurate and always kept up to date. Inaccurate data should be corrected or deleted immediately when it becomes apparent that it is inaccurate. Decisions based on inaccurate data may be challenged.

⌐ **Integrity & Confidentiality**: take appropriate protective measures to prevent inaccurate processing, accidental loss, deletion or corruption of the collected data. This also includes raising awareness of staff and information managers. Apart from the legal rules, you must therefore also observe a number of general principles. Among other things, make sure you use an **SSL certificate** for your website to help prevent data leaks.

⌐ **Accountability**: organisations must themselves be able to demonstrate that they have taken the necessary technical and/or organisational measures to comply with the GDPR. That responsibility therefore no longer lies with external control bodies. Again, information sharing is crucial here. Accountability applies to all the previous rules.

---

66 What is personal data? (European Commission):
    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/

Attention!

Anonymisation and **pseudonymisation** are two different concepts.
The pseudonymisation of personal data (for example, changing a name into a unique code for that person) does not necessarily lead to strictly anonymous data. In principle, and especially in citizen science, all data related to one person - whether or not in combination with other data - can lead to the ultimate identification of that person. Avoid guaranteeing people absolute anonymity if you have not included data aggregation (preferably before the data is stored) in your data protocol. Pseudonymisation and anonymisation should preferably take place as soon as possible after the data has been collected, in order to reduce the chances of data leaks. As long as you also keep a list that links the pseudonyms/codes with peoples' identification data , you cannot speak of absolute anonymity - after all, the identity of the people can be traced back this way.

Most citizen-science projects include a link to the privacy statement in the 'small print' at the very bottom of the web page. In some projects, you will only get to read the statement when you actually register as a participant in the project. Make sure that your privacy statement is easy to find on your website, that it is complete and correct, and that it is also concise and clearly understandable for everyone. It is useful if you also provide these in PDF form so that people can download them.

Tip

Do you want to share general information with your citizen scientists about the GDPR, and their rights? Then easily refer to this brochure from the European Commission:
https://ec.europa.eu/info/sites/info/files/virtual_identity_nl.pdf

Tip

Universities, too, often have a well-developed information page (e.g. this one from Ghent University[71]) about the GDPR or privacy in general. In addition to clear information about the legal rules, they often have their own 'rules of procedure' about privacy and ethics in research. At Ghent University, for example, they are clearly explained on this page:
https://www.ugent.be/en/research/datamanagement/policies/privacy.htm.
For comprehensive information on privacy in general, you can also visit the website of the Data Protection Authority[72] (DPA).

Tip

When drawing up a data management plan (DMP, see *principle 13*) you can use online tools and templates (e.g. DMPOnline[73]), which also include checklists related to GDPR. Be sure to use them.

[71] https://onderzoektips.ugent.be/nl/tips/00001763
[72] https://gegevensbeschermingsautoriteit.be/burger/privacy/wat-is-het
[73] https://dmponline.be/

Practical example

Privacy statement

The citizen-science project Fietsbarometer[74] is working together with schools to map the safety of cycling routes to schools. The project has a clear, understandable privacy statement[75], but also a clear page[76] with additional explanations about what happens to all data and how long it is kept, as well as a separate cookie statement.[77]
The privacy statement also explicitly states that no **sensitive personal data** ('special data') is stored or processed.

Other examples of clear privacy statements are those of Curieuzeneuzen in de tuin[78], Astrosounds[79], and SOS-Antwerpen[80], which are also available in PDF format.

Anonymisation
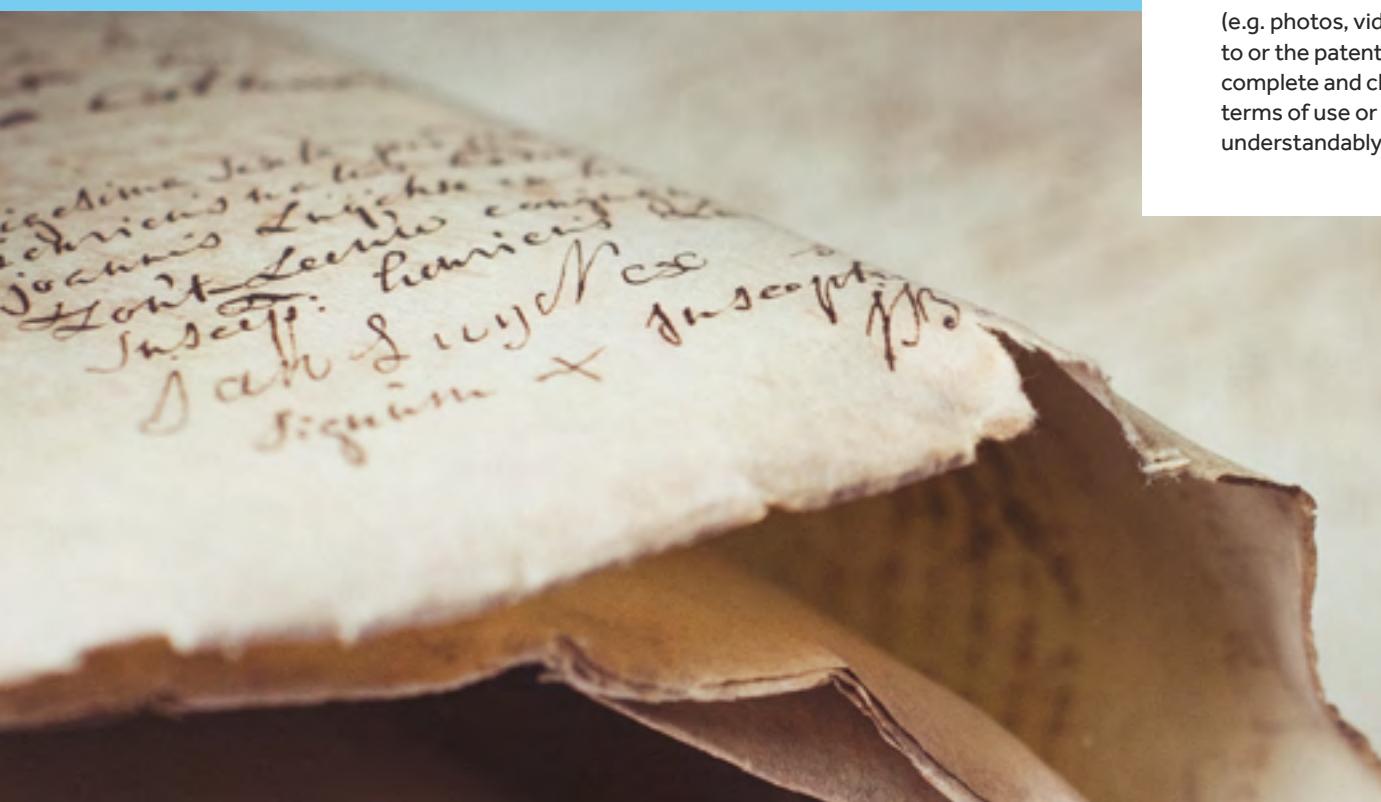
To protect the privacy of the participating students, Fietsbarometer uses pseudonymisation (each student is given a unique code) in combination with aggregation: cycle routes or segments are only displayed on the platform from the closest intersection to the home, and only when at least two students have cycled past the segment. For example, it is impossible to trace the home-school cycle route of a single pupil.

In the citizen-science project MamaMito[81], participants drew up their maternal family tree and their mitochondrial DNA was examined and compared with the DNA of other participants who were related through their 'foremothers'. This information is particularly sensitive to privacy. Volunteers from Histories.vzw and Familiekunde Vlaanderen were available to help and validate the family trees drawn up by the participants. Data less than 100 years old (and therefore easily verified by the participants themselves, e.g. via marriage records and in memoriam cards) was not verified by external parties.

[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. *Communicate clearly about intellectual property and copyrights.*

Before starting your project, think carefully about the various aspects that may give rise to questions or tensions about intellectual property. This concerns both intellectual property rights to what citizen scientists contribute (e.g. photos, videos, self-written texts), as well as intellectual property rights to or the patenting of one or more outputs of your project. Write correct, complete and clear terms and conditions of participation (and if necessary terms of use or purchase), which you communicate clearly and understandably to your participants before they contribute.

Intellectual property

Intellectual property can be a stumbling block to the further dissemination or use of the research results in some citizen-science projects. The likelihood of these problems occurring depends very much on the type of contributions made by the citizen scientists in the project. When citizens submit **photos, videos or written observations**, the likelihood of tensions is greater than when citizens transcribe texts, for example, or type data into online forms. The inclusion of personal data may raise privacy concerns, but does little to change intellectual property issues, with the exception of, for example, portrait rights. However, when citizens contribute through **data analysis or the development** of a product or protocol, questions of intellectual property rights arise more quickly. So think carefully, long before the start of your project, about the intellectual property and participation conditions of your project, as your chosen research protocols may differ accordingly. For example, it may mean allowing citizens to make categorised observations instead of their own free descriptions of an observation. It is also important to have clear written terms of use for the dissemination or use of the results of the project.

A third type of intellectual property rights that you should pay attention to are the so-called **3rd party copyrights**. These are the proprietary rights of, for example, existing platforms or websites that you use before you

collect or process data with or by citizen scientists. Please note that the General Terms and Conditions of such platforms may impose additional restrictions (or rights) on both contributors and recipients of the results of your project.

In Scassa & Chung (2015) you can read all about the management of intellectual properties in citizen science. Please note: the authors are Canadian and therefore often refer (e.g. in the case of patents) to the American or Canadian legal system.

> **Tip**
>
> Do you have a project where there is a good chance that the results will be patented? Then be careful with the (often intellectual) degree to which you ask for input from employees.

> **Tip**
>
> Some projects ask participants to make a financial contribution in order to participate. You must also provide correct, complete and clear information about this in the General Terms and Conditions.

Attention!

If a citizen or organisation contributes large amounts of data to your project, that data may be considered a 'database'.[82] In Europe, databases fall under the protection of the EU Database Directive[83], which states that databases that do not fall under classic copyrights can still be protected for a period of 15 years, and renewable every time the database undergoes a substantial update. This protection is violated as soon as someone extracts a 'substantial part' of the database.

Attention!

Although in the field of copyrights (usually valid up to 70 years after the death of the copyright holder/author) universal agreements have been made to harmonise the regulations worldwide, there may still be differences from one country to another in interpreting the rules. If your project includes input from different countries, be extra-vigilant (Scassa & Chung, 2015).

---

[82] *A database is defined in the EU Database Directive as a collection of independent works, data or other materials arranged in a systematic or methodological way, each of which can be accessed independently.*
[83] *https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009*

Practical example

The user segment[84] of the biodiversity platform Waarnemingen.be[85] clearly states the rights and obligations of citizen scientists and/or project coordinators who offer or take data from the platform. In addition, the text is also concise and readable at here.[86] The agreement includes (a) general provisions, (b) rights and obligations when using multimedia materials from the platform, (c) map material, and (d) the desirable method of source acknowledgement for the platform and for the use of individual multimedia materials (photos, sounds). A new user publishes his/her data under a CC-BY-NC licence by default, but can change that licence. Thus, each user chooses whether his/her multimedia material is available to third parties or not. Waarnemingen.be does ask that the platform is always mentioned as the source, even if, for example, one publishes photos that one has uploaded to the platform oneself. Subsets of the waarnemingen.be database can be downloaded for non-commercial individual use only. For professional use, a separate permission or agreement is needed.

Open Street Map[87] is not a citizen-science project but is very often used in projects, such as Fietsbarometer (Ghent University) and Vespa-Watch[88] (INBO). It is covered by the open licenceOdBL[89], which makes it easier to use for the visualisation of citizen-science results or data. However, the platform does stipulate in itsGeneral Terms and Conditions[90] that derivative works must be published under the same conditions (licence) as those that apply to the existing maps of Open Street Map. For some researchers this may conflict with their task of disseminating results, and may therefore be a reason not to work with Open Street Map.

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
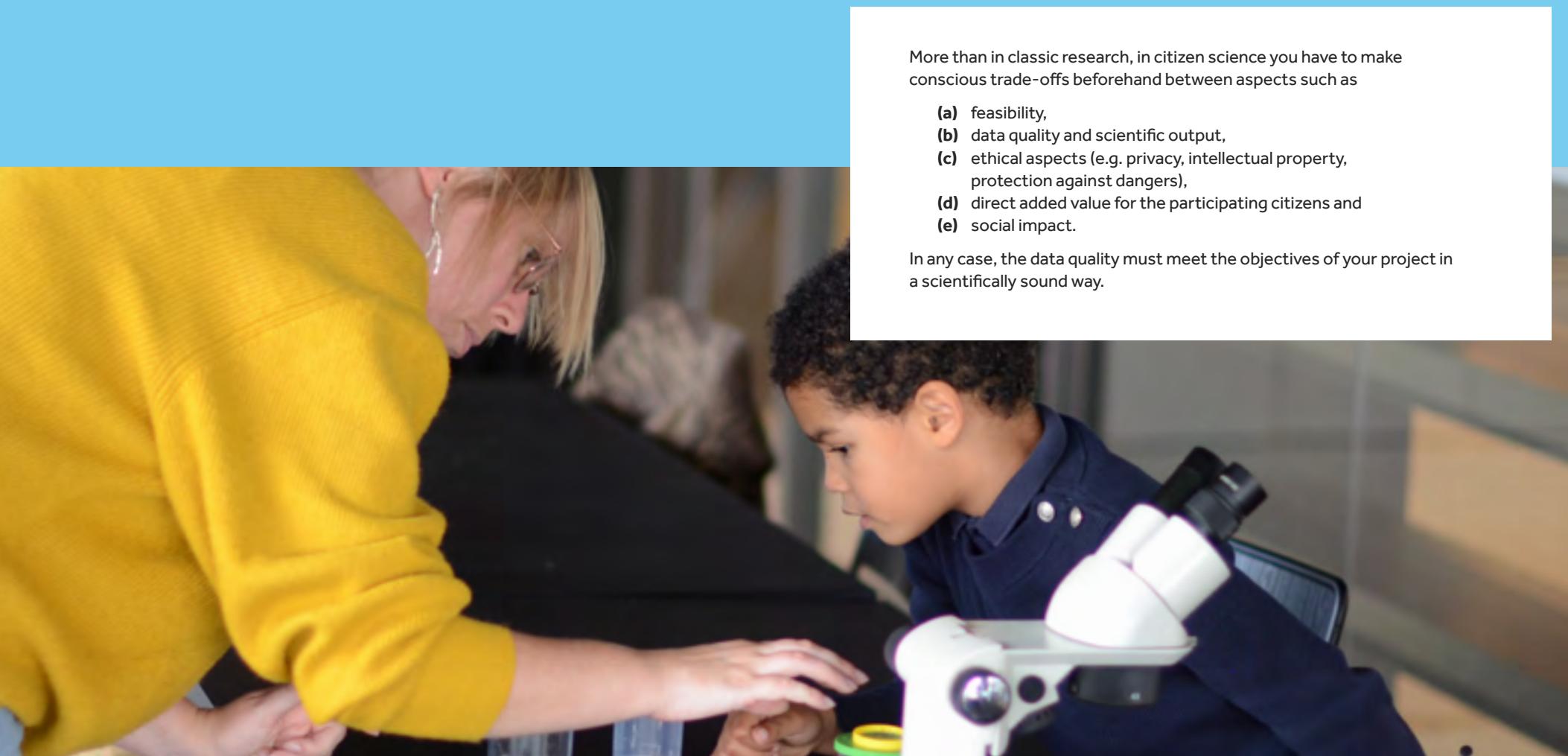
10. *Carefully weigh your data quality and quantity against ethics, feasibility and project goals.*

More than in classic research, in citizen science you have to make conscious trade-offs beforehand between aspects such as

(a) feasibility,
(b) data quality and scientific output,
(c) ethical aspects (e.g. privacy, intellectual property, protection against dangers),
(d) direct added value for the participating citizens and
(e) social impact.

In any case, the data quality must meet the objectives of your project in a scientifically sound way.

You can achieve a sufficiently high data quality for your project in various ways. This is explained more concretely in **_principle 15,_** where the trade-off between quality and quantity is made according to the intended results and the feasibility of the chosen method. However, obtaining scientific data or analyses is not the only purpose of citizen science: projects often also have an educational and social value. Some projects, which (if carried out with sensors) tend to be called 'citizen sensing'[91], primarily serve a social purpose, namely to raise an environmental problem with a certain authority. Although the data generated must be sufficiently reliable to serve the project purpose, this does not always lead to sufficient data for a robust study that can be published in scientific journals, and that is therefore not always necessary. It is important, however, that you communicate the methods used and their robustness properly.

Ethical constraints, as mentioned in **_principles 8_** and **_11_**, together with feasibility (e.g. constraints due to financial barriers, or due to situational barriers such as a COVID lockdown) should be considered when determining the methods to achieve good data quality.

Attention!

The selection of your participants can also have an impact on both your data quality and your ethical standards. Choose wisely according to your project goal: is the emphasis on social impact by reaching hard-to-reach groups? Or do you need citizen scientists who are highly skilled in a specific field (e.g. professional associations, hobby associations)?

---

[91] See _https://research.tilburguniversity.edu/en/publications/sensing-the-risk-in-search-of-the-factors-influencing-the-policy- for more information about Citizen Sensing._

11. *Beware of ethically undesirable side effects of releasing your data.*

Consider at each step in your project whether it is ethical to release the data collected. Sometimes you cannot make data publicly available because it might endanger a rare or endangered species, for example, or because it would violate people's privacy or harm people. Deliberately pay sufficient attention to this.

Sometimes you collect **'sensitive'** data: that is, data about, for example, vulnerable groups of people, children, elderly, sick or psychologically vulnerable people. Sensitive data can also include, for example, the location of an endangered or rare animal or plant species. In such cases, it is often in the interest of that person, or that organism, to keep their location secure internally, but not to disclose it to third parties (i.e. not to make it Open). This consideration often has to be made more in citizen science than in classic scientific research. Similarly, if an organisation cannot financially or technically maintain its citizen science database without claiming revenue from the release of data, that can be a reason not to release certain datasets.

Practical example

Waarnemingen.be aims to make the data collected by citizen scientists as widely available as possible for general use. However, some observations or multimedia are not published openly on the platform, for example

when an endangered species is involved. In doing so, Waarnemingen.bewants to avoid such a sighting generating a public interest that would threaten the observed animal.

At Fietsbarometer (Ghent University, see also *principle 8*), among other things, bicycle segments of pupils between the home of the pupil concerned and the next intersection on the way to school where the pupil no longer appeared as the only one on the route were omitted.

12. *Adopt a respectful, equal attitude towards citizen scientists.*

Following on from, and in line with, ***principle 11*** , when planning and executing your project, make sure you know what motivates your (future) participants. Within the limits of your scientific context and objectives, try to accommodate them as best as you can, and deal with them respectfully. This includes full, timely and clear communication with your participants about the use of their data (regardless of GDPR rules). You should also take great care when thanking or naming citizen scientists for their contributions to your project, when publishing your results, whether or not with a co-authorship, and/or through other media.

Citizen-science projects distinguish themselves from other scientific research through the participation of citizens in one or more phases of the research. Citizens typically participate out of (a) a desire to learn about a particular topic, (b) being part of a 'community' and having a purpose, (c) having a sense of being able to contribute to science, and/or (d) concern about a particular situation, and thus the ambition to influence policy. Those who contribute logically expect to be able to consult the results afterwards.

Tip

The Scivil communication guide[92] for citizen science deals extensively with this theme.

Practical example

The best-known practical example is the naming of 'Hanny's Voorwerp', a reflection nebula discovered in 2007 by the Dutch teacher Hanny Van Arkel. She found the previously unknown astronomical object in a photograph she was analysing for the Galaxy Zoo[93] project. To this day, astronomers around the world use the name 'Hanny's Voorwerp' to refer to the nebula (e.g. Keel et al., 2012).[94]

Teacher Christof Wiedemair, and some of his pupils, have co-authored numerous scientific publications through his contributions to astronomical observations (e.g. Wiedemair et al., 2020).[95] Some of his students have also been included as co-authors in relevant publications. Palaeontologist Mark Bosselaers[96] also regularly writes (and contributes to) scientific publications.

Even if a larger group of people have contributed to your project, they can still be included, either as co-authors or in the acknowledgements of the article or report. Harlin et al. (2018)[97] wrote, for example, a chapter entitled 'Turning Students into Citizen Scientists' in a book on citizen science. The students of the Leysin American School in Switzerland were jointly included as co-authors in this chapter.

In the Stiemerlab[98] project (City of Genk), local residents of the Stiemer valley are very closely involved in different phases of the research into the water quality of the Stiemerbeek. There is a strong focus on communication and community building, including organising workshops, sending out regular newsletters, creating an attractive website where people can access the data, and the close involvement of the city administration.

In the B&Bee[99] project, which aims to study the occurrence of wild bees in our region, equipment (a bee hotel and sampling material) is made available to participants, and it is ensured that the scientists are readily available in case of questions from participants. The Astrosounds[100] project, on the other hand, focuses on captivating content and clear explanations tailored to the participants.

[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
[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/

III. Data hygiene

13. *Draw up a data management plan (DMP).*

In a data management plan (DMP), you describe how, during and after the completion of a project, you ensure that all data that comes into play in a research project is of reliable quality, is stored securely and is made as findable, accessible, reusable and linkable as possible. Describe carefully how you will achieve this in your project and use appropriate templates that are available. A DMP is more than mere administration: it has a central supporting role in your project. Also appoint a (data) manager who maintains an overall perspective and can act as the data contact person in your project (see also ***principle 7***).

44

Data management plan (DMP)

A good data management plan (DMP) helps you to plan your project more effectively and to estimate the costs of your data management more accurately in advance. Moreover, parts of your DMP correspond to what you need to comply with the General Data Protection Regulation (GDPR) and to draw up your privacy statement correctly.

Creating a general uniform data management template for citizen science as a whole is difficult, as each plan is tailored to the particular project and research area. Broadly speaking, DMPs contain information about the project and about the data(sets): what data is generated, what is reused? What are the (meta)data formats? Which **data standards** (see *principle 17*) are you using? How is access to the data organised? How do you protect the data from accidental loss? What budget do you anticipate for data management and storage? Finally, how will you open up or share the data, even after the project has ended?

Tip

Ghent University has a useful, understandable and complete web page on Data Management Planning[101], with links to tools and lists. At European level too, summary information[102], additional tools and examples of data management plans are given via the OpenAIRE[103] project.

Tip

The UK Environmental Observation Network wrote an accessible guide[104] on creating a data management plan in the context of citizen science. In addition, the data working group also produced a number of recommendation documents[105] on data management, including a document[106] with a simplified data roadmap for citizen-science projects.

Tip

Have you checked every principle in this charter against your project? Then you probably already have enough material with which to fill a DMP.

Tip

Reuse **existing templates**.

Many Flemish research institutions[107] use the online tool 'DMP-Online'[108] to draw up data management plans for research projects. The tool contains several templates for creating such a plan, with all the relevant questions to be answered before starting your project or applying for funding. In DMP-Online you can create your own data management plan for internal use, you can use a template of your own institute (e.g. INBO, UGent) or you can follow a template of a funding body (e.g. FWO-Vlaanderen, Belspo, European Commission) that is included in the tool.

Are you connected to an institute that uses DMP-Online?
Then you can log in directly through your organisation. As an external party, you can gain access via your **Orcid ID**[109] and, (only) at the invitation of someone from an affiliated institute, participate in a joint data management plan.

Is your research institute not connected to DMP-Online? Then check whether it has developed its own DMP tool (e.g. KU Leuven: https://dmponline.kuleuven.be/).

Is that not the case either? Don't worry: most funding bodies also make their template available separately, e.g. in a Word document[110] (see FWO-Vlaanderen[111], Belspo[112]) or a web page[113] (e.g. Horizon 2020 programme).

You can also draw inspiration from the list of examples of data management plans publicly available on the original UK DMP-Online website: https://dmponline.dcc.ac.uk/public_plans.

---

[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/cswg-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

[107] The following institutes are connected to DMP-Online: ILVO, Sciensano, UAntwerpen, UCLouvain, UGent, UHasselt, ULB, ULiege, UMons, UNamur, UZGent, VITO, VLIZ, VUB.
[108] https://dmponline.be/
[109] Available free of charge from 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)
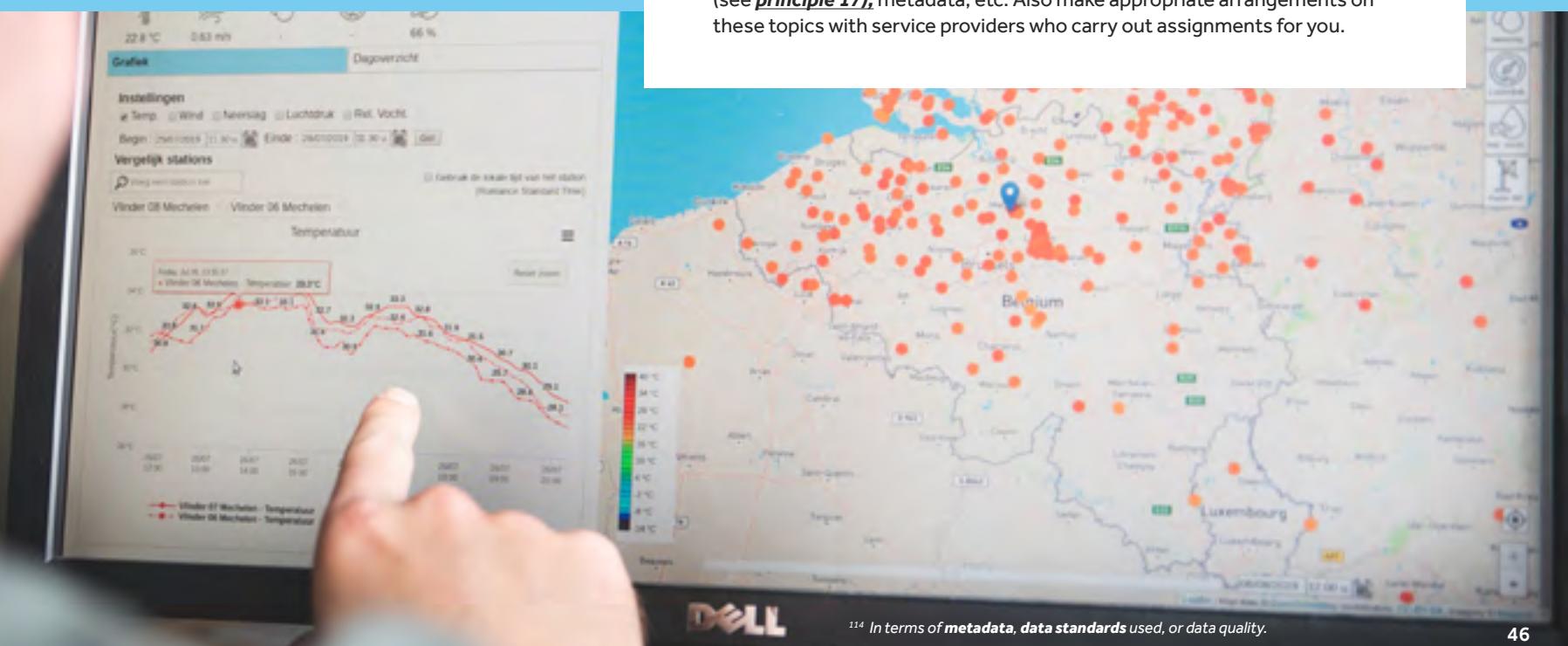
14. *Take into account the cumulation potential of data: 'Treat a small dataset like a large one'.*

At the start of a **citizen-science project**, the organiser should pay extra attention to data hygiene. The underlying idea is: rather five well-documented[114] elements than 50 poorly documented ones. Ensure active knowledge sharing with your citizen scientists and co-creators (e.g. co-project writers) about data management, **data standards** (see *principle 17),* metadata, etc. Also make appropriate arrangements on these topics with service providers who carry out assignments for you.

[114] *In terms of **metadata**, **data standards** used, or data quality.*

Tip

Educate and raise awareness among your **participants** about data hygiene.

More than in any other classic research, citizen science has a very great diversity in terms of size of datasets or quality of measurements or observations. This is why it is particularly important to pay specific attention to data hygiene in terms of potential interoperability or integration into other projects' datasets. It is important, not only for increasing the interoperability of the data itself, but also for increasing the motivation and engagement of your participants, to strongly emphasise this aspect in communication to participants. Above all, remember that if you collect only a few metrics, they can eventually become a crucial part of a dataset to which important scientific findings can be linked.

Tip

Make appropriate arrangements with **service providers** when outsourcing a job.

Practical example

Many biodiversity projects, such as Teek a Break[115] (UAntwerpen) or Bugs 2 the Rescue[116] (VUB) link up with existing crowdsourcing platforms for biological observations, such as (in this case) waarnemingen.be. In doing so, they are acknowledging the importance of the interoperability of their datasets, even if it seems small compared with some of the 'bigger' projects.

[115] https://www.uantwerpen.be/nl/projecten/teek-a-break/
[116] https://www.bugs2therescue.be/

15. *Pay special attention to the quality of data..*

Data quality forms an integral chapter within a data management plan, but deserves extra attention within the context of citizen science: the quality of your data should meet the minimum required for its own use in your project, provided that it is well described by correct metadata (see ***theme V***).

To achieve good data quality for your project, you should consciously choose the profile of your participants, carefully weighing it against any educational or social objectives of your citizen-science project. Take into account the principles from Theme II. Finally, also look for the most appropriate balance between quality and quantity of individual measurements or analyses to achieve optimal quality of project results (see also ***principle 10***).

Quality versus quantity

In citizen science, there is always a trade-off between individual data quality (resolution, precision and correctness) on the one hand, and scaling up to larger amounts of observations or measurements (with better statistical representativeness) on the other.

Broadly speaking, you can focus on different aspects:

(a) You ensure a very large number of samples, measurements or observations by involving a **large number** of participants in your project. You can either use a maximum number of participants or a smaller group that carries out multiple observations or measurements. The advantage of 'large numbers' is that different participants can **duplicate** certain observations, measurements or analyses.

(b) You pay extra attention to **training** the citizen scientists, so that they can make more accurate observations or work more independently. For example, you can provide extra guides, protocols, an extensive website or a teaching package, but also multiple contact moments or extra guidance by 'ambassadors' from your field (e.g. strong hobbyists, well-trained amateurs).

(c) You focus on 'revision': you work together with researchers or with the aforementioned 'ambassadors' in your field of study, so that they can review and validate the work of citizen scientists.

In some projects, strong emphasis is placed on individual coaching and training of citizen scientists, which allows for greater independence for the citizen afterwards, while still providing reliable results. Other projects benefit more from simple protocols that are widely disseminated in order to obtain a larger number of measuring points. This depends, as also mentioned in ***principle 10***, on your project goals and the context (e.g. financial) in which you are working.

Subsequent statistical corrections can also help to monitor the reliability of the data as a whole.

Tip

The American Environmental Protection Agency[117] (EPA) published an extensive guide and practical examples[118] on data quality management in citizen-science projects.
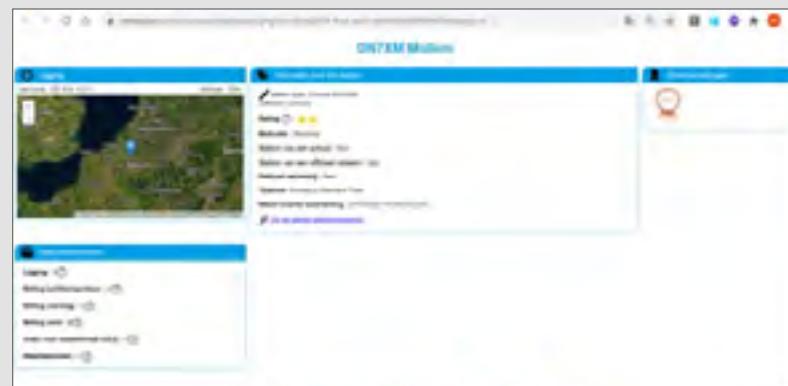
---

[117] *https://www.epa.gov/*
[118] *https://www.epa.gov/citizen-science/quality-assurance-handbook-and-guidance-documents-citizen-science-projects*

Practical example

In a project such as CurieuzeNeuzen[119], in which the NOx content in Flanders was monitored for the month of May 2019, the focus was on the power of numbers: 20,000 participants provided a particularly large number of measurement points. That said, detailed measurements with sensors were not possible for all those participants, for reasons of practicality. In the MamaMito[120] project, in which participants reconstructed a family tree from their mother's line, the guidance of citizen scientists was strongly emphasised. This was done in cooperation with Familiekunde Vlaanderen, the Davidsfonds, and certain local history circles. In the VLINDER[121] project (Vlaanderen in de Weer), 60 high-quality weather stations were placed in well-considered locations in Flanders. Validation of data entered on the DoeDat[122] platform of Meise Botanic Garden is carried out by the scientists themselves. They still consider this way of working to be more efficient than if they were to make their observations without the help of citizens.

The international weather platform WOW[123] uses **metadata** from weather stations to assign a quality label to each station.
Citizen scientists who send data from their weather station to the platform can also enter the specifications and environmental factors of their station, after which the web platform automatically assigns a star rating.
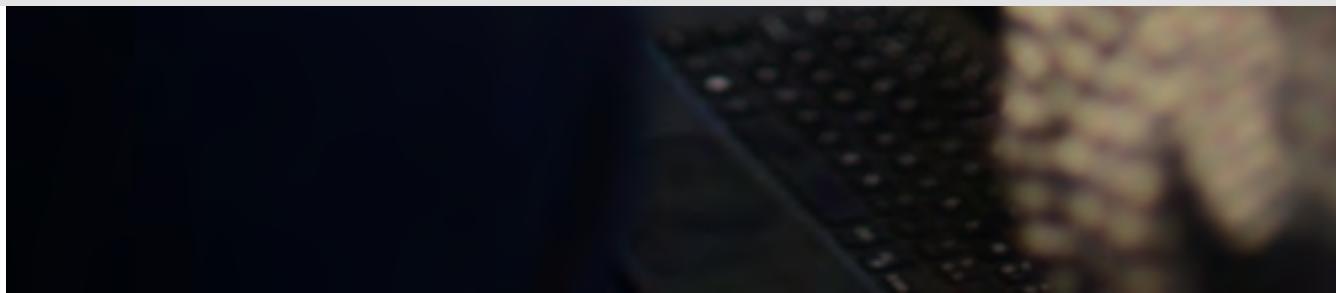
[119] https://2018.curieuzeneuzen.be/
[120] https://historiesvzw.be/projecten/mamamito/
[121] https://vlinder.ugent.be/
[122] https://www.doedat.be/
[123] https://wow.meteo.be/nl/
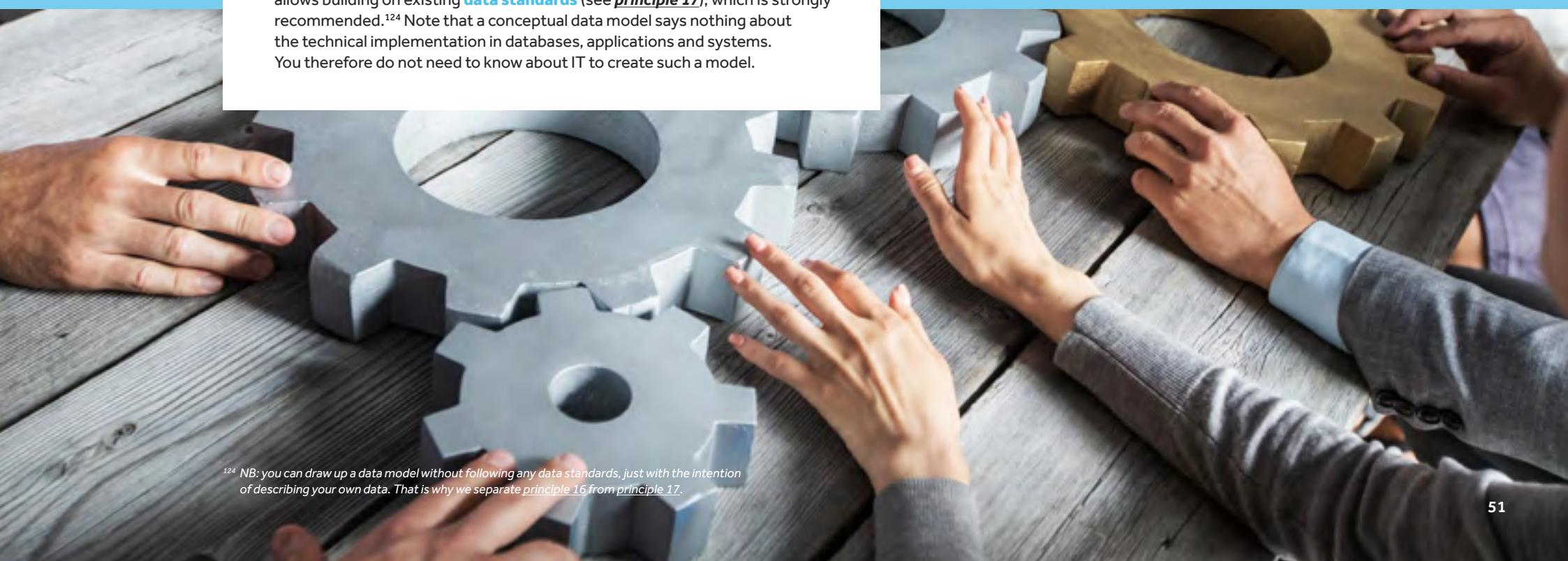
16. Create a conceptual data model for your project.

A conceptual **data model** (also known as an information model or domain model) is a way of unambiguously and logically describing the information or data to be collected or reused in a project. This is necessary to be able to communicate the meaning of the data both inside and outside the project, and helps prepare for the physical storage and publication of data. It also allows building on existing **data standards** (see *principle 17*), which is strongly recommended.[124] Note that a conceptual data model says nothing about the technical implementation in databases, applications and systems. You therefore do not need to know about IT to create such a model.

[124] NB: you can draw up a data model without following any data standards, just with the intention of describing your own data. That is why we separate *principle 16* from *principle 17*.

Describe at least the entities, attributes and relationships

Any collection of information (or: data) in the world can be described using three elements:

(a) *entities* [= real objects or concepts, things with meaning],

(b) *attributes* [= the characteristics of those entities], and

(c) *relationships* [= what the connection is between those entities].

With these three elements you can easily draw a 'conceptual data model' that describes a certain set of data (see Figure 4). Afterwards, you can add additional nuances to such a diagram.
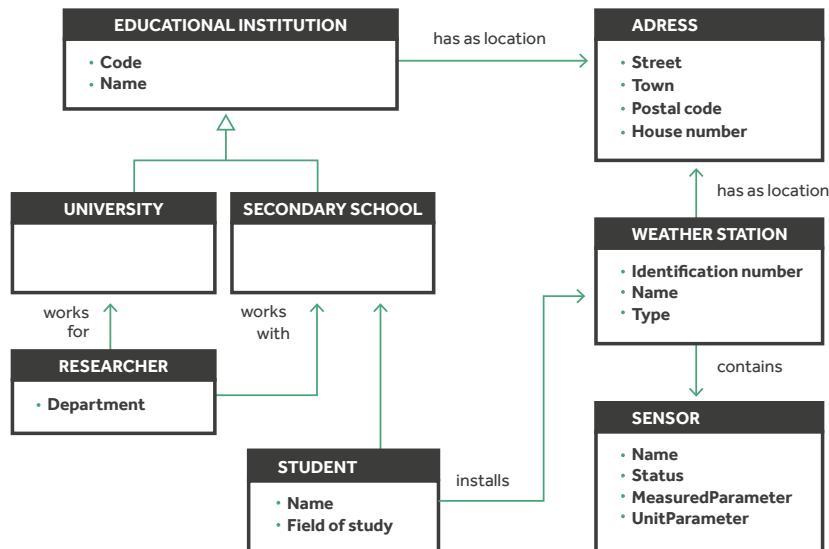
Figure 4
Data modelling applied to the domain of the VLINDER citizen science project: each rectangle describes an entity (grey), possibly with attributes. The connecting lines indicate relationships between entities. The hollow arrows indicate an 'is a' relationship: 'A University is a type of Educational Institution, and therefore inherits its attributes (characteristics).'

Provide clear definitions

Entities, attributes and relationships are defined in an unambiguous way so that there can be no misunderstandings about the meaning of the data collected. When such definitions are captured and recognised by a broad group of domain experts, they are collectively called a **vocabulary** , and can become part of a **data standard**.

Data standards

The first reflex should be to use existing data standards (see also *principle 17*) in your model, and to extend these standards where necessary.

It makes no sense to invest a lot of time and energy in describing entities, attributes and relationships for which **data standards** already exist. Whether you want to model locations or sightings of species, the chances are very good that others have done this before. By following existing standards, you not only save time, but you can also immediately start exchanging your data according to that standard, which enhances cooperation and **interoperability**.

Also try to reuse and combine as much as possible existing data models, and then only add those entities, attributes and relationships that are needed to describe the data from your project.
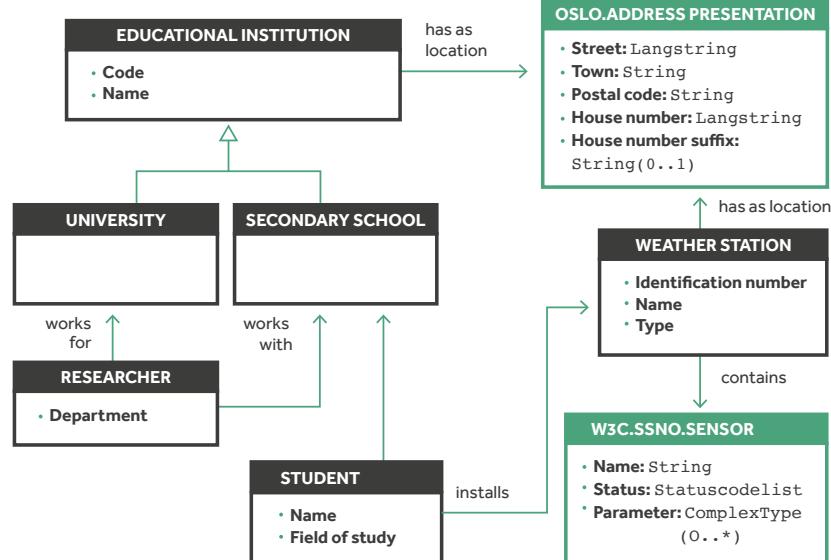

UML class diagram

Use a UML class diagram

Although there are various modelling languages that are suitable for conceptual data modelling (e.g. **ERD** and **ORM**), '**UML** Class diagrams' are the most appropriate way of creating conceptual data models for three reasons:
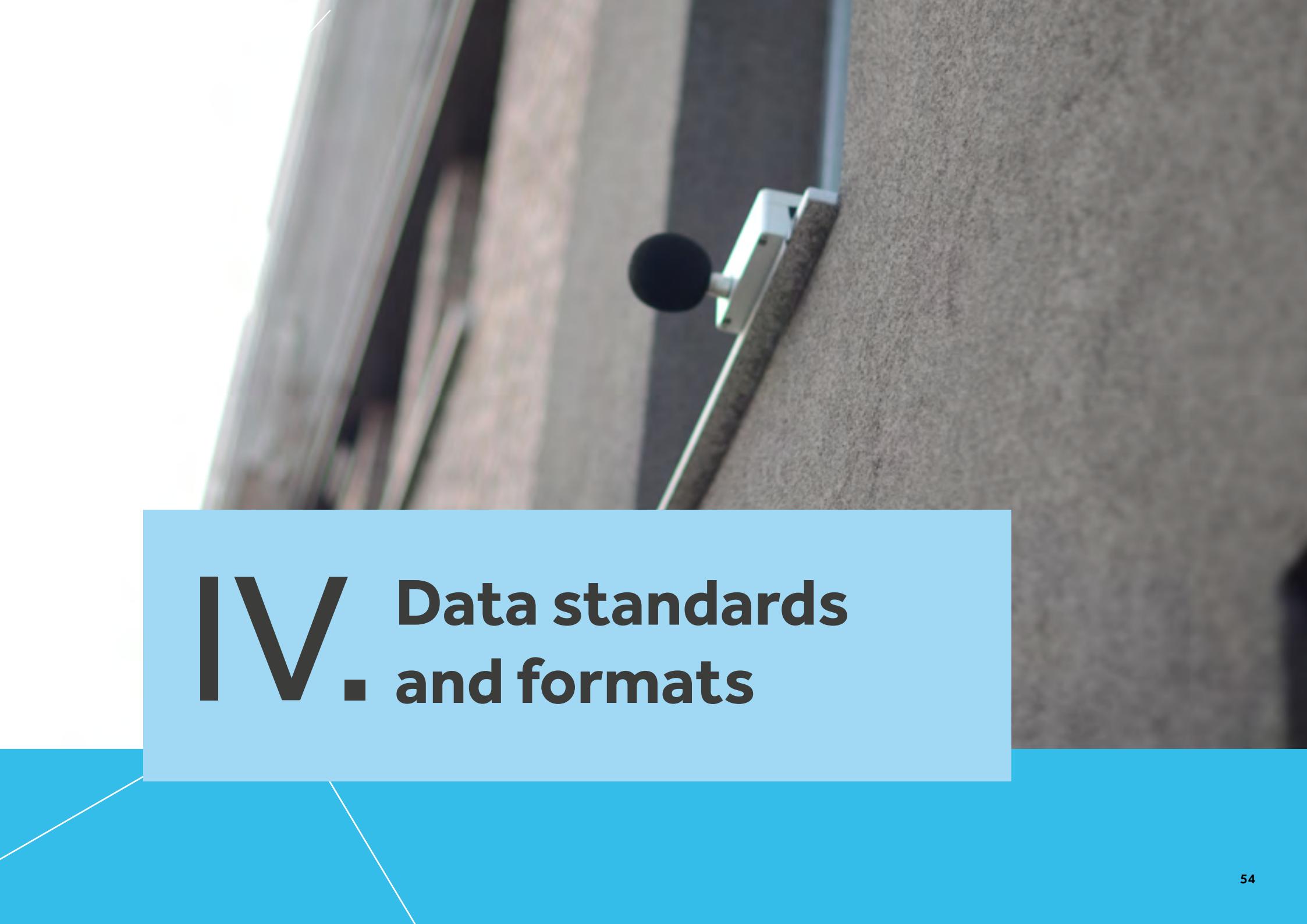
1. 'UML Class' is more comprehensive than ERD and ORM and is also an ISO standard.

2. The entry threshold for a beginner is low (using entities, attributes and relationships is sufficient), while advanced and expert users have every opportunity to add nuances or to automatically convert the data models into code or documentation.

3. Any self-respecting software package for information modelling supports the UML modelling language. UML is therefore very common in the world of information modelling and data standards.

> **Tip**
>
> Overview of the building blocks of a UML class diagram: https://www.uml-diagrams.org/class-diagrams-overview.html.
>
> Video tutorial on UML class diagram (10 minutes): https://youtu.be/UI6lqHOVHic.
>
> Free, online drawing environment where you can immediately start modelling: https://www.draw.io/.

Figure 4
Same data model as in the previous figure, but now we use existing data standards for the two blue entities. From the OSLO application profile
'address' was reused as 'address representation', and a link was also made to the description of the entity 'sensor'' from the W3C-SSNO standard.

IV. Data standards and formats

17. *Build on existing data standards relevant to your project.*

Data standards record the meaning and structure of exchanged information in **vocabularies** (with definitions), data models (with entities, attributes and relationships) and **code lists** (with uniformly usable codes). They are often domain-specific, and are agreed upon by domain experts. The use of data standards is crucial to achieve **'semantic interoperability'**. If different (citizen-science) projects use the same data standards, this facilitates the exchange of information and knowledge sharing. As far as possible, use the existing data standards from your research area.

It is impossible to draw up a list of all relevant data standards, firstly because citizen-science projects operate in different domains (air quality, biodiversity, mobility, health, linguistics, etc.) and secondly because in certain specific research domains such standards are still being developed. We are talking here about **domain-specific standards**, the most well-known of which are the **TDWG** standards and the **INSPIRE** data specifications (see below). Specifically for citizen science, more appropriate derived standards are developed for each research area by working groups of experts from each area (see *'practical examples'*).

TDWG standards
https://www.tdwg.org/

Biodiversity standards, of which the Darwin Core'[125] (with the entities 'organism', 'specimen', 'taxon' and their attributes) is probably the best known. Used, among others, by the GBIF platform. The taxonomic register Aphia[126] lists all known marine species in the world and provides an overview of the accepted taxonomic names (including synonyms). This register is used worldwide as a standard list.

INSPIRE data specifications
https://inspire.ec.europa.eu/data-specifications/

These standards were developed with a view to sharing spatial data between European Member States in support of environmental policies. These English-language data specifications (with UML models, definitions and code lists) cover various environment-related domains (e.g. 'atmospheric conditions', 'habitats and biotopes', 'species distribution', 'human health and safety'). Their application is supported by extensive documentation, toolkits and an international community. These standards are used, for example, in the geoviewer of the coastal portal[127] (Flanders Marine Institute).

In addition, there are also standards that are relevant irrespective of the domain, for example because they describe **basic or location data** to which domain-specific data can be linked, or because they are **more generic (cross-domain)** in nature. Here is a non-exhaustive overview of some interesting data standard portals:

ISO standards
https://www.iso.org/standards.html

These international standards could be interesting for a citizen-science project. For example, the ISO 19131 standard[128] is well suited to describing geographical data, and the language codes in ISO 639[129] can be used to describe languages or language families.

W3C standards
https://www.w3.org/TR/?tag=data

W3C standardises the internet worldwide. From there, recommendations and vocabularies were also developed that may be of interest to citizen-science projects. Examples are the 'Semantic Sensor Network Ontology'[130](describes sensors and their observations), the 'Data Catalog Vocabulary'[131] (DCAT, suitable for metadata) and 'Best Practices for Publishing Linked Data '.[132]

OSLO data standards
https://data.vlaanderen.be/

The OSLO standards are compatible with international standards (ISO, INSPIREISA,ISA,[133], etc.) but are easier to apply because they were drawn up in Dutch by Flemish domain experts. Currently there are 94 recognised standards[134], and 42 candidate standards. The OSLO standards contain vocabularies, but also **'application profiles'**. Standards that relate to basic data, such as 'addresses', 'buildings', 'roads', 'public domain', 'people' and 'contact preferences' can be useful for many citizen-science projects, as can standards that focus on a specific (scientific/life/policy) area, such as 'soil and subsoil' or 'air and water'.

---

[125] *https://www.tdwg.org/standards/dwc/*
[126] *http://www.vliz.be/nl/taxonomic-register-marine-species*
[127] *https://www.kustportaal.be/nl*
[128] *See https://www.iso.org/standard/36760.html. This is used, for example, by Digitaal Vlaanderen, when publishing geographical raster and vector data.*
[129] *See https://www.iso.org/iso-639-language-codes.html. Are used, for example, in website addresses, but can also add value in other applications.*
[130] *https://www.w3.org/TR/vocab-ssn/*
[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 data standards
https://core.citizenscience.org/docs/observation

The Data & Metadata Working Group[135] of the Citizen Science Association (CSA) has been working for some time on standards for data from projects with public participation. The **PPSR** core consists of an 'Observation Data Model' (ODM) and a 'Common Data Model' (CDM). In addition, two PPSR metadata models were also prepared (respectively for datasets and for projects; see *principle 24*).

In the **Observation Data Model** (ODM), the entity is 'an observation', and the attributes are all kinds of things seen, heard, noticed or reported by a citizen scientist carrying out this 'observation'. The ODM describes different types of 'observations'. A series of those 'observations' together form a dataset. Sometimes you can use attributes from one domain for another domain, which is good, because then datasets from different domains can be combined more easily. Op the website of PPSR-ODM[136] you will find a list of domains in which standards have already been agreed; the most well known are the Darwin core (for biodiversity) and Ecological Metadata Language[137] (EML) for ecological research.

The **Common Data Model** (CDM) is a **PPSR** umbrella model that stitches together the other PPSR data and metadata models. This model provides a context for groups of projects, which can be linked together, for example, within a given framework programme or a given campaign.

Tip

If you want to create policy impact with your citizen-science project, first check with the relevant policy bodies what conditions/standards your data must satisfy to be easily integrated with policy data. That way you increase the potential impact of your data and of your project.

Practical example

Within the field of Biodiversity (TDWG), the Citizen Science Working Group is looking into the application of the Darwin core standard in citizen-science projects. More information can be found on the Github[138] page of this working group, which includes representatives of the Research Institute for Nature and Forest (INBO).

[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
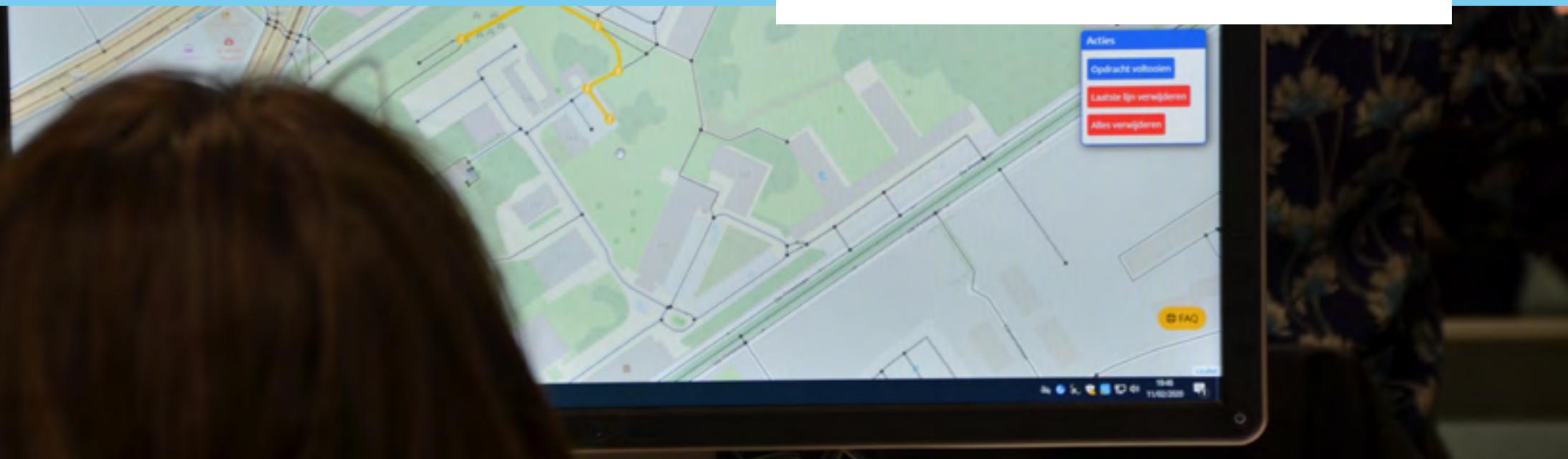
18. *Use machine-readable, open formats.*

Providing data in open (i.e. specifications not kept secret or licensed by one organisation[139], but released and/or standardised) and machine-readable (i.e. structured) formats is a step towards 'technical interoperability'. More software applications will be able to read and process your data. You also avoid the risk of '**vendor lock-in**'.

[139] Note that a format may be openly documented, but still proprietary because subject to licences.

Tip

Before publishing your data, pay sufficient attention to choosing the most suitable file format. This format must not only support the intended applications, but also encourage reuse by being machine-readable and open.

Here are some examples of file formats[140] that you should replace with an 'open' alternative:

Proprietary format	Example of 'open' alternative
.doc (text with formatting)	.odt / .docx
.xls (data in table form)	.ods /.csv / .xlsx
.dwg (CAD drawings)	.dxf
.shp (geographical vector data)	.gml
.bmp (images)	.tiff
.wma (sound)	.mp3 / .flac
PHP or Java **data serialisation**	.json / .xml

Practical example

The projects VLINDER[141] & Leuven.Cool[142], which offer their data via the WOW-be[143] platform, do so not only in graphs (figures) and xls format, but also in .csv. The API that gives access to the data of VLINDER[144] is also being made open and available in json format, via Github.

---

[140] You can find more information on these file formats by searching at 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_
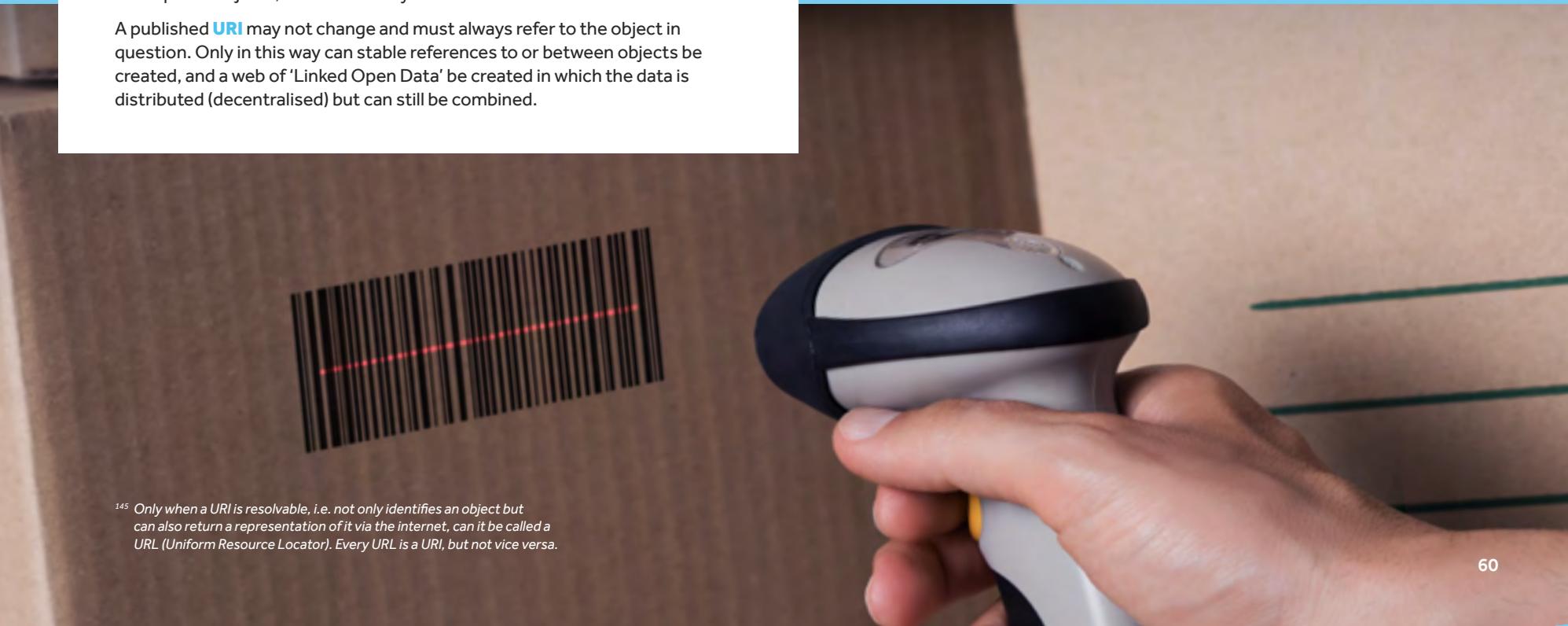
19. *Assign globally unique and persistent identifiers to your data.*

By assigning a unique identification code to each concept or measurement in your dataset, you make communication about (the meaning of) data, both between people and between machines, much clearer and easier. Choose **Uniform Resource Identifiers** (URIs) as globally (literally: worldwide) unique identifiers. These 'web addresses' serve primarily to uniquely identify data concepts or objects, and secondarily to access information about them.[145]

A published **URI** may not change and must always refer to the object in question. Only in this way can stable references to or between objects be created, and a web of 'Linked Open Data' be created in which the data is distributed (decentralised) but can still be combined.

[145] *Only when a URI is resolvable, i.e. not only identifies an object but can also return a representation of it via the internet, can it be called a URL (Uniform Resource Locator). Every URL is a URI, but not vice versa.*

URIs

How can you provide URIs to refer to your project data?

Let's take waarnemingen.be as a (fictitious) example to provide a URI step by step. On this website, volunteers in Belgium and the Netherlands can register observations of plant and animal species.

1. Decide which data (images, measurements, locations, annotated documents, alphanumeric data, etc.) will be fully managed by your project. Or in other words: what data are users looking for from you that they cannot find anywhere else? Please refer to your domain model (see *principle 16*).

   → Applied: species descriptions, observations, photos, sounds, etc.

2. Think about the **domain under which you want to offer your data**. Since the domain name will be a part of the URIs of your project data and data concepts used, you must have absolute guarantees that you can 'claim' this domain, keep it **stable** and manage it yourself.

   You can also choose to offer your data on an online platform as long as it offers persistent URIs for the uploaded data. In that case, you do not need to own the domain name, and you can leave the publication of your project data with URIs (or as **RDF triples**, for example) to this (international) platform. The Flemish government is also setting up initiatives[146] to facilitate access to (sensor) data.

   A third possibility is to use online services[147] that offer persistent URLs (PURLs), and have these redirect to your own service URLs.

   → Applied: the domain 'waarnemingen.be' is already allocated to Natuurpunt[148] and is therefore under its own management, so it is a possible candidate for use in URIs.

3. Define the **pattern** your URIs should follow. This is particularly useful when you want to see from the URI structure whether it is …

   - a separate definition of a data concept (object, attribute or relationship) – e.g. /ns;
   - the identification of a data object – e.g. /id;
   - a representation of the data object – e.g. /doc.

   For inspiration for your own pattern, the 'Flemish URI standard for data'[149] can be used, which imposes binding rules on the Flemish authorities.

   → Applied: it is decided to use `https://waarnemingen.be/id/{soort,observatie,foto...}/{numeriekveId}` as a pattern for the identification of species, observations, photos, etc. It will be ensured that these URIs are resolvable, by showing an associated web page with more information (under the same URI) when calling up these URIs in a browser.

4. Provide URIs not only for your data (points), but also for the concepts behind your data (the entities, attributes and relationships from your domain model), so that users and systems can find a (machine) readable definition via that URI.

   Note: this step is a little less obvious than offering URIs to identify data points, but it is a prerequisite for arriving at 4-star open data and making your data accessible via the **RDF** format.

   You can create these definitions yourself and offer them under a URI on your own domain, but you can also reuse existing **LOD** databases and **ontologies** online (e.g. DBPedia, Wikidata, Geonames, Darwin Core[150],etc.) to obtain URIs for definitions of data concepts.

   → Applied: it is decided not to create one's own definitions and make them accessible URIs, but to reuse the Extensible Observation Ontology.[151] It already defines, among other things, an observation with all its characteristics. The URI for the definition of 'observation', for example, is http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#Observation.

---

[146] Context broker/data pipeline, sensor data platform, etc.
[147] For example http://identifiers.org, http://www.purlz.org/ and https://archive.org/services/purl/.
[148] See https://www.dnsbelgium.be/nl/whois/info/waarnemingen.be/details
[149] https://data.vlaanderen.be/cms/VlaamseURI-StandaardVoorData_V1.0.pdf
[150] See https://dwc.tdwg.org/text/ for an application of Darwin Core in text-based files such as CSV.
[151] https://bioportal.bioontology.org/ontologies/OBOE

5. Use the persistent URIs instead of the classic (alpha)numeric identifiers in all your access forms (in datasets, web services and on websites). This will encourage users to use the URIs to refer to your data. After field names you can place the URI that refers to the definition of the data concept in the same way.

→ Applied: on the website 'waarnemingen.be', observations are offered in the following form:

Observation	Bird species	Location
https://waarnemingen.be/id/observatie/1	Crane – *Grus grus*	Westkapelle
https://waarnemingen.be/id/observatie/2	Black-throated loon – *Gavia arctica*	Ostend
https://waarnemingen.be/id/observatie/3	Common loon – *Gavia immer*	Willebroek

6. For advanced users: offer your project data in **RDF**, **Turtle** or **JSON-LD**[152] format. RDF is the native language of the semantic web. Just as a database can be queried with SQL, RDF sources can be searched with SPARQL.

Practical example

Use of URIs to refer to data concepts and objects

https://orcid.org/0000-0001-8888-635X
The 'ORCID iD' is a unique number that can be requested by **researchers**. This URI can, for example, be mentioned in all scientific publications of that person. This way, you can never confuse two researchers with the same first and last name when you want to consult their list of publications.

https://www.uniprot.org/uniprot/P98161
This URI identifies a human **protein** in the UniProt database.

http://dbpedia.org/resource/Julius_Caesar
This URI identifies the **topic** 'Julius Caesar' in the DBpedia Knowledge Base.

https://data.vlaanderen.be/id/adres/983512
This URI identifies the **address** of Scivil in the Flemish address register.
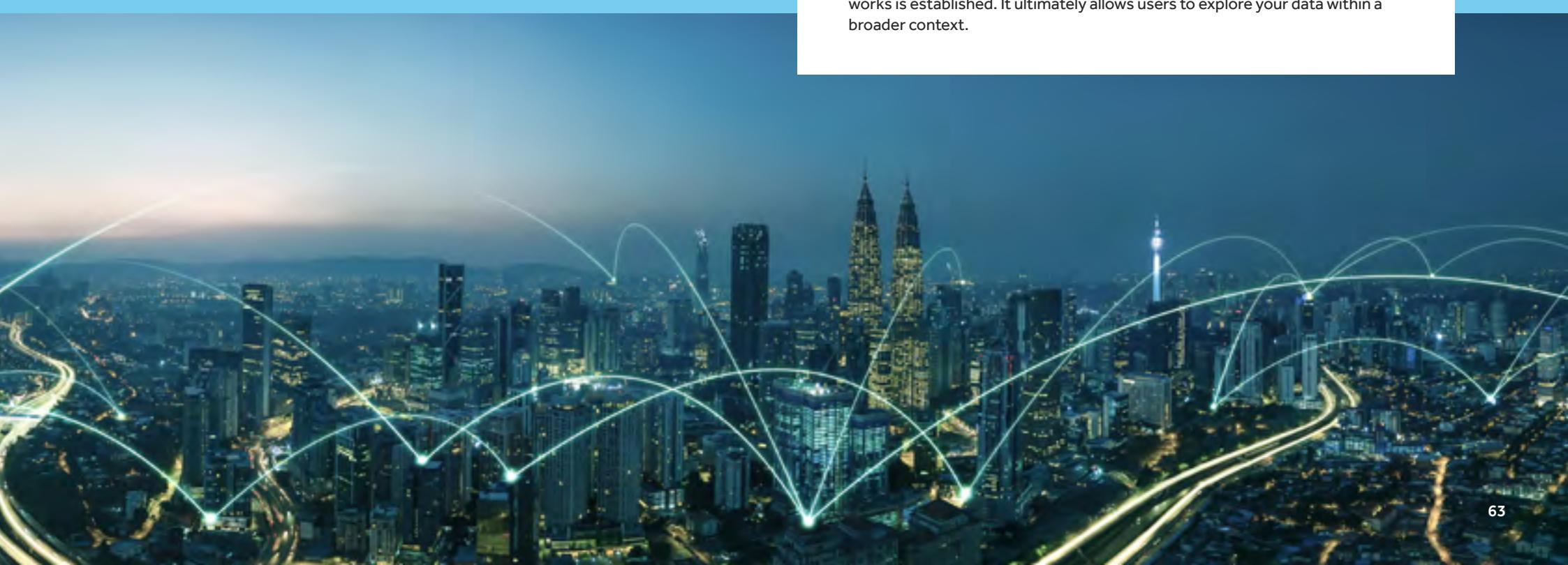
https://data.vlaanderen.be/ns/weg#EuropeseWeg
This URI identifies the **definition** of a European road in the context of the Flemish road register.

---

[152] See https://www.w3.org/RDF/, https://www.w3.org/TR/turtle/ en https://json-ld.org/.

20. *Incorporate your project data into the Linked Open Data network.*

Have you provided your project data with globally unique, persistent identifiers? Then you can make your data even more valuable by linking it to other information sources on the web. In this way, you can offer richer datasets, often without much extra effort, in which the relationship between the project data you have collected and other (scientific) sources or reference works is established. It ultimately allows users to explore your data within a broader context.

IV.

Data standards and formats

Linked Open Data

How can you make your project data part of the Linked Open Data network?

Building further on the example of waarnemingen.be:

1. Investigate what additional information would enrich your own dataset.
   → Applied: it would be desirable to provide more context to the locations where observations were made, so that targeted, geographical enquiries can be made later. In addition, the aim is to link bird species with information about them on Wikipedia.

2. Reuse existing **LOD** databases (see https://lod-cloud.net/) to search for sources that offer the information sought, including persistent URIs. This own data is then linked with these URIs.
   → Applied: for the locations, the official municipalities in the Flemish address register and the Dutch **BAG** will be used (GeoNames was also considered); for the Wikipedia content, Wikidata can be used. The URIs provided by these sources are used to enrich own data on the website.

Observation	Bird species	Location
1	Crane – *Grus grus*	Westkapelle
2	Black-throated loon – *Gavia arctica*	Ostend
3	Common loon – *Gavia immer*	Willebroek

3. Since you are now referring to external sources, which are not under your own control, it is important to draw up a **policy** for **dealing with this external information**, and also with broken or incorrect hyperlinks. For example, you can check whether incorrect information can be reported at the source, or use a tool that looks for 'broken links'. Also be careful about the sources you link to: what minimum quality do they need to be?

Practical example

Conversion of dataset to Turtle and linking to DBpedia elements

As part of the 'Nieuws van de Groote Oorlog' project of Het Archief, a script was made (https://github.com/viaacode/hetarchief2lod) to to make the metadata of the archived newspapers suitable for querying on the semantic web, namely by offering this metadata as Turtle **triples**. Using automatic character recognition in combination with DBpedia Spotlight[154], text from photographs was linked to elements from DBpedia. This linked its own LOD dataset to another LOD source, allowing queries across both sources. Read more at https://meemoo.be/nl/publicaties/historisch-krantenarchief-gepubliceerd-als-linked-open-data. Although not a citizen-science project, this project, especially in combination with possible other linked data, lends itself perfectly to citizen science.

2007 **2010**

2014 **2020**
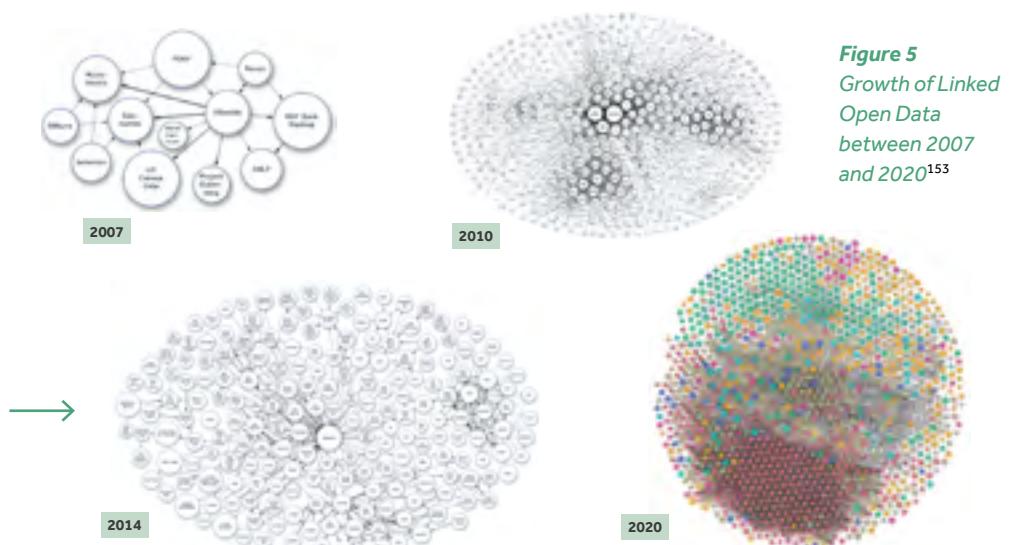
Figure 5
Growth of Linked Open Data between 2007 and 2020[153]

---

[153] Bron: https://lod-cloud.net/#diagram Licence: CC-BY. Check also Frank & Zander (2019) for more information.
[154] See https://www.dbpedia-spotlight.org/

V. Metadata

21. *Provide your data with the richest and most accurate metadata possible.*

Metadata means 'data about data'. It constitutes the 'leaflet' for (published) research data and results and is crucial for making your data findable.
The more richly the data is described by means of metadata, the easier it is for a potential user to ascertain the origin, quality and 'scope' of the dataset. It also allows the user to assess the suitability of the data for his/her intended purpose, and to contact the publisher. Since data sets are subject to change, the associated metadata must also be kept accurate and up to date.

Common core of basic information

- Most metadata standards share a **common core** of basic information that, regardless of the domain in which the citizen-science project is located, should be included when creating metadata for your dataset:

- **Identifier of the described dataset**
  Usually metadata, and the dataset it describes, are separate files. The link between the two must be made explicit by including the (globally) unique identifier of the dataset in the metadata. This could be, for example, an identifier that is automatically generated when describing your dataset in a (meta)data portal. Note: where in *principle 19* we spoke of persistent identifiers for individual concepts or measurements in a dataset, here we mean the identifier for the whole dataset. This can be a **GUID** or a **URI**.

- **Title, description, themes and keywords**
  These fields allow your dataset to be found on portals describing hundreds or thousands of datasets. It is therefore advisable to pay due attention to this.

  > **Tip**
  >
  > Make sure that the title is identifiable across all datasets (series) and their versions, that it contains legible Dutch (avoid abbreviations) and is not too long. Also make it future-proof by including some kind of versioning in the title if you expect further versions, observations or recordings to follow.

  > **Tip**
  >
  > You do not need to include your project name in the title: you are better off using other fields (such as 'author', 'publisher' or 'description') for that.

Field	Bad example	Good example
Title	Leuvenair PM2.5/PM10 '19	Westkapelle
Description	Sensor data	Sensor data (PM2.5/ PM10) from the 'Leuvenair' citizens' initiative for the mapping of air quality in Leuven
Themes	Environment	Environment, Health
Keywords	(not filled in)	Open data, air quality, citizen science

- **Author and/or publisher with contact information**
  Provide up-to-date contact information (project coordinator, data manager and/or organisation) and working email addresses, telephone numbers and websites so that interested users or data suppliers can contact you for questions, suggestions and ideas.

- **Information on (usage) rights**
  Under this heading you can specify your open licence (see *principle 2*), with any restrictions on the use of the dataset. The desired source/name can also be included here.

- **Language, format, topicality and (spatial/temporal) boundary of the dataset**
  This is important metadata to get a good picture of the data and its usability. For example, one user may specifically be looking for recent data (past two years) on Flemish territory, while another user is mainly interested in historical data on a single city. The format (see *principle 18*) can also be given here, which gives an idea of the software applications with which the data can be viewed or processed. Topicality can translate into a 'last updated' date, or the indication of the update frequency of the dataset.

- It is strongly recommended that you **go one step further** in describing metadata by enriching it with the following information:

- **Sources, collection methods and process steps**
  How was the data created or acquired? Especially for citizen-science projects, it is an added value to describe the methodology (in which citizens play a role) concisely but completely. Were other data sources used, and if so, what transformation or enrichment steps were taken? The metadata may contain references to other, related or complementary sources.

- **Data quality**
  With what accuracy and precision were measurements made? Is the data raw or post-processed? (How) were calibrations done? How large was the sample? What sensors were used? Compliance with data standards (i.e. the dataset follows the semantics described in a recognised information model) is also worth mentioning.

- **URLs of web services**
  If the project data are not (only) offered as a dataset, but (also) via a web service, then the metadata of the dataset can refer to the URLs of these services. For example, geographical data sources are typically accessed not only as datasets, but also through OGC-compliant map services, such as a **WM(T)S**.
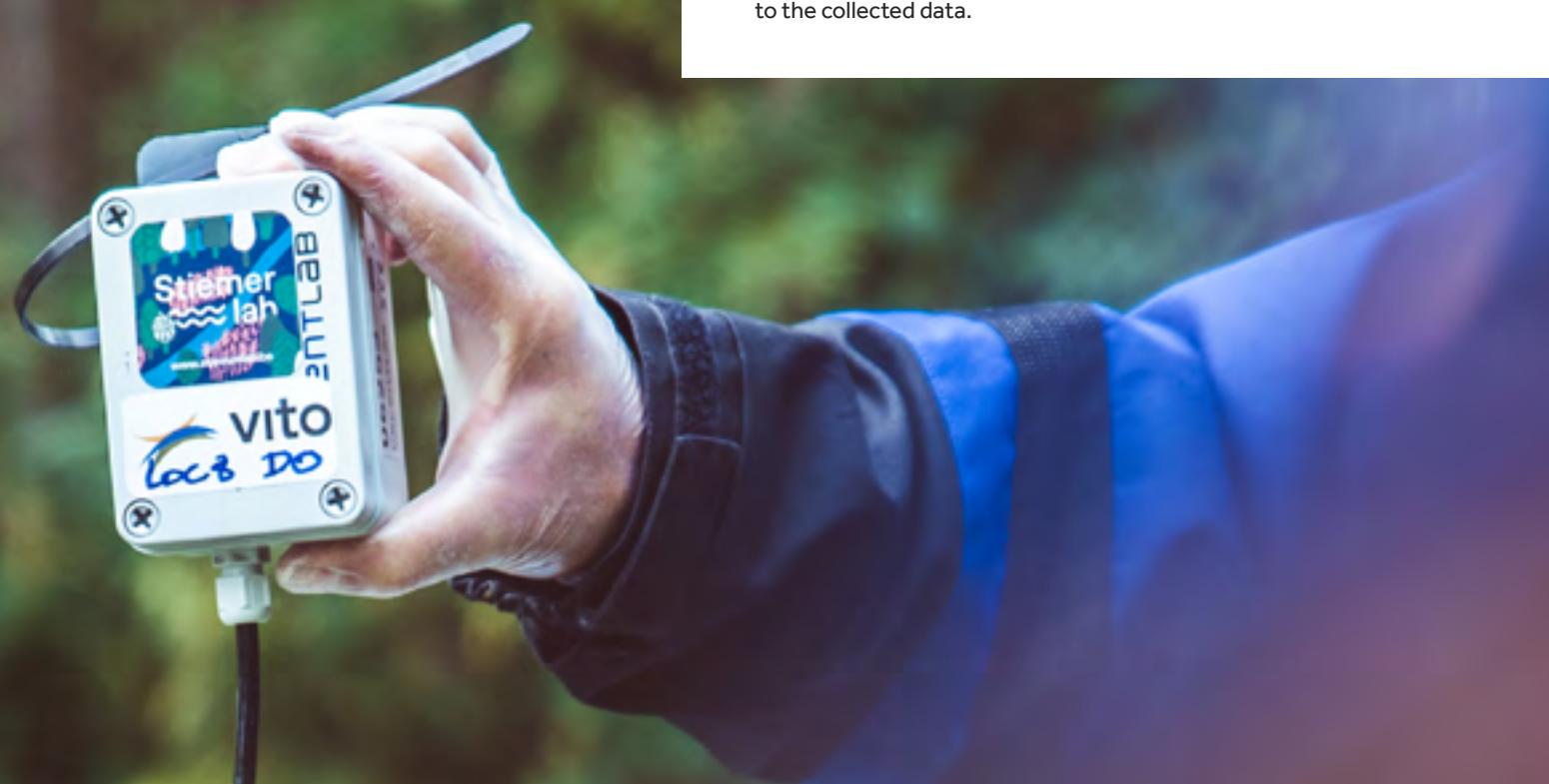
- **Application profile or object catalogue**
  Last but not least, it is very useful to work on a documented **application profile**[155] or a neat object catalogue[156]. This takes the form of a website or a loose document. It describes the entities, attributes, relationships and code lists that are part of your conceptual data model (see *principle 16*) and/or dataset, accompanied by their definitions and data types. This enables a user to look up the exact meaning of an attribute or code, for example.

---

[155] *Example of application profile for Building Register:* https://data.vlaanderen.be/doc/applicatieprofiel/gebouwenregister/
[156] *Example of object catalogue for Basiskaart Vlaanderen (GRB):* https://metadata.vlaanderen.be/srv/dut/cataog. search#/metadata/9650dfb1-9e7f-4ce2-8ee9-98a1b7733ee4

22. *Remember to capture the metadata as close to the source as possible.*

As project leader of a citizen-science project you make sure that all necessary metadata is captured with the data that is collected or analysed in your project. Unlike in regular research, you do not always have control over which measuring device or analysis tool is used in certain citizen-science tasks. If that information is relevant (and not in contradiction with the GDPR), you technically create possibilities for your participants to add this metadata to the collected data.

For example, when participants add photos to their observations, you could have the participant indicate in a form when registering which type of camera he/she uses. Another way to get hold of this metadata is to ask the camera settings to be adjusted so that parameters such as (f-stops, shutter speed, ISO, GPS location) are automatically stored with the photos themselves.

So think beforehand not only about the way in which data will be collected, but also about how it can be provided with metadata immediately, and as close to the source as possible.

Practical example

As stated in **_principle 15_**, the international weather platform <u>WOW</u>[157] collects **metadata** from weather stations to assign an overall quality label to each station, and display the station's properties for the website user. Citizen scientists who send data from their weather station to the platform can also enter the specifications and environmental factors of their station. Information provided includes: the motivation of the citizen scientist, whether the station belongs to a school or an official network, whether there is a webcam, what time zone the station is in, and when the most recent observation was made. Furthermore, the dashboard also displays the location (open/closed), the observation hours and an index showing whether the station is in an urban area or not.

[157] _https://wow.meteo.be/nl/_

70
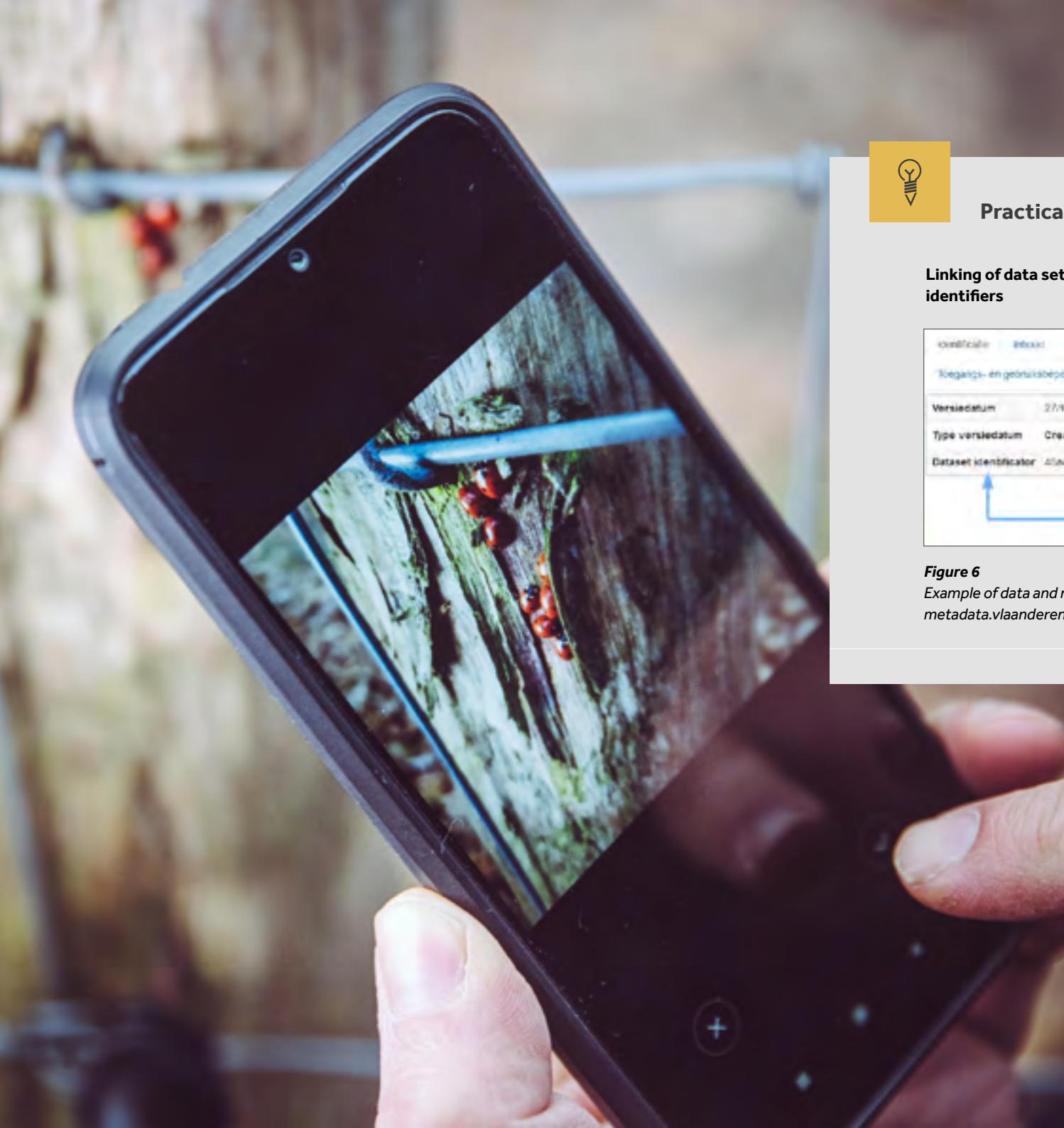
23. *Assign a globally unique and persistent identifier to your set of metadata.*

As with the dataset, the associated set of metadata should also be referenced with a globally unique and persistent identifier. Here, too, a **GUID** or **URI** can be chosen that is automatically generated by the metadata portal. Through these identifiers, data and metadata can be inextricably linked, and dataset and metadata set can be referred to unambiguously.

Practical example

Linking of data set to associated metadata set via globally unique identifiers

Figure 6
Example of data and metadata linkage from
metadata.vlaanderen.be

24. *Use the right standard for your metadata.*

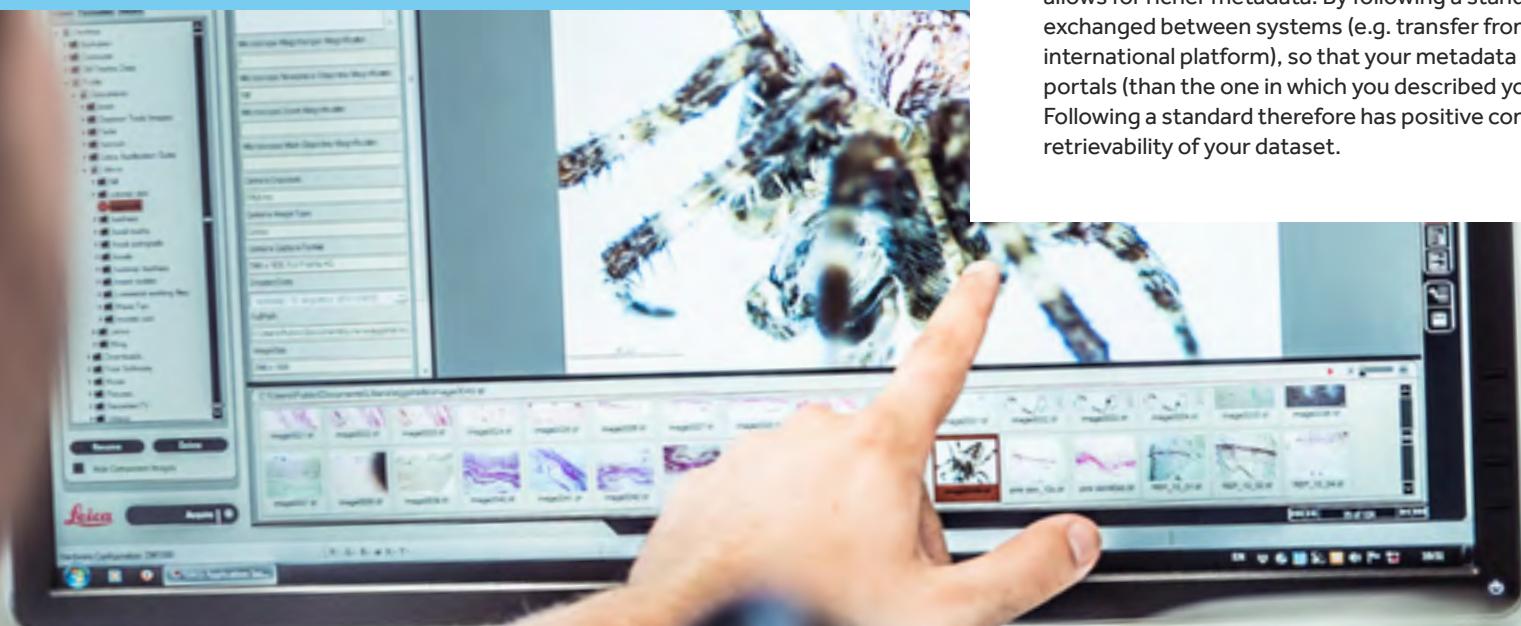There is also an abundance of standards for metadata descriptions. Many are extensions or specialisations of the <u>Dublin Core</u>[158], a set of 15 elements (later standardised as **ISO** 15836) that allows metadata to be attached to resources across domains. The fields in the Dublin Core represent the lower limit. However, it is better to choose another (domain-specific) standard that allows for richer metadata. By following a standard, metadata can be easily exchanged between systems (e.g. transfer from regional/national to international platform), so that your metadata can also be found in other portals (than the one in which you described your metadata the first time). Following a standard therefore has positive consequences for the retrievability of your dataset.

[158] *Drawn up at a workshop in 1995 in* <u>*https://www.wikidata.org/wiki/Q988937.*</u>

Metadata standards

Metadata is also prepared according to **(domain-specific) standards**. Depending on the research field to which your project belongs, scientific data standards may have been developed.

> **Tip**
>
> Try to make your metadata conform as much as possible to the established standards in your research field. Achieving compliance with such a metadata standard is usually a matter of choosing a portal that supports the standard in question and offers guidance on filling in the necessary fields.

Here is a non-exhaustive overview of metadata standards:

- *Dublin Core (Metadata Element Set)*
  https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
  Very generic. See the 15 'elements'. Suitable for describing physical and digital resources (such as books or images).

- **DCMI**
  Metadata Terms
  https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
  Extension of the 'core' with additional 'terms'. Equally application- and domain-neutral.

- *DataCite*
  http://schema.datacite.org/
  Intended for the standardised citation and retrieval of sources (e.g. publications). Uses DOI numbers.

- **SDMX**
  https://sdmx.org/
  Extension of the 'core' with additional 'terms'. Equally application- and domain-neutral.

- **DCAT**
  https://www.w3.org/TR/vocab-dcat-2
  Comprehensive W3C standard intended for describing datasets and data services in (government) catalogues. Builds on DCMI.

- *Metadata DCAT*
  https://data.vlaanderen.be/doc/applicatieprofiel/metadata-dcat/
  Application profile based on DCAT that focuses on gathering information about generic datasets, distributions and services made available by a government.

- *DCAT-AP-VL* [159] https://data.vlaanderen.be/doc/applicatie-profiel/DCAT-AP-VL
  Application profile derived from Metadata DCAT for the publication of generic open data in the Flanders Open Data Portal (VODAP).

- **DDI**
  https://ddialliance.org/
  Intended for describing and attaching metadata to data from surveys and other observation methods used in the social, economic, behavioural and health sciences.

- **EML**
  https://eml.ecoinformatics.org/
  Standard developed for earth and environmental sciences and ecology. Suitable for taxonomic and geographical data and for citing sources. Used by GBIF platform.

- *GDI-Vlaanderen metadata profile* [160]
  http://www.geopunt.be/nl/voor-experts/metadata
  Profile based on relevant ISO standards and the INSPIRE Framework Directive, specifically intended for the publication of geographical datasets (coordinates, zones, distances, height measurements, etc.). Mappable on the GeoDCAT standard.

---

[159] Citizen-science projects can make use of this provided that the metadata is published by a Flemish municipality or government agency.

[160] Combination of ISO 19115 (metadata for geographical datasets), ISO 19119 (services), ISO 19110 (object catalogue) and 'INSPIRE Implementing Rules for Metadata'. Publication required by government here too.

In the citizen-science world, a number of metadata standards have been developed for both projects and datasets. These are based as far as possible on existing metadata standards. It is therefore more than worthwhile consulting them.

- **PPSR** *CORE* **DMM**
  https://core.citizenscience.org/docs/dataset/
  Suitable for attaching metadata to datasets in citizen-science projects. A dataset is defined here as a collection of data collected in a project by citizen scientists using a single protocol. You can therefore associate a project with multiple datasets.

- *PPSR CORE PMM*
  https://core.citizenscience.org/docs/project
  Suitable for attaching metadata to citizen-science projects.
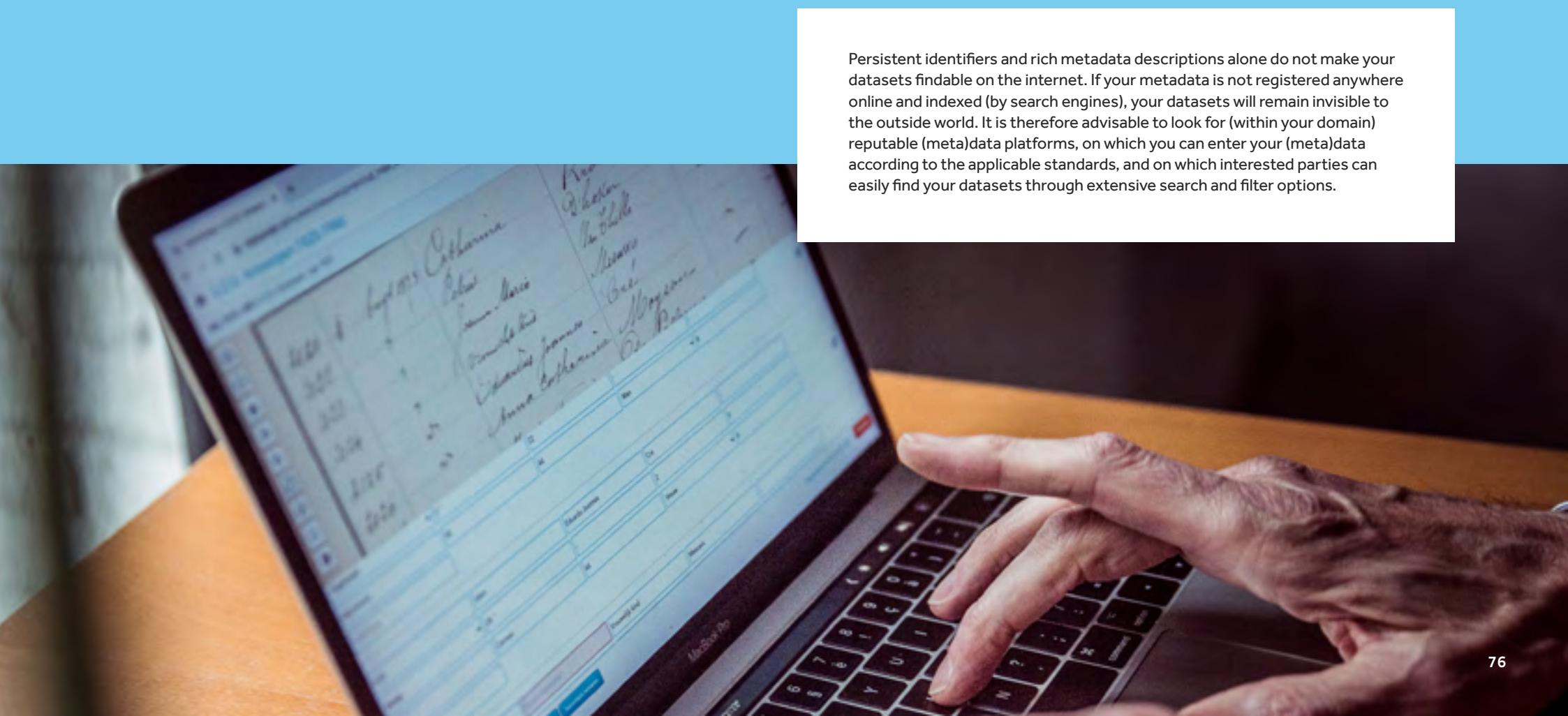 Important elements included in this model are:

  - *The title and description of the project*
  - *Ownership and contact details*
  - *The duration of the project (start/end)*
  - *Project partners*
  - *Funding programme or source*
  - *Information about public participation (in what way do citizen scientists participate? In what role?)*
  - *Links to other (project-related) sites and resources*
  - *Project-related graphic elements*

Practical example

The **PPSR** Core **PMM** is already being used in the following citizen-science web platforms to exchange project data: Scistarter.org[161], CitSci.org[162], Atlas of Living Australia – Biocollect[163], CitizenScience.gov.[164] A more detailed example of such a data exchange is described on the website of the **PPSR** Core **PMM**.

[161] https://scistarter.org/
[162] https://www.citsci.org/CWIS438/Websites/CitSci/Home.php?WebSiteID=7
[163] https://ala.org.au/biocollect/
[164] https://www.citizenscience.gov/#

25 *Make your metadata, and thus your datasets and research results, findable by registering them on a searchable portal.*

Persistent identifiers and rich metadata descriptions alone do not make your datasets findable on the internet. If your metadata is not registered anywhere online and indexed (by search engines), your datasets will remain invisible to the outside world. It is therefore advisable to look for (within your domain) reputable (meta)data platforms, on which you can enter your (meta)data according to the applicable standards, and on which interested parties can easily find your datasets through extensive search and filter options.

Metadata platforms

It is not possible to make a complete list here either, but the following **metadata platforms** are worth mentioning:

- **FRIS** *research portal of the Department of* **EWI**
  https://researchportal.be/nl
  Over 35, 000 new publications are registered in FRIS every year.[165]
  The metadata of these scientific articles and book (contributions) is searchable by knowledge institution, type, year and discipline. The actual data is linked, e.g. via **DOI** numbers.

- *Metadata Vlaanderen*
  https://metadata.vlaanderen.be
  This portal can be used by citizen-science projects to track down interesting (government) data, but also to attach metadata to their own project data and make it findable. You can describe geographical (according to GDI-Vlaanderen metadata profile) and open data (according to DCAT-AP-VL), and soon also open APIs (according to DCAT metadata). Here, too, links are provided to the data at your chosen location. The metadata entered is automatically shared with federal and European metadata nodes.
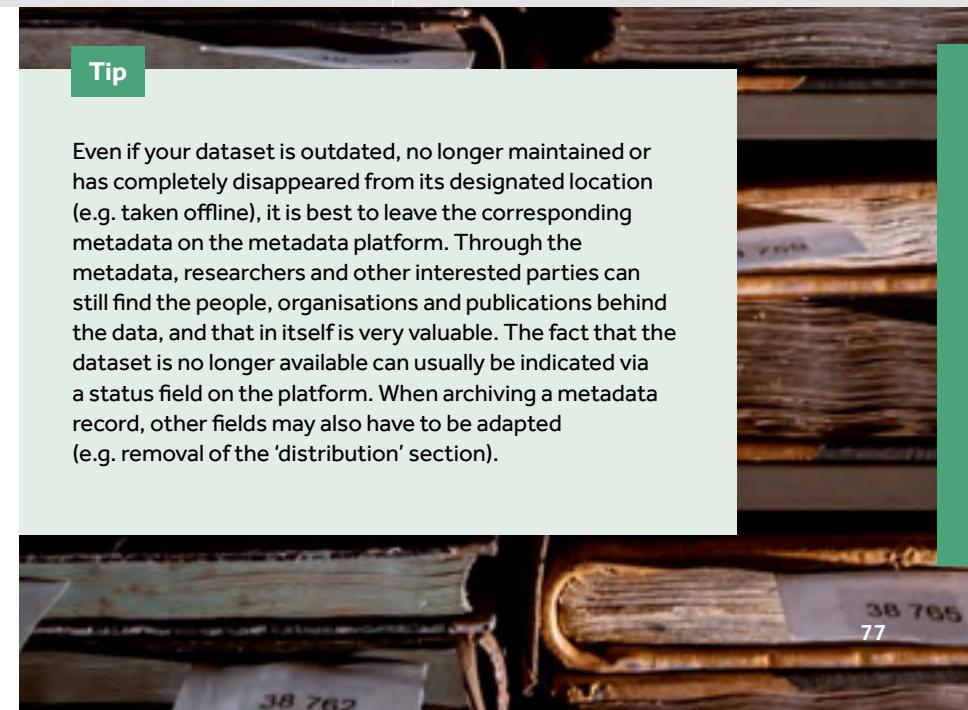
Practical example

Metadata from the website 'waarnemingen.be'[166] (Natuurpunt) flows (semantically operable thanks to the use of standards, and technically interoperable via documented **APIs**) automatically through to the GBIF portal, making the data collected by volunteers findable for a very wide audience of research institutions and researchers.

The impressive list[167] of more than 18 million observations on Belgian territory.

Meise Botanic Garden[168] is one of the research institutions that supplies datasets to the GBIF portal.[169] The estimated 2 million observations (15 datasets) can be used by others for analyses. Scientists who extract data from the portal for research and publish their results afterwards refer to the source of their data. This has already earned the Botanic Garden 423 citations in publications.

Tip

Even if your dataset is outdated, no longer maintained or has completely disappeared from its designated location (e.g. taken offline), it is best to leave the corresponding metadata on the metadata platform. Through the metadata, researchers and other interested parties can still find the people, organisations and publications behind the data, and that in itself is very valuable. The fact that the dataset is no longer available can usually be indicated via a status field on the platform. When archiving a metadata record, other fields may also have to be adapted (e.g. removal of the 'distribution' section).

---
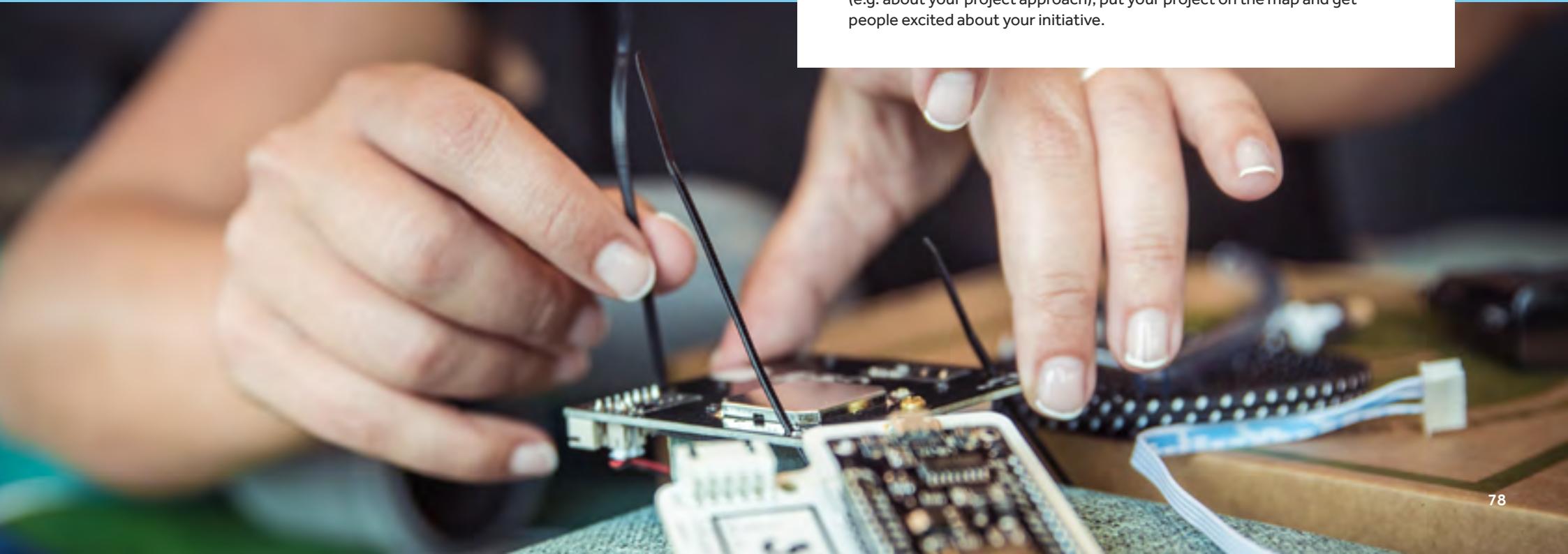
[165] Bron: *https://researchportal.be/nl/cijfers*
[166] *https://waarnemingen.be/*
[167] *https://www.gbif.org/occurrence/search?country=BE&publishing_org=4d3ceea8-5699-439d-a899-decac9cbbdac*
[168] *https://www.plantentuinmeise.be/nl/*
[169] *https://www.gbif.org/publisher/a344ee9f-f1b7-4761-be2c-58ee6d741395*

26. Make your citizen-science project findable for colleagues and volunteers.

There are metadata standards and portals for describing datasets and publications, but also for making citizen-science projects and the people and organisations associated with them findable.

The importance of this should not be underestimated. After all, it allows you to get in touch with fellow project coordinators, share insights and knowledge (e.g. about your project approach), put your project on the map and get people excited about your initiative.

How do you make your project findable?

Some websites on which you can make your citizen-science project findable, or expand your network:

- *Iedereen Wetenschapper*
  https://www.iedereenweten-schapper.be/
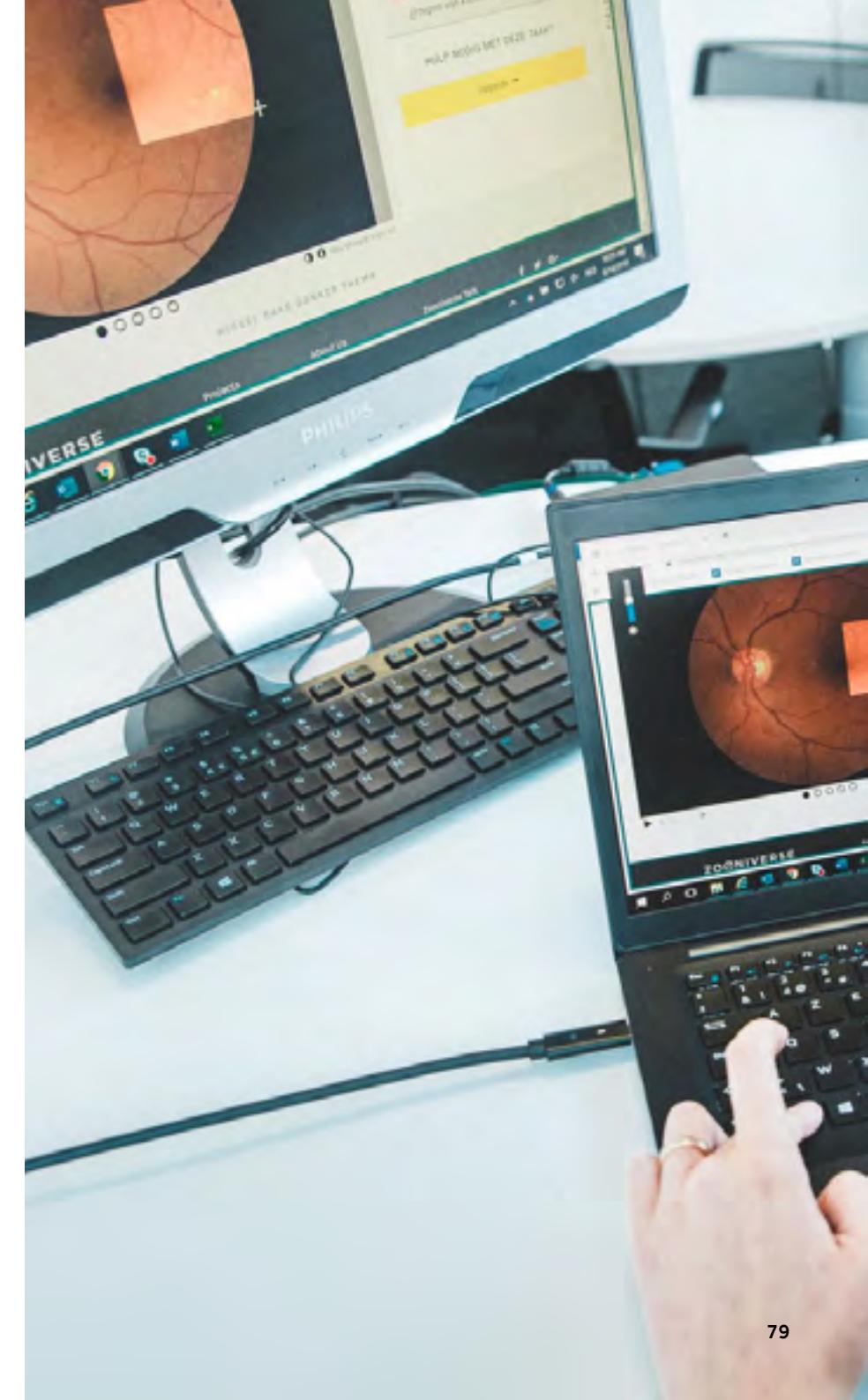 At Flemish level, this website, set up by Eos Wetenschap and the Jonge Academie, is the place par excellence to make citizen-science projects visible and involve committed volunteers.

- *EU-Citizen.Science*
  https://eu-citizen.science/
  A database for citizen-science projects has recently been set up at European level, in which metadata from projects can be entered. The website also offers resources, tools and training for your project.

- *SciStarter*
  https://scistarter.org/
  The website SciStarter originated in America, but now contains a very extensive database of citizen-science projects worldwide. Citizen scientists can search SciStarter using different search terms, and can also monitor their individual contributions to different citizen-science projects.

- **FRIS** *research portal of the Department of* **EWI**
  https://researchportal.be/nl
  In addition to publications, FRIS also features organisations, researchers and scientific projects.
  The counter has now reached more than 40,000 registered projects that can be searched by knowledge institution, fund provider, discipline and start date, among other things.

> **Tip**
>
> The PPSR Core metadata standards provide a separate PPM model[170] for attaching metadata to citizen-science projects (see ***principle 24*** for further explanation). These are already being used by various international citizen-science portals such as SciStarter.

---

[170] *More info: https://core.citizenscience.org/docs/project*

References

Buyle R., De Vocht L., Van Compernolle M., De Paepe D., Verborgh R., Vanlishout 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), 16 pp. DOI: https://doi.org/10.1088/0004-6256/144/2/66.

Porter W.T., Barrand Z.A., Wachara J., DaVall K., Mihaljevic 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 scientst. 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. Final report within the framework of the Citizen Science call of the Department of Economy, Science and Innovation (EWI) of December 2017. Honeybee Valley and Research Institute for Nature and Forest, Brussels. 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

List of abbreviations

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

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

Term	Definition	Sources & references
Anonymisation	Anonymisation, also known as data masking, is a method whereby personal data is processed in such a way that it can no longer be used to identify a person. This operation is irreversible. According to the General Data Protection Regulation (GDPR), anonymised data is therefore no longer personal data. Often, the anonymisation process involves a combination of pseudonymisation and data aggregation.	*https://www.fgsupport.nl/wat-is-anonimisering*
Application profile	An application profile is a specification for data exchange for applications (software applications) that fulfil a certain use case. It allows the imposition of additional restrictions, such as the definition of cardinalities or the use of certain code lists, in addition to a shared semantics (see 'data model' and 'vocabulary'). An application profile can serve as documentation for analysts and developers.	*https://data.vlaanderen.be/ns/*
Authentic data source	An authentic data source is the most complete, high-quality collection of data, maintained electronically, which is useful or necessary for public authorities in carrying out the public interest tasks entrusted to them or in complying with obligations imposed on them.	Article III.66 §1 of the Administration Decree: *https://codex.vlaanderen.be/PrintDocument.ashx?id=1030009&datum=&geannoteerd=false&print=false*
Citizen Science	See 'Citizen Science'	
Code list	Literally: a list of codes. Used in data models when a property (attribute) does not take on a single value (such as a piece of text or a number), but can be given different possibilities (codes) from a predefined (but extendable) list. Example: - A sensor has a name. That name can take as value a piece of text ('string'). - A sensor has a status. The status must be selected from a code list with the following possibilities {online, offline, unknown}.	(own definition)

Term	Definition	Sources & references
Citizen Science	Citizen Science is scientific research carried out wholly or partly by non-scientists (citizens), often in cooperation with or under the guidance of professional scientists. The citizens who participate in a citizen-science project are called citizen scientists. Citizen science is a broad term, which covers many meanings. Citizen scientists can perform a wide range of tasks depending on the project they are contributing to; they carry out counts and observations, build and install sensors, annotate or transcribe images and texts, carry out data analyses, devise research questions and research methods, and report or disseminate research results.	*https://www.scivil.be/aan-de-slag/voor/citizen-science-iets-voor-mij*
Data Management Plan	A data management plan (DMP) is a formal document that describes how researchers will deal with the collected data before, during and after the research: how data will be collected; the size and format of the data; the way in which the data will be stored; who will have access to the data both during and after the research. Such a plan also contains information about the legal and ethical aspects of data.	*https://www.ugent.be/nl/onderzoek/datamanagement/datamanagementplan*
Data model	A data model describes how the data in an information system is structured. A distinction is made between • the conceptual data model; • the logical data model; • the physical data model.	*https://nl.wikipedia.org/wiki/Datamodel*

List of terms

Term	Definition	Sources & references
Data standard	Data standards are necessary when different parties want to cooperate and exchange data with each other, but there is a confusion of terms (the 'jargon') within the domain in question. In this case, domain experts may decide to define concepts from reality together, and to define the structure of (or relationships between) data. A data standard is the documented result of this exercise. If the various parties use the same concepts and definitions from the data standard, data exchange is possible and (semantic) interoperability is achieved. A data standard can take the form of a vocabulary, ontology, conceptual data model (= domain model) or application profile, or a combination thereof. In data models, many nuances can be added through the chosen modelling language to describe the domain in detail and unambiguously.	(own definition)
Data aggregation	The merging of data (e.g. by taking an average value, or the minimum and maximum value).	(own definition)
Data serialisation	The conversion of a data object into a format that can be can be stored (sequentially) or transmitted over a computer network. In addition to the data itself, (part of) the data structure is also delivered in the serialised format. This metadata is used to interpret the data correctly in case of deserialisation.	*https://nl.wikipedia.org/wiki/Serialisatie*
DBpedia	A machine-readable, 'linked data' version of the articles on Wikipedia.	*https://nl.wikipedia.org/wiki/DBpedia*
DOI number	Digital Object Identifier: a unique number assigned to a journal article.	*https://library.uic.edu/help/article/1966/ what-is-a-doi-and-how-do-i-use-them-in- citations#:~:text=A%20DOI%2C%20or%20Digi- tal%20Object,to%*

Term	Definition	Sources & references
FAIR	To fully exploit the potential of research data, it is necessary to make them as Findable, Accessible, Interoperable and Reusable as possible in the research ecosystem (Force 11, 2014; Wilkinson, et al., 2016; GoFAIR, n.d.). The FAIR principles consist of 15 facets. The main thing is that research data must be FAIR not only for people, but also for machines. The FAIR principles have become an indispensable part of the data management landscape and are at the basis of the blueprint for the 'European Open Science Cloud'.	To: *https://datasupport.researchdata.nl/start-de-cursus/i-in-vogelvlucht/begrippen/rdm-open-en-fair* *https://datasupport.researchdata.nl/start-de-cursus/iv-oogstfase/european-open-science-cloud*
FAIR data management	FAIR data management is the set of decisions and measures taken during the research data life cycle to deliver research data as FAIR as possible. What is important here is the choice of data licence and data format, the definition of metadata and data documentation, naming conventions, etc.	*https://datasupport.researchdata.nl/start-de-cursus/i-in-vogelvlucht/begrippen/rdm-open-en-fair*
Sensitive personal data	Sensitive or special personal data is data that, according to GDPR legislation, must be given increased protection. This is data about a person's ethnic or racial origin, political, religious or philosophical beliefs or trade union membership, genetic information, biometric information (e.g. the iris of an eye, or a fingerprint), health data or data relating to a person's sexual orientation. Data relating to criminal prosecutions also belongs to this category.	To: *https://www.gegevensbeschermingsautoriteit.be/professioneel/themas/gevoelige-gegevens* *https://onderzoektips.ugent.be/nl/tips/ 00001772/*
GUID	Globally Unique Identifier, a number and letter combination automatically generated by software and assumed to be globally unique. This makes GUIDs ideal for use as identifiers, since identifiers must refer uniquely to an object. The typical form of a GUID is as follows: `{3F2504E0-4F89-11D3-9A0C-0305E82C3301}`	To: *https://nl.wikipedia.org/wiki/Globally_unique_identifier*

List of terms

Term	Definition	Sources & references
Interoperability	Products, systems or organisations are interoperable if they can work together without limitations. To achieve this, standards, protocols and procedures are needed. Interoperability is referred to as 'semantic' (definitions of data concepts), 'technical' (technical guidelines), 'organisational' and 'legal'.	To: *https://nl.wikipedia.org/wiki/Interoperabiliteit* *https://vloca-kennishub.vlaan*
Linked Data	Structured data (data) that is linked to other data and is therefore more usable in semantic queries. The method is based on the technology of HTTP URIs and RDF. 'Semantic queries' are queries on the semantic web, which is a further development of the World Wide Web in which the meaning of data is embedded in web pages so that automated systems can also perform intelligent searches.	To: *https://nl.wikipedia.org/wiki/Linked_data* *https://nl.wikipedia.org/wiki/Semantisch_web*
Metadata	Metadata is data describing the characteristics of certain data. So it is actually data about data. For example: when you measure temperature, you can include in the metadata that the unit in which you measure is Celsius, not Fahrenheit.	*https://nl.wikipedia.org/wiki/Metadata*
Ontology	In information science, an ontology represents the representation of a set of concepts in a (knowledge) domain. An ontology establishes definitions of concepts and relationships between concepts. The Web Ontology Language (OWL) is a W3C standard for defining web ontologies.	*https://www.ensie.nl/bibliothecarium/ontologie*
Open data	• 'Open' means that data can be freely consulted, used, modified and shared by anyone, with conditions imposed at most to safeguard its origin and openness. • 'Open data' is a term that is used to describe freely available information. The conditions under which this information is available are described in licences and conditions of use. In the case of open data, the aim is to minimise the restrictions on reuse.	To: *http://opendefinition.org* *https://okfn.org/opendata/* *https://opendata handbook.org/gu ide/nl_BE/what-is-open-data/* *https://nl.wikipedia.org/wiki/Open_data*

Term	Definition	Sources & references
Open Licence	A licence is a kind of promise you make to others that they can consult, use, distribute your work (data, database, software, or any other work covered by the licence) without you taking any action against them. That is, if they comply with the conditions in your licence agreement. So you choose what may happen to your information, and the licence makes that explicit.	(own definition)
ORCID iD	The 'ORCID iD' is a unique number that can be requested by researchers. This URI can, for example, be mentioned in all scientific publications of that person. This way, you can never confuse two researchers with the same first and last name when you want to consult their list of publications.	
Preprint Service	A preprint service is a web portal where scientific articles are available before they go through the peer review process. By sharing their preliminary text online, scientists present their findings to each other even before the full process of publication is completed. NB: preprint services do not review or quality-check the content of the articles, please be aware of this. Examples are ArXiv, BiorXiv and MedrXiv.	(own definition)
Pseudonymisation	Pseudonymisation is a method in which identifying data is replaced by encrypted data (the pseudonym) using a specific algorithm. The algorithm can always assign the same pseudonym to a person, allowing information from different sources to be combined. After encryption, it is therefore still possible to identify the person concerned: the algorithm can be used to reverse the encryption. Because the encryption process is reversible, pseudonymised data about an identifiable person is still personal data under the General Data Protection Regulation (GDPR). After pseudonymisation, the data can be traced back to a person.	*https://www.fgsupport.nl/wat-is-anonimisering*

Term	Definition	Sources & references
PSI Directive	Public Sector Information Directive of the European Commission, also known as the Open Data Directive. This directive defines the conditions under which public sector data must be released (including for licensing).	To: *https://overheid.vlaanderen.be/organisatie/informatiemanagement/omzetting-psi-richtlijn* *https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32019L1024&from=EN*
Semantic interoperability	Achieving cooperation in terms of terms used (naming) and meaning of the data exchanged. See also 'Interoperability'.	(own definition)
SSL certificate	SSL stands for Secure Sockets Layer: this is an extra security that you can apply to make your website safer against data leakage. Websites with an SSL certificate have a URL that starts with 'http**s**://'.	To: *https://nl.wikipedia.org/wiki/Transport_Layer_Security*
Triple	A semantic 'triple' is a set consisting of three elements that contain meaning about data. The form is (subject, predicate, object). The subject is essentially the object being described. The predicate describes a property, characteristic or relationship. Finally, the object is the value of that characteristic. Example: (`"Italië","heeftAlsHoofdstad","Rome"`)	To: *https://en.wikip edia.org/wiki/Semantic_triple* *https://nl.wikipedia.org/wiki/Resource_Description_Framework*
URI	A uniform resource identifier (URI) is an internet protocol element. It is a one-off name for a 'source', a piece of information, data or similar. URIs follow the format. `protocol:` `//[gebruiker:wachtwoord@]host(naam)[:poort]/path`	*https://nl.wikipedia.org/wiki/Uniform_resource_identifier*
Vendor lock-in	'Vendor lock-in' makes a customer dependent on a supplier for products and services, as he is unable to change suppliers without substantial switching costs or inconvenience. Applied to software, it means, for example, that file formats or protocols are incompatible with those of competing software.	*https://nl.wikipedia.org/wiki/Vendor_lock-in*
Vocabulary	A word stock, lexicon, vocabulary [vocabularium] or vocabulary [vocabulaire] is the collection of words used by persons (within a domain).	To: *https://nl.wikipedia.org/wiki/Woordenschat*

SCivil
Citizen Science
Vlaanderen

Kapeldreef 75, 3001 Leuven   info@scivil.be   www.scivil.be

Flanders
State of the Art   EN

DEPARTMENT OF
ECONOMY
SCIENCE &
INNOVATION