

Digital Object Identifiers (DOIs)

A Digital Object Identifier (or DOI) is a character string used to uniquely identify an object. Metadata describing the object is stored in association with the DOI name, including an URL which leads to where the object can be found.

Why assign a DOI to my dataset?

In science DOIs have been widely used by publishers of peer reviewed journals for over 10 years to uniquely identify a specific article. Associating DOIs to scientific publications has not only increased the traceability of the cited literature but also simplified the maintenance of citation indexes which serve today to **assign academic credit** to scientists for their work.

As is true for claims based on information from other publications, in scholarly literature, **whenever** and wherever a **claim relies upon data**, the **corresponding data should be cited**. Moreover there is growing international support for the idea that **dataset citations** should also lead to **academic credit**. Assigning DOIs to your dataset can facilitate these processes for the following reasons:

- a. The DOI system can serve to **identify** and locate a **specific version** of a dataset
- b. Datasets to which a DOI is assigned should to stay **persistently accessible**.
- c. After assigning a DOI to a dataset, the data can no longer be changed. This will assure that claims based on data can always be checked.
- d. Minimum metadata - including the names of the data author(s) – needs be associated with the DOI during registration. Having the data authors associated with the DOI will assure that they can receive the **appropriate credit and recognition** for their work
- e. DataCite will store the associated metadata and disseminate it to other initiatives like the **Data Citation Index** (Thomson Reuters) and ORCID. The Data Citation Index will **link** the indexed **datasets to publications** citing them. ORCID allows researches to claim authorship of their datasets, which will help resolve confusion caused by researches with a similar name.
- f. by assigning a DOI with the dataset, the data publisher attests that the dataset answers to certain **quality standards** and contains all the necessary metadata to make the dataset usable by other scientists. This will encourage other scientists to use it for their analyses and then cite your dataset in their publications.

Requirements

There are some requirements for the data to be published as a data publication using the VLIZ systems.

- ➔ A dataset formally published with a DOI will be stored by VLIZ and will remain publically available indefinitely.
- ➔ The availability of dataset may remain restricted for a pre-defined period. For example if the DOI is requested for the submission of a paper in a peer reviewed journal, the dataset can remain restricted until publication of this paper.
- ➔ The data provider has the option to submit corrected or updated versions and have a new DOI assigned to them. A new version will however not replace any earlier version to which a DOI has been assigned.

After submitting the dataset, VLIZ will create an IMIS discovery record for the dataset using the information provided in the submit form. The data provider will be requested to review and approve this record. The data provider will also be asked to provide any missing mandatory metadata.

Before publishing VLIZ will check whether the dataset meets basic quality standards and contains all the necessary metadata to make the dataset re-usable by other scientists. Parameters description can be provided in another sheet of the same spreadsheet (MS Excel), in the table fields (MS Access), or in a separate text file. Methodologies can be provided by referencing associated papers. VLIZ provides a checklist which can help make sure all information is available.

Data submit guidelines

Dataset description

Datasets submitted to VLIZ will be described in the Integrated Marine Information System (IMIS) using the information provided in the [online submit form](#). The IMIS record of the dataset will primarily serve to inform other scientists about the existence of your dataset. If you decide to make your dataset publicly available, a [DOI can be assigned](#), and the IMIS record will contain a link from where the dataset can be freely downloaded.

A good dataset description contains at least following elements:

- 1) **Title:** VLIZ recommends that the title gives an indication of the content of the dataset; the time period and the region of sampling.
For example: "Macrobenthos collected in the Westerschelde between 1965 and 1974"

- 2) **Citation:** The proposed dataset citation should reference the dataset, and not a paper using or describing the dataset. Therefore the dataset title should be included in the dataset citation. VLIZ recommends the following format.

Data Creator(s) Person; Data Creator(s) Institute (division); (Publication Year). Title.
Identifier

For example: Braeckman, U.; Marine Biology Research Group - Ugent, Belgium (2014). Ocean acidification effects on nitrification in natural sediment communities from Belgian part of the North Sea. <http://dx.doi.org/10.14284/1>

Data Creator(s) Person or *Data Creator(s) Institute* can be omitted if not applicable. However, the dataset citation needs to contain at least 1 data creator.

- 3) **People involved:**

- a) **Contact person:** is the person who should be contacted in cases of questions regarding the content of the dataset or any data restrictions. This is also the person who is most likely to stay involved in the dataset the longest.

It is highly recommended that this refers to a person, or if this is not possible (e.g.: someone is working in an institute only for a short term: a project, a master thesis,...) it can refer to a division of an institute.

- b) **Data creators:** All people and/or institute (divisions) who are responsible for the creation of the data(set). Data creators should receive credit for their work and should therefore be included in the citation.

4) **Abstract:** An abstract is a short summary (max 1000 characters = about 5 lines) of the dataset. It contains an indication of its content.

5) **Availability:** VLIZ advocates free data exchange. If you choose to make your data freely downloadable, it can be published formally with a DOI, under one of the different [Creative Commons licenses](#).

You can choose to make your dataset not publically downloadable, in this case no DOI will be assigned.

VLIZ recommends completing all the additional information requested in the submit form. If there are **publications** (or **websites**) based on/or describing the dataset or other datasets related (or preceding) to this dataset it is highly recommended that they are referenced as well. A link will be made between the publications, (related) datasets, projects and websites.

Dataset Checklist

When preparing your dataset for submission please take care to provide all necessary information for other scientists to interpret the data correctly. VLIZ provides a checklist which can help make sure all necessary or available information is included:

Technical metadata

1. **Parameters:** Is it for all parameters unambiguously clear what was actually measured (e.g. wet weight vs ash free dry weight)?
2. **Units:** For all parameters, are the associated units documented?
3. **Used instruments, protocols and information on calibration:** this information can be provided in an extra tab in an excel file, a separate file or by referencing a publication.
4. **Georeference stations:** Do all stations have coordinates? Is the used coordinate system indicated? Are measuring depths or heights recorded?
5. **Dates and time:** Is the time zone specified (UTC, GMT, MET, CST, ...)? For notation [ISO format 8601](#) is recommended.
6. In **what framework** (project, context,...) was the data collected, or gathered? Where can the user find additional information?,...

QC process

Although VLIZ will perform a basic technical quality control on the data submitted for a data publication, we recommend the data provider to carry out a first quality control before submission.

For Marine Biological data, taxonomy can be standardized by referencing the matching `aphia_id` from the World Register of Marine Species. This way, differences in spelling, or accepted species will be corrected. <http://www.marinespecies.org/aphia.php?p=match>. Additional data quality checks (Georeferencing, data format (Darwin Core, OBIS),...) can be performed using the LifeWatch e-lab <http://www.lifewatch.be/en/e-lab>

Versioning

Once the submitted dataset has been published with a DOI, it will stay publically available as is. Therefore to avoid unnecessary versioning VLIZ recommends that your dataset is submitted at key moments (e.g. the end of the project or the study).

Datasets related to long term (monitoring) projects can be dynamic. Not only the content and the structure might change, but also the researchers and institutes involved and the context in which the data is gathered. If data is continuously being added, updated versions can be submitted at regular intervals. These different **snapshots** of the dataset will be available from the same IMIS metadata discovery record. When the dataset has changed in a large extent (e.g. it was integrated with another database, it was split up in different subsets, its maintenance was handed over to a different institute, ...), it can be preferable to consider it a new dataset (version) with a **new metadata discovery** record. In the new metadata discovery record you can highlight this change by using a new dataset title (and citation), and you can change the metadata in such a way that it would not be applicable to the previous version. The discovery records will be linked, as to provide an overview of the history of the dataset and avoid confusion.