

Checklist for the general review of a data package

This checklist serves to guide a *general review* of a data package submitted to [ERIC](#). A *General review*, as opposed to a *domain specific review* can be conducted by people without expertise in the scientific field the data package relates to.

1 Bibliographic metadata

DORA-id or DOI In case this is a publication data package and if the paper is already published, this field contains a URL that resolves to the landing page of the associated paper.

Title The title of a publication data package is: "*Data for: [title of paper]*".

Authors All persons that can claim scientific credit for the data are in the author list. Authors are given in the correct format.

Keywords A publication data package contains at least the keywords contained in the associated paper (unless the paper covers additional topics unrelated to the data).

2 Domain specific metadata

Variables All measured variables that pertain to the main question of the research are listed. If there are relevant variables not in the list, they have to be given in the field *Notes* at the bottom.

Substances All chemical substances that are represented in the data are listed. Both, "*scientific names*" and "*general terms*" are given.

Organisms All organisms represented in the data, from which data was derived, or which are in the focus of the associated research are listed, both, in the field "*Taxa*" and in the field "*Organisms*".

Systems In case the question "Where or in which medium did the measurements or observations take place?" can be answered in an obvious way, the field "*Systems*" is not empty.

Timerange is as specific as possible.

Spatial Extent and Geographic Name(s) If the data has (a) geographic reference(s) (i.e. measured in the field, not in the lab),

1. "*Geographic Name(s)*" is filled out exhaustively.
2. "*Spatial Extent*" contains valid and correct GeoJSON or a GeoJSON-representation obtained from geojson.io is in the field "*Notes*" at the bottom of the form.

Visibility and Status are set to Eawag and (truthfully) to complete, respectively.

Curator and Usage Contact are set to people who are expected to stay at Eawag for a longer period of time (e.g. not PhD students).

Notes contain information about meta-data that could not be properly entered into the form, such as variables measured that are not offered for selection in the field "*Variables*".

3 Resources

There is a resource (file) `README.txt` or `README.md` that contains *specific metadata* (see [Documentation "scientific metadata"](#) in the *Eawag archiving guide*, <https://doi.org/10.25678/000066>).

All resources (files and URLs) are at least implicitly mentioned in the README-file.

Resources that are URLs resolve to the expected target.

Files have sane file-names (see [File naming conventions](#) in the *Eawag archiving guide*, <https://doi.org/10.25678/000066>).

Files open as expected.

The character encoding for text files (`.txt`, `.csv`, `.md`, ...) is UTF-8.

There are no proprietary file-formats, if can be avoided (e.g. `.doc` and `.docx` should almost always be converted to PDF).

There are no spurious files that do not belong to the package (e.g. `.DS_Store` and `__MACOSX`)

Tabular data is well-structured:

- It is rectangular.
- It has a header row.
- Numerical values have units.
- The provenance of the data is clear: There is a reference to a publication that contains a description of the methodology, or the package itself contains that description.
- Third party data is properly credited.