Publish your data

A step by step guide on how to create a record in Zenodo for the CLEX Data Collection

Table of contents

Table of contents (continued)

Here we will show step by step how to create a new record for your data in Zenodo.

If your data is already published on another repository then you can probably reuse most of the information already provided. You should contact us before starting to fill in the record form and we will advise on the best way to proceed. In some cases we might be able to map the metadata of the original record to the zenodo form.

Even if you have created already a record in Zenodo for your dataset, this step by step guide will still help you make sure all the information required for your record to be accepted in the CLEX Data Collection is displayed correctly.
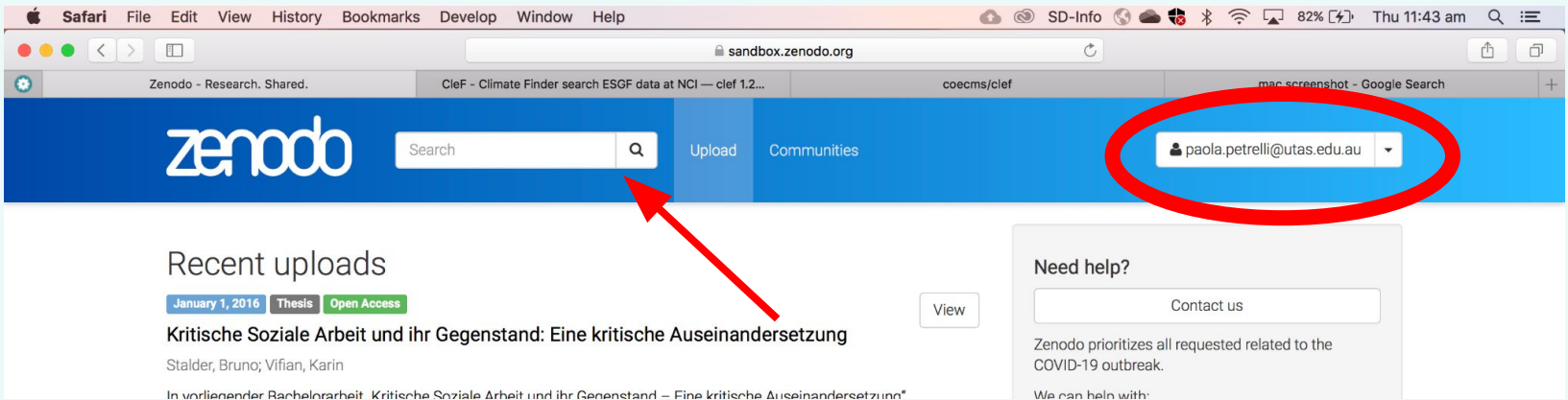
Login into your Zenodo account if you haven't already, your name should appear at the top on the right.

Click on the **upload** button

The **upload** page shows all your records that have not yet been published.

You can click on any of them to edit them.

At the top right of the page you can click on the **New upload** button to create a new record

The **New upload** page shows the form to fill for your new record

The first step is uploading your code. You can 'drag and drop' your files or open the upload window by clicking on the **Choose files** button.

- You need at least one file, if you have already published your code on another repository you should upload only a readme file here.
- Overall your files cannot be more than 50 GB in size, if this is not enough for your data please contact us to discuss other options, also you can review our [publishing guidelines](#).

You can now see the file you added and their size. At this stage they have not yet been uploaded. Once you added all the files you can do so by clicking on the **Start upload** button

The **Communities** box is next. You can search for a community and select it simply by typing its name or part of it. As you can see above typing CLEX shows the CLEX Data Collection, that we want to select in this case.

Pay attention to the **required** fields

Next required field is the **Upload_type**, with actual dataset its straightforward just select "Dataset".

If the record is describing data documentation, than you have a few choices depending on the documentation format:

Publication, Poster, Presentation, Video/Audio, Lesson

In the **Basic Information** section, you will enter all the essential information to describe your data. Most of the fields in this section except **Description** will be used to identify your data.

The **Description** field is going to be the first thing a user would look at to decide if your data is potentially useful to them.

The **Digital Object Identifier** field should be left empty unless a DOI already exists for the record.

Otherwise a DOI will be added to the record at the moment of publication.

If you want to know what your DOI will be before the actual publication, for example to add a reference when submitting a paper. Then you can click on the **Reserve DOI** button.

This will now assign a DOI to the record, so it cannot be used elsewhere, but the DOI will not be active until the record is actually published.

The **Publication date** is automatically assigned as the current date, if the record was previously published then, as for the DOI, the official one should be used.

!!! remember that re-publishing a data that already as a DOI is a breach of our Collection policies and also of the Zenodo policies. So use the original DOI wherever applicable.

Title *

Weather Research and Forecasting (v3.6) model outputs from 2-km Kuala Lumpur urban climate experiments

Required.

It is worth thinking carefully about a **Title** for your record, it is what will be used by anyone who reference your data. While you can potentially change the **Title**, even after a record has been published, it is important to get it right and be as descriptive as possible.

Most people will skim quickly through a list of records returned by a query, unless they actually click on one specific record, all they will see is the **Title** and the start of the description. Also, any word in the **Title** is used as a keyword by the search engine. In the example above the author tried to include:

- the model name and version - Weather Research and Forecasting (v3.6)
- the kind of data - model outputs
- simulation resolution - from 2-km
- spatial domain- Kuala Lumpur
- subject - urban climate
- that there is more than 1 experiment - experiments

NB that having used the full model name rather than an acronym clarified to any user not familiar with WRF that this are weather forecast related experiment

The **Authors** field includes 3 columns:
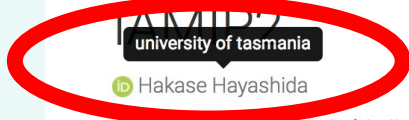
- the first is for the name to be entered as "Surname, name"
- the second is the affiliation
- the third is the orcid

The first two are required , the orcid is optional but highly recommended.

The authors name will be shown under the title, the affiliation will appear when hovering with the mouse over the name. If an orcid was supplied there will be a small orcid icon next to the name linking to the orcid account.

To add more authors just click on the "*+ Add another author*" button

This dataset contains model outputs from the Weather Research and Forecasting model 3.6 over the Malay Peninsula at 2 km spatial resolution and covering the period 2008-2012.

It includes two different simulations: an experiment with urban areas and a second simulation where cities are replaced with the dominant surrounding vegetation type. Precipitation and 2m temperature were directly extracted from the raw model outputs. Near-surface water vapor mixing ratio divergence was calculated using 2-m mixing ratio and 10-m wind components from the raw outputs. 4D variables (i.e. temperature, water vapor mixing ratio, cloud water and ice mixing ratio and wind speed) were extracted from the raw outputs and interpolated to height above the ground levels from the model native levels.

The project that generated this data was aimed at understanding how cities modify precipitation extremes in a tropical region. We make use of a convection-permitting model coupled with an urban canopy model and in-situ observations to quantify the effect of cities on heavy rainfall, focusing on precipitation intensities at sub-daily scales.

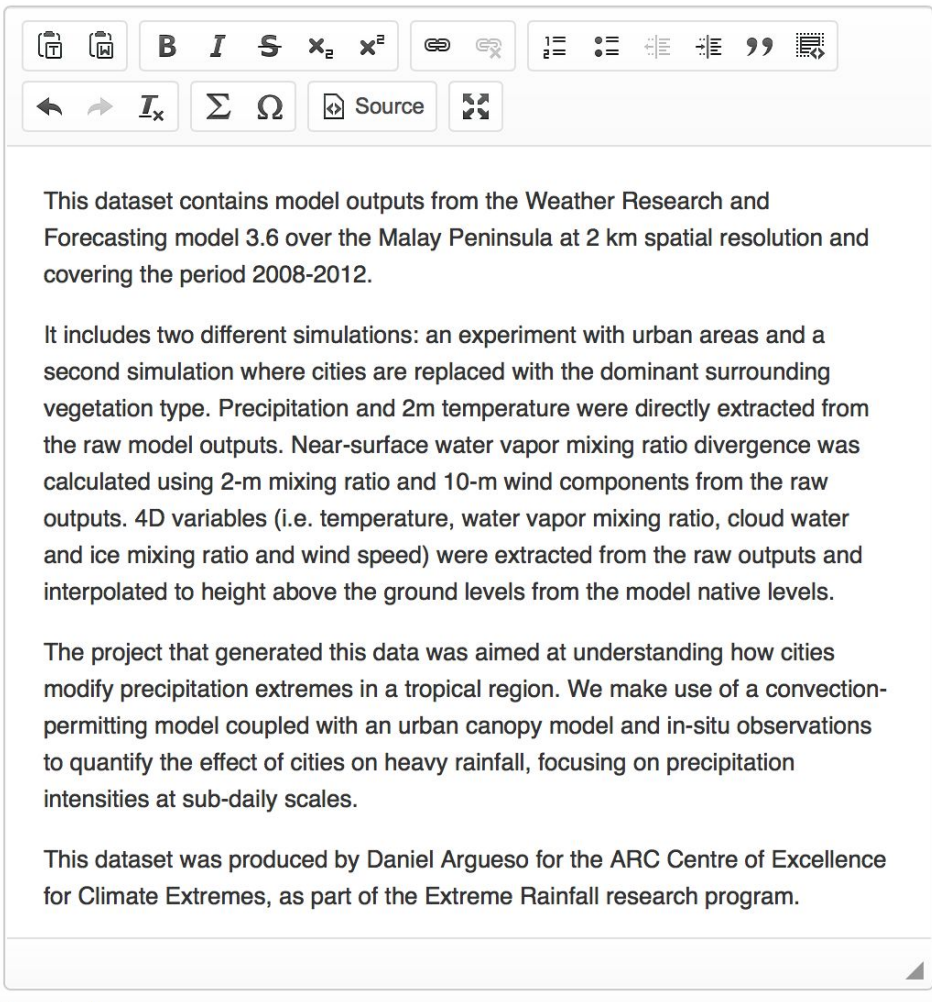This dataset was produced by Daniel Argueso for the ARC Centre of Excellence for Climate Extremes, as part of the Extreme Rainfall research program.

Description should include:

- Context in which data was created
- Data and code used to generate the data
- Spatial and temporal extension and resolution
- Data format
- If possible, a list of all the variables included in the dataset
- If this is a new version how it is different from previous ones
- What is the best use for this data and, more importantly, if the data is not suited to a particular use
- Any existing documentation and/or training resources

Links to any related material should be added also in "Related Information".

The **Version** field is really important to identify your data, as well as being required. Even if you are not planning new releases now, you might need to correct or extend the data in the future.

Conventions around data versioning are still been developed. The ARDC provides some guidance.

Here are some suggestions to work a versioning strategy for a "stable" dataset, for datasets which are continuously updated please refer to next slide:

- any change in the actual data should be accompanied by a new version. NB if you're main data is stored on Zenodo adding or removing files will force a new version to be created
- changes to the metadata do not need a new version
- Most often numbers or time stamps are used to identify versions, whichever approach you choose, it is important to be consistent. It is impossible for a user to work out which version is the latest if two versions are called v2.0 and v2020.

Datasets which are continuously updated require particular attention. As a starting point you should apply the same versioning strategy used for a "stable" dataset. However, you also need to consider the following:

- The documentation should report how the dataset is updated and how frequently
- The actual data cannot be hosted on Zenodo as this would force the creation of a new version with any file change
- It is better to add new files, if possible, rather than updating existing files
- It is even more important to have a versioning strategy to distinguish between the constant change to the data, allowed as part of the same version, and what circumstances will warrant the creation of a new version
- When the way the data is produced changes, then there should be a new version, even if the older data in the timeseries is not affected by the change in methodology.

As a DOI should always refer to exactly the same object there is not consensus on how to treat continuous data. Some of the common strategies are:

- Publishing the data at regular time intervals, i.e. every year. This is done at times delaying the data release, so the data is always covered by one DOI or updating the data continuously and leaving gaps in the DOI coverage. In both cases each new DOI will be reflected in a new version
- Publishing the DOI and updating the data continuously until a change in data production warrant the creation of a new version. Users are then required to add a timestamp indicating when the data was accessed, in the dataset citation.

Language indicates the language used to write the record, you can use eng for English

Keywords can be added by clicking on the **+Add another keyword** button.

Any word included in the title is already used as keyword by the Zenodo search and so it does not need to be added here.

We have a controlled vocabulary for **Keywords**, this is to encourage using the same terms across the collection.

If you execute a query from the main collection page then the query will include only the collection records, and we want the keywords in this space to be consistent.

You should always include the data type, other categories to use wherever is applicable are:

- If the data is derived or has a strong connection to another dataset/model you should use the official acronym

- discipline

- domain

- frequency

- data re-use (ie. forcing, analysis.)

Additional notes is an optional field that you can use to highlight some important information.

In this example it highlights the preferred citation for this dataset and how the data can be accessed as its main publication site is the NCI thredds server.

You could also use it to add a disclaimer.

As this field is displayed in a more prominent way it is a good idea to include here also the contact details or how to retrieve them, for example from the ORCID.

Contact details are a requirement for the record to be accepted in the Collection.

We are covering all the options to handle them in a separate slide.

License is the last of the required sections and both of its fields are required.

Access rights are by default **Open Access, this is a requirement for your record to be accepted in our Collection**. However, you have the option of applying an Embargo if you want to delay access until a later date, that you should specify in the **Embargo date** field. This is shown only when you select Embargoed Access.

License - we recommend Creative Commons Attribution 4.0, or you can choose another open access license. Creative Common licenses are the default in Zenodo. You can start typing the name of the license to see more options.
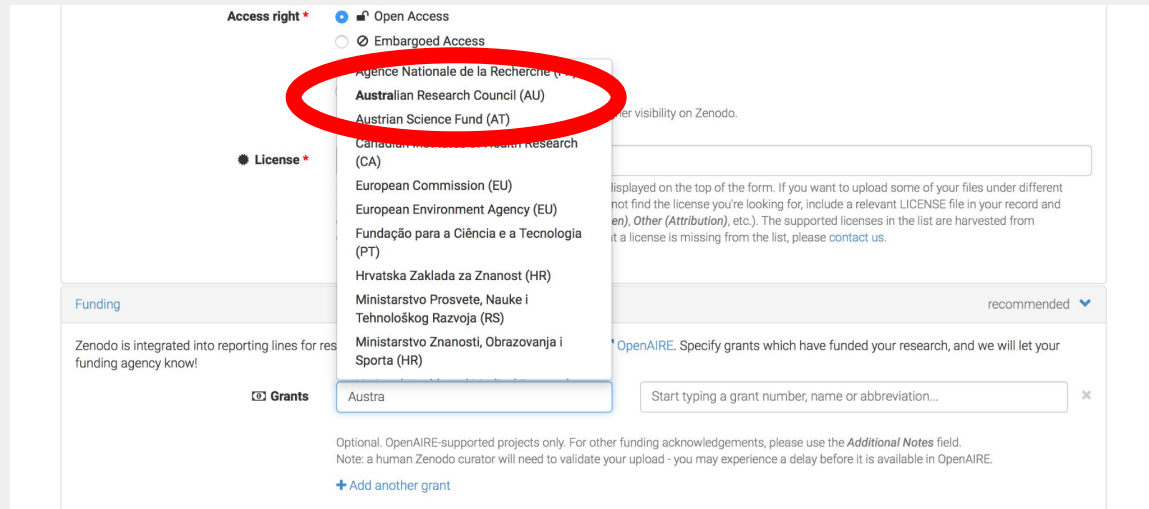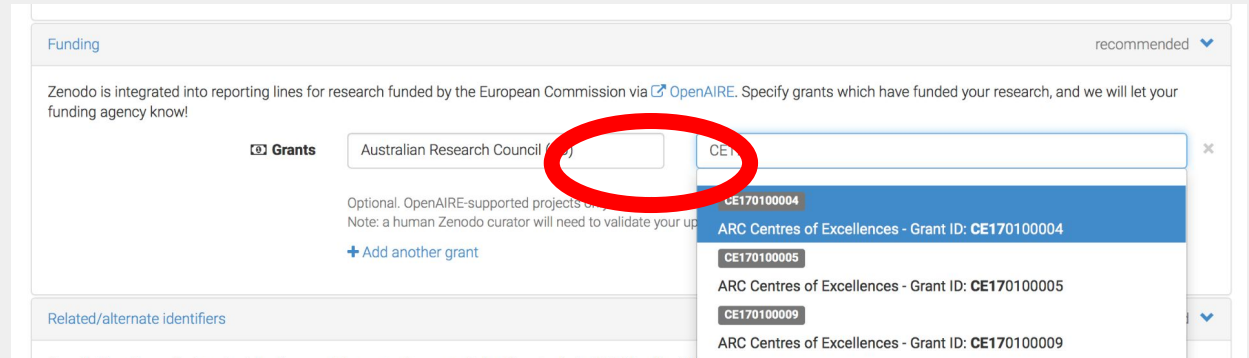
If your data was funded by a grant you can list the funder and the grant code in **Grants.**

There are funders listed from all over the world. In this case, as I want to list the CLEX grant, I choose the ARC as funder.

When I start typing the Grant ID for CLEX, Zenodo shows me the list of available ARC grants to choose from.

The Grant ID for CLEX is

CE170100023

Unfortunately currently only funders and grants listed can be chosen. If your grant is not available, you can still add it as a **Related Identifier** (see next slide)

Related/alternate identifiers are used to add links to related resources. There are 3 columns:

- the first is to list the **resource url**

- the second is for the **relationship** between your code and the linked resource.

These are both required and there is an extensive controlled vocabulary to define the relationship

- the third, which is optional, also uses a controlled vocabulary to define the **resource type,** an example of relevant types is "Journal article" as visible above

Rn = H + LH + G

When the latent heat flux (LH) is converted to the same unit as the water budget terms, the hydrologic variables satisfy:

P = LH + Q + deltaS

The uncertainty of each flux is computed from its discrepancy with in-situ observations

This dataset was created by Dr Sanaa Hobeichi as part of her PhD with the ARC Centre of Excellence for Climate Systems Sciences, and as part of the ARCCSS research program - The role of land surface forcing and feedbacks for regional climate.

Preferred citation: Hobeichi, Sanaa, 2019: Conserving Land-Atmosphere Synthesis Suite (CLASS) v1.1. NCI National Research Data Collection , doi: 10.25914/5c872258dc183 Access to the data is via the NCI geonetwork record in related identifiers, details are also provided in the readme file.

Files (3.1 kB)   ⌄

Name	Size
CLASS_v1-1_readme.txt	3.1 kB	Download

md5:0f2cd5aaa374e719cd7be52158ecf358 ⍰

Citations ⍰ 0   ⌄

Show only: ☐ Literature (0) ☐ Dataset (0) ☐ Software (0) ☐ Unknown (0)   Search 🔍

DOI
DOI 10.25914/5c872258dc183
Keyword(s):
Climatology   Surfacewater Hydrology
Surface terrestrial energy budget
Surface terrestrial water budget
Grants:
Australian Research Council:
ARC Centres of Excellence for Climate System Science (CE1101028)
Related identifiers:
Source of
10.1175/JCLI-D-19-0036.1 (Journal article)
Alternate identifiers:
https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f4282_2934_6733_5289
https://researchdata.ands.org.au/conserving-land-atmosphere-v-10/1360835
Communities:
CLEX Data Collection: collection of climate science data
License (for files):
☐ Creative Commons Attribution Non Commercial Share Alike 4.0 International

Related identifiers will appear in a prominent position on a right side panel, as shown in the image. They should be used wherever applicable, examples with a suggested relationship are shown in the next slide.

Resource	Suggested relationship	Resource	Suggested relationship
data repository	is cited by this upload	A code or other software required to use this data	is required by this upload
Online documentation or training	documents this upload	A dataset used as base for this	is the source this upload is derived from
official metadata record, if the data is already published in another repository	is alternate identifier of this upload	A previous/new version of the same data (NB these links are automatically generated if they are versions of the same Zenodo record	Is previous version of this upload is new version of this upload
publications related to the data: technical reports, paper describing the data	describes this upload documents this upload	Paper referencing/citing the data	references this upload cites this upload
published data derived from this dataset	has this upload as its source	scientific paper based on the data	has this upload as its source

Contributors can be added the same way as authors but with an additional field to clarify their role.

Roles are defined by a controlled vocabulary, there are plenty of choices.

It is important to distinguish between authors and contributors, from Zenodo point of view only authors are used in citations.

You can refer to the CLEX Data Collection authorship policy to see when to use one or the other.

You have a few options to add your **Contact details** to the record

1) you add an ORCID, for at least one of the authors, and the email can be retrieved from their ORCID account. This is a good way to make sure the email stays current.

2) You can specify a contributor with the role of "Contact person" and in the "affiliation" box you can put an email. This will add a Contact person details under the authors. The email will be shown when you hover on top of the name, as shown in the image on the side.

3) Put directly the contact email in the **Additional notes** section, as:

   " Contact details: <email> "

In any case you should clarify in the **Additional notes** section how the contact details were shared, as shown below. This is important because, the ORCID link shown next to the authors' names is not very visible (circled in red above).
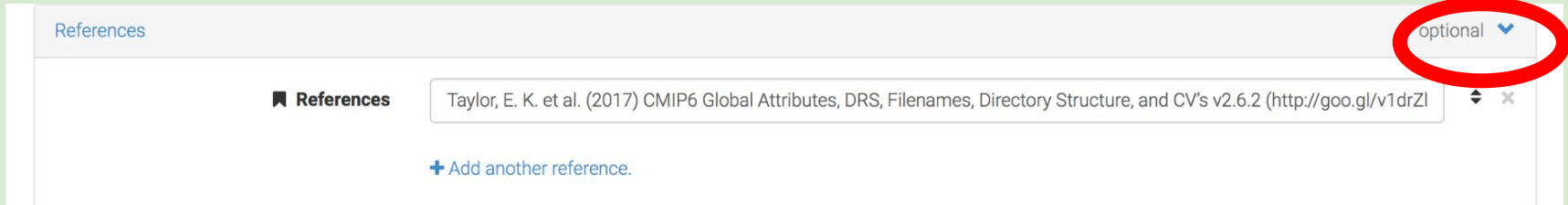
There are a few more optional fields, potentially the most relevant are **References** and **Subjects.**

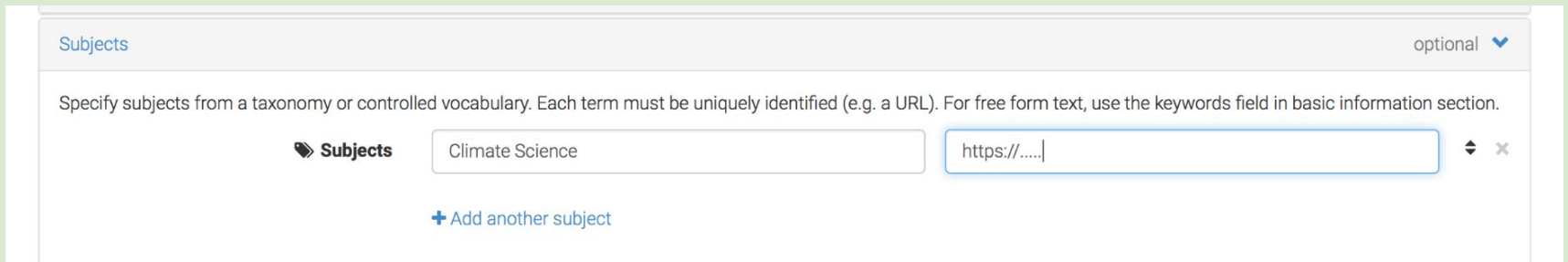Click on the arrow next to optional to expand this section and on + Add another reference to show the input box.

References can be used to add references as you would for a paper.

Subjects can be used to add terms from a Controlled Vocabulary. The first box is for the actual Term, the second for the identifier. Please note the identifier has to be a valid URL.
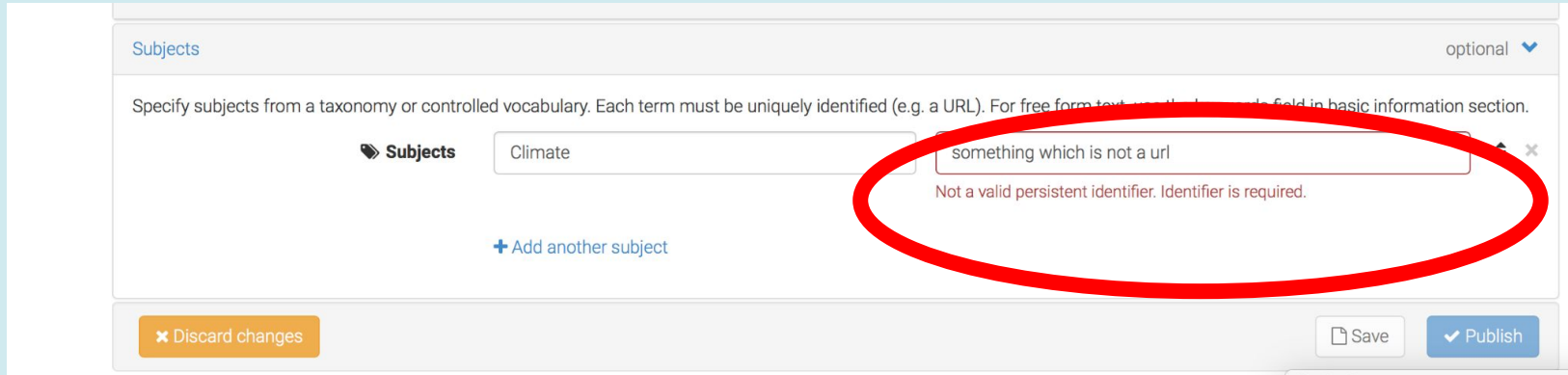
At the bottom of the web form you will find 3 buttons **Discard Changes**, **Save** and **Publish.**
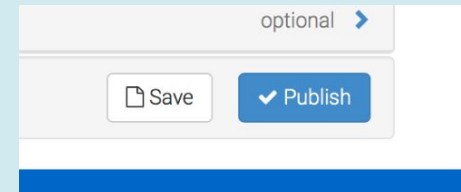

You can save your record at any time, if one of the fields have errors, as shown above, the record won't be saved. The error will be shown in red next to the relevant field. You have to fix the error and Save again.

The **Publish** button is disabled (transparent as in picture above) until:

- there are no errors

- at least 1 file is uploaded

- all the required sections and fields are filled

When the Publish button is enabled it will show as a solid blue color.

Once published your record will appear in Zenodo, a DOI will be assigned at this stage unless you reserved one or used an existing one. The publication date will appear at the top left of the page, as highlighted

You can edit your record at any time, you can change any metadata field including the title, the DOI and the publication date will remain the same.

Only if you change, add or remove any of the files, will Zenodo automatically create a new version and assign a new DOI.

This is the same as using the **New version** button. A new record will be created and a new DOI assigned. The form however will be pre-filled with the current version information by default.

Even if you choose the CLEX Data Collection Community in your form, your record will not immediately appear as part of the collection.

When you publish your record, Zenodo will send an email to the Collection curators informing them that a request has been submitted.

The record will be reviewed, and if anything needs changing or adding, one the curators will be in touch with you.

More information on the review process is available from the Collection policies page.

There you can also find the Guidelines for contributors and the Authorship policy, upon which these slides are based.

For feedback and further questions please contact

cws_help@nci.org.au