climate extremes

ARC centre of excellence

Publish your code

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

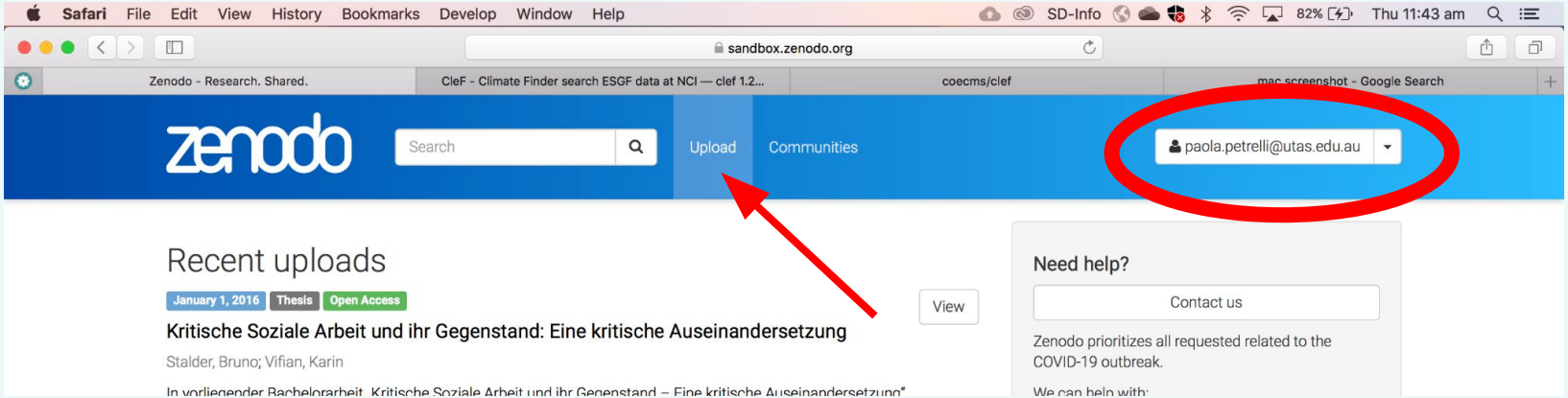
Table of contents

Table of contents (continued)

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

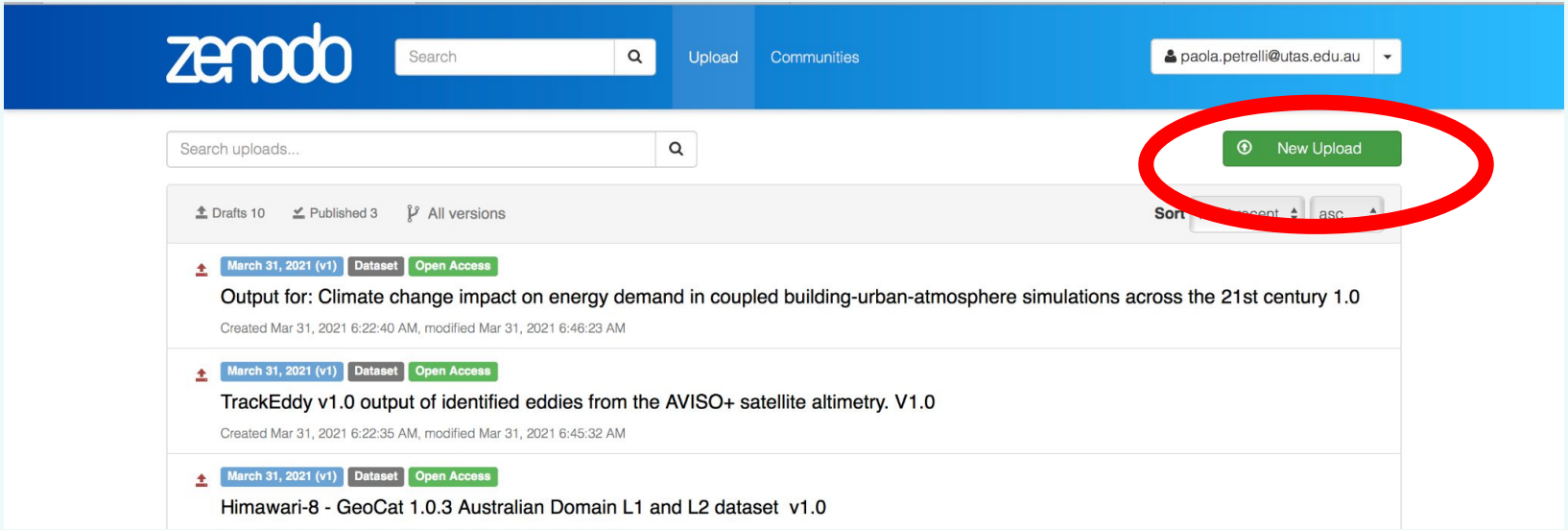
If your code is hosted on a public github repository then you can potentially link your account and repository directly to Zenodo. This has the advantage that for any new release of your code, Zenodo will automatically create a new version of your record. This process is covered in this blog from GitHub and we won't cover it here.

Even if you have created already a record for your code, this step by step guide will still help you make sure all the information required for your record to be accepted in the CLEX Code Collection is displayed correctly.



Login into your Zenodo account if you haven't already, your name should appear at the top on the right.

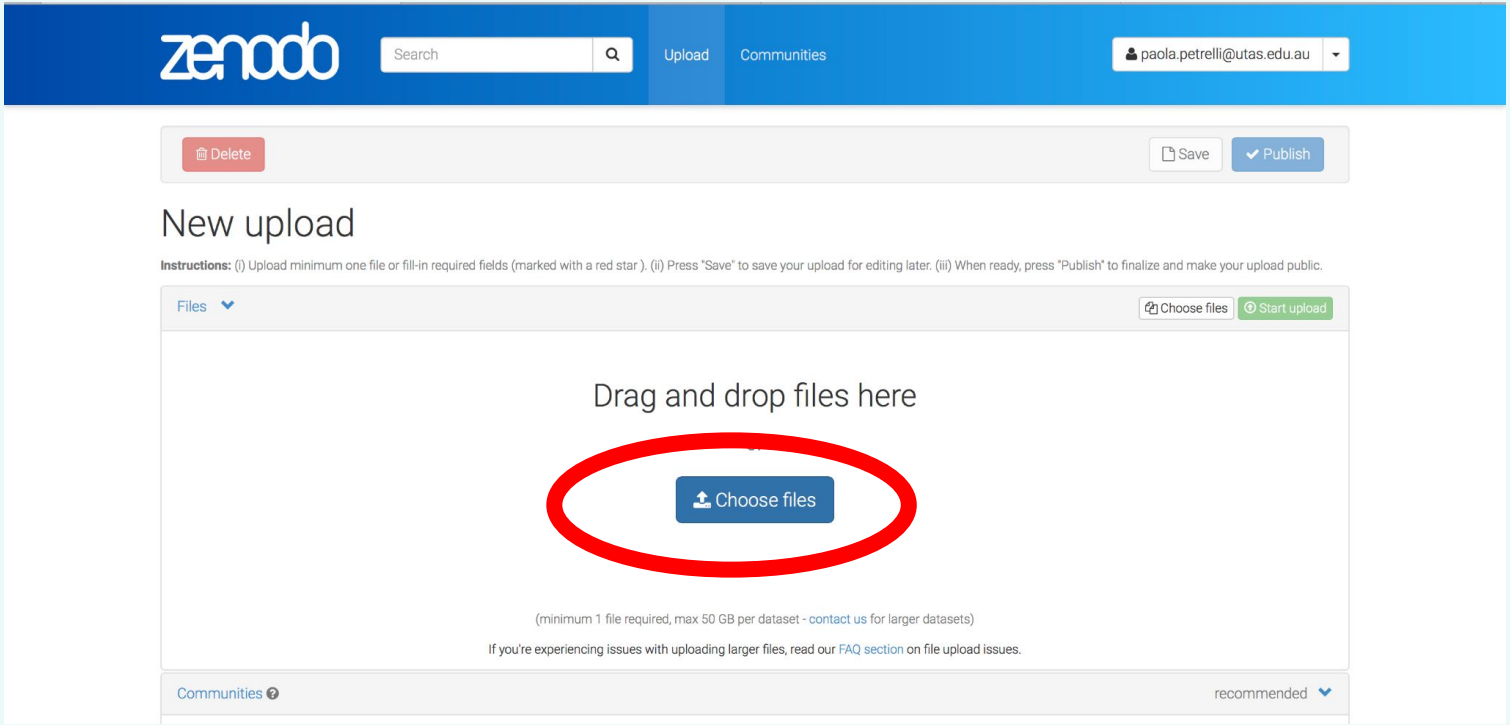
Click on the **upload** button



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

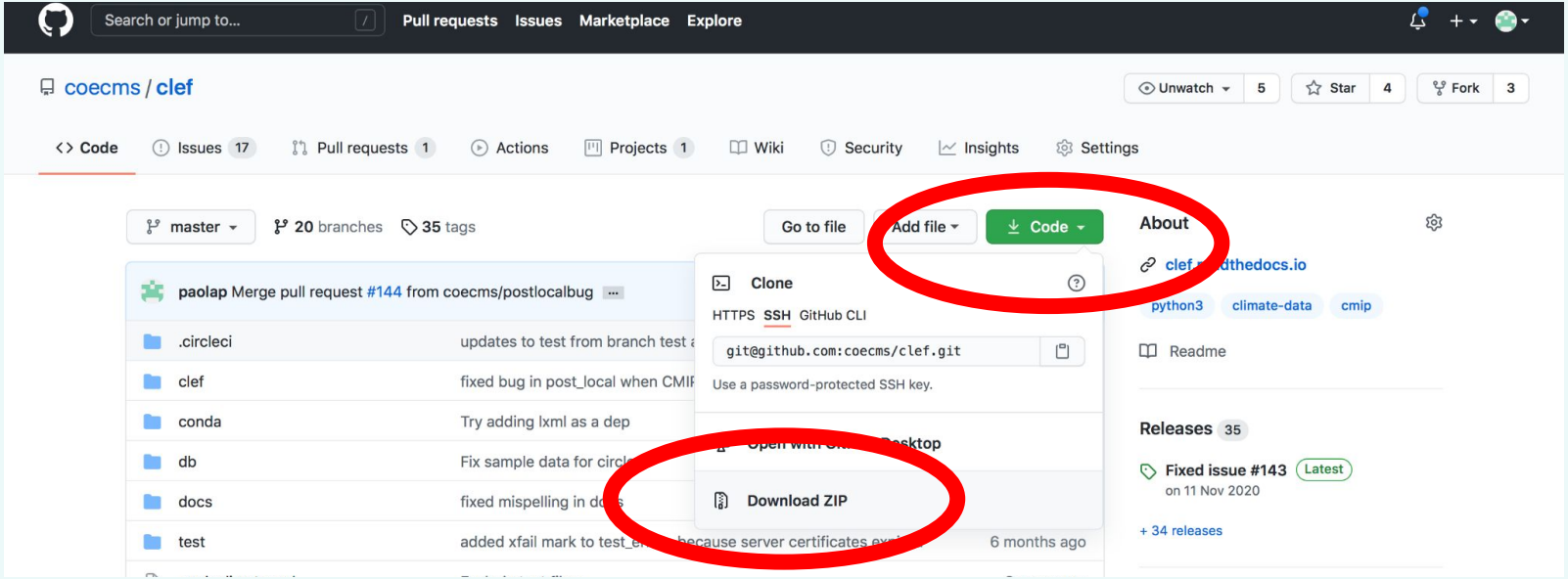
You can click on any of them to edit them.

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

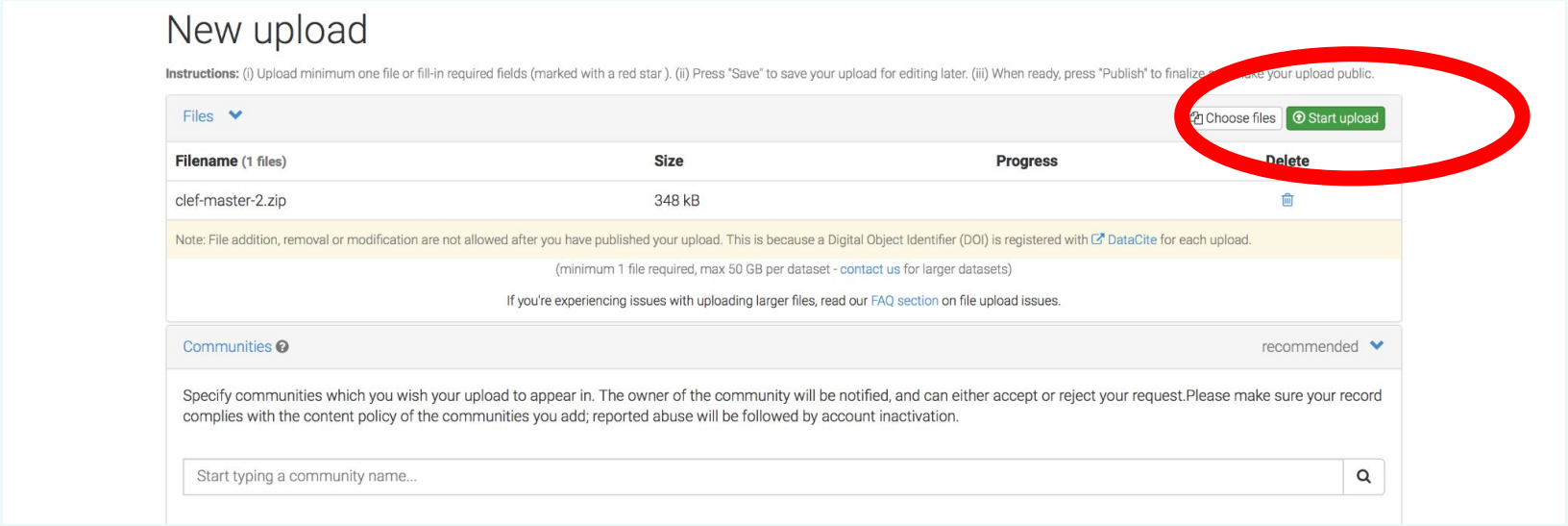


The **New upload** page shows the form to fill for your new record

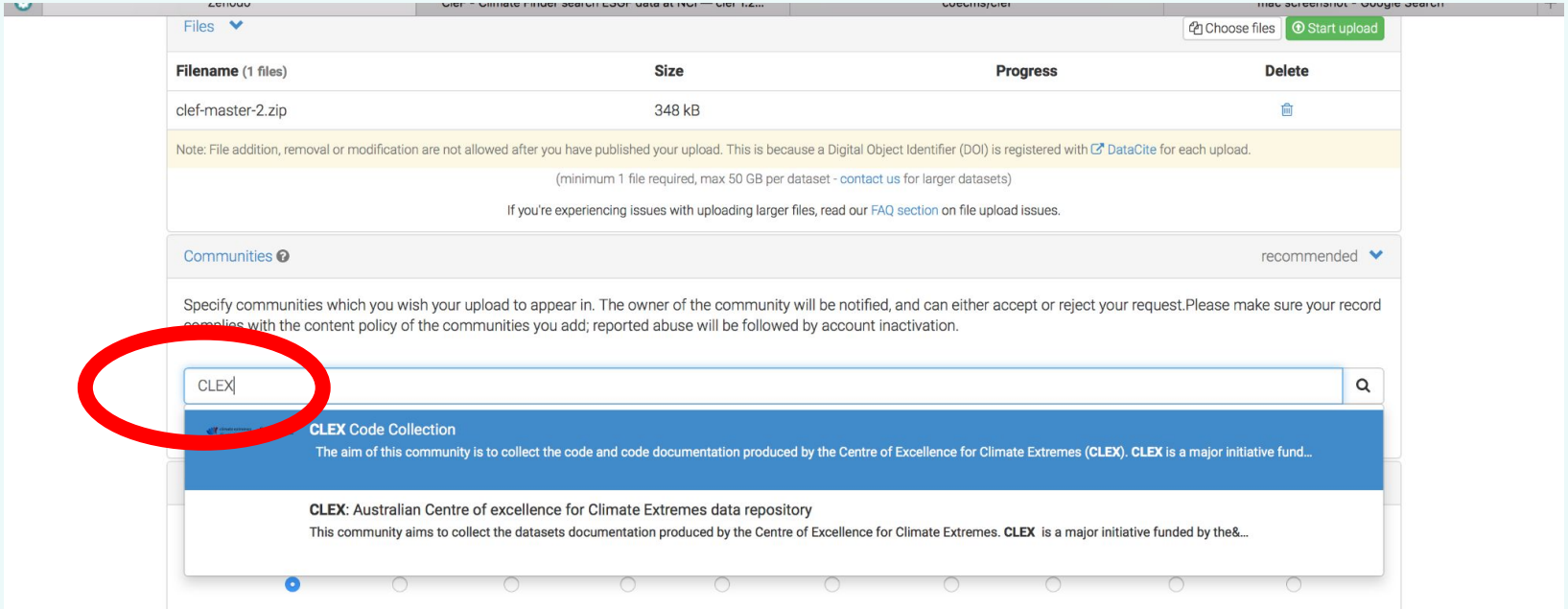
The first step is uploading your code. You can ‘drag and drop’ your files or open the upload window by clicking on the **Choose files** button.



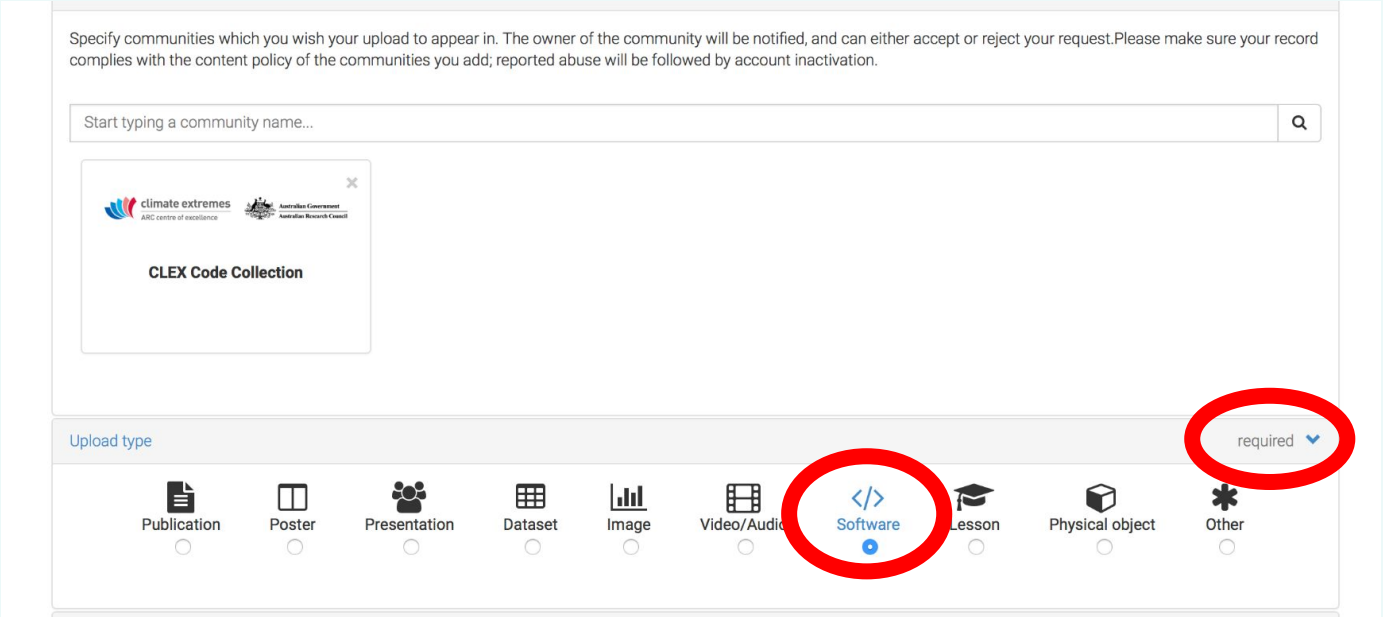
- If your code is on GitHub you can download your code as a zip file, and upload it as one file. The archive will be automatically expanded once the record is published.
- You need at least one file, if you have already published your code on another repository you can upload just a readme file here.
- Overall your files cannot be more than 50 GB in size, which should be plenty for code.



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 Code Collection, that we want to select in this case, and the CLEX data collection.

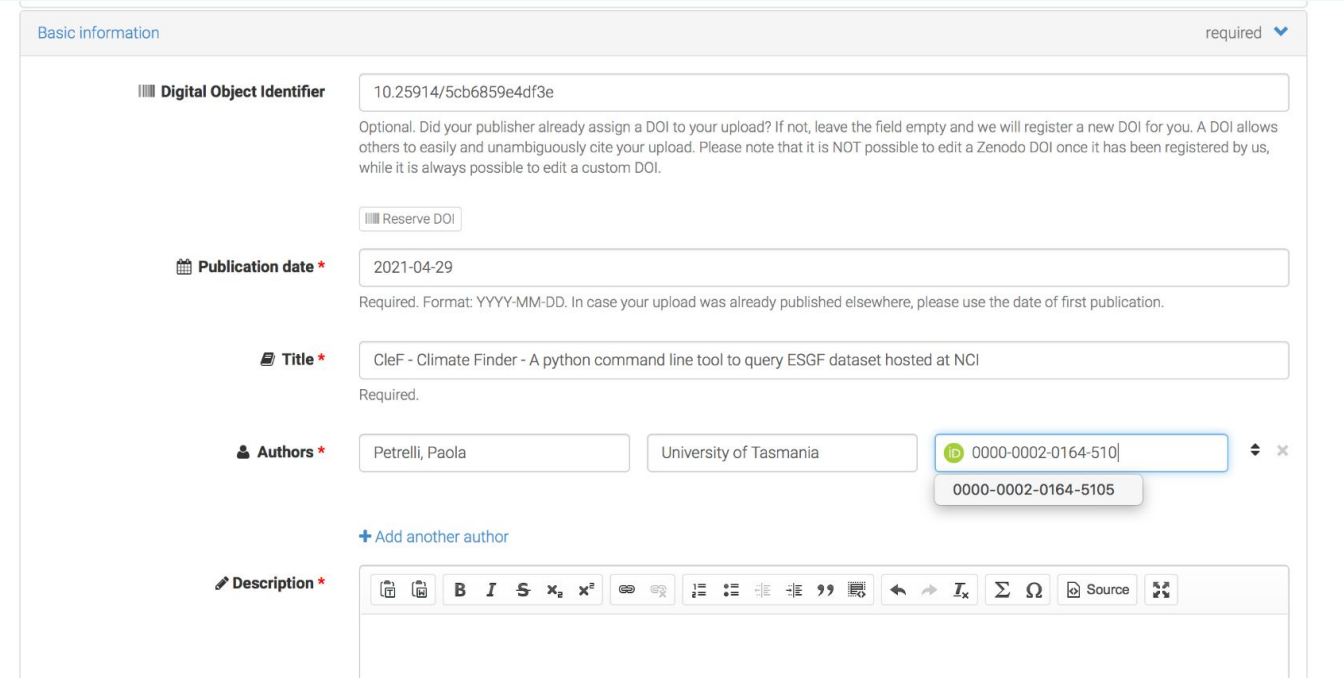


Pay attention to the **required** fields

Next required field is the **Upload_type**, with actual code its straightforward just select “Software”.

If the record is describing software 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 code. Most of the fields in this section except **Description** will be used to identify your code.

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

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

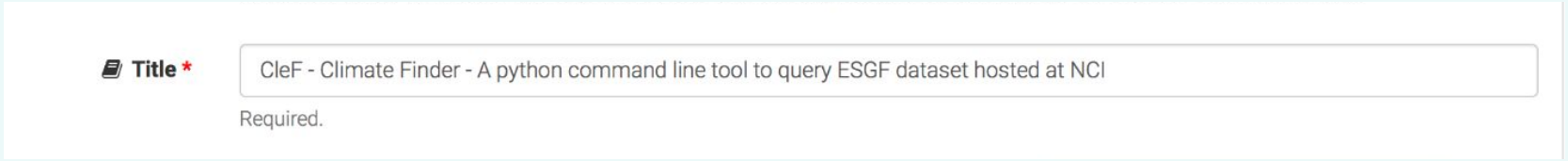
Otherwise a DOI will be added to the record at the moment of publication.

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 code that already has a DOI is a breach of our Collection policies and also of the Zenodo policies.



It is worth thinking carefully about a **Title** for your record, it is what will be used by anyone who reference your code. 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. All they will see initially is the title and the start of the description, until they actually click on one specific record.

Also, any word in the Title is used as a keyword by the search engine.

In the example above I tried to include:

- the code acronym and full name: CleF - Climate Finder
- that it is python based - A python
- that it is executed as a command line - command line tool
- what it does, including the data to which is applied: to query ESGF dataset
- The specific environment for which it was built: hosted at NCI

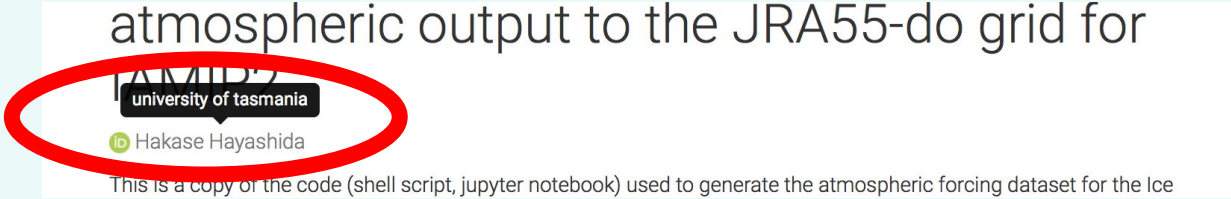
The **Authors** field includes 3 columns:

- 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



CleF searches the Earth System Grid Federation datasets stored at the Australian National Computational Infrastructure, both data published on the NCI ESGF node as well as files that are locally replicated from other ESGF nodes.

Currently it searches for the following datasets:

- CMIP5 in NCI projects: rr3, where NCI is the primary publisher and al33 for replicas
- CMIP6 in NCI projects: oi10 for replicas, fs38 where NCI is the primary publisher
- CORDEX (ESGF only) in NCI projects: rr3, where NCI is the primary publisher and al33 for replicas

The search returns both the path of data that is already available at NCI as well as information on data that is on external ESGF nodes but not yet available locally.

The search works like the ESGF search website, e.g. https://esgf.nci.org.au/search/esgf_nci. Results can be filtered by using flags matching the ESGF search facets.

Four optional flags are available for the **cmip5** and **cmip6** commands to change the output or submit a data request:

> - *`clef --remote <dataset>` returns all the ESGF for the dataset matching the constraints, it is the equivalent of doing a search online on an ESGF node*
> - *`clef --local <dataset>` finds local files accessing directly the NCI's clef.nci.org.au database so it will also return older versions or datasets that might be temporarily offline.*
> - *`clef --missing <dataset>` finds files on ESGF that haven't been downloaded to NCI*
> - *`clef --request <dataset>` create and pass to NCI a request to download the missing files*

CleF is a python module, it can also be imported directly in your own python codes. It was designed specifically to work on the NCI servers and it is already installed and available in a shared conda environment on these.

This version (1.3.0) adds the *cordex* sub-command to query CORDEX data published on the ESGF.

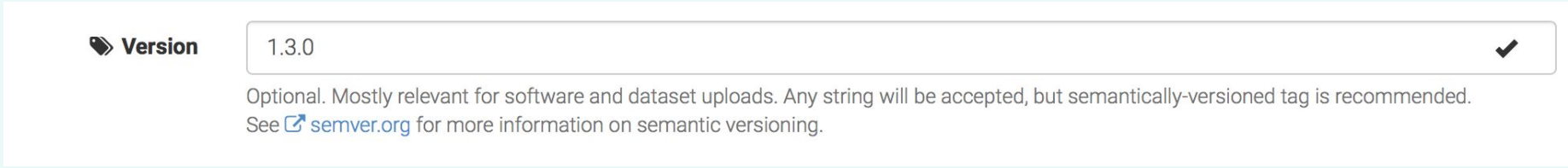
For more information, tutorial and the full list of options check the Read the Docs documentation.

If you find a bug or have a comment on the code, please create an issue on GitHub.

Description should include:

- code scope and context
- software needed to run the code
- how to run or compile
- any installation on a specific server
- if this is a new version how it is different from previous ones
- preferred way to handle bugs and feedback
- any documentation and/or training resources

Links to the code repository and other related material should be added in related information. However, it is recommended to also mention them here.



The **Version** field is really important to identify your code, as well as being required. Even if you are not planning new releases it's fairly common to have updates with code.

Consider following the Semantic Versioning convention, this scheme uses a 3 part version number,

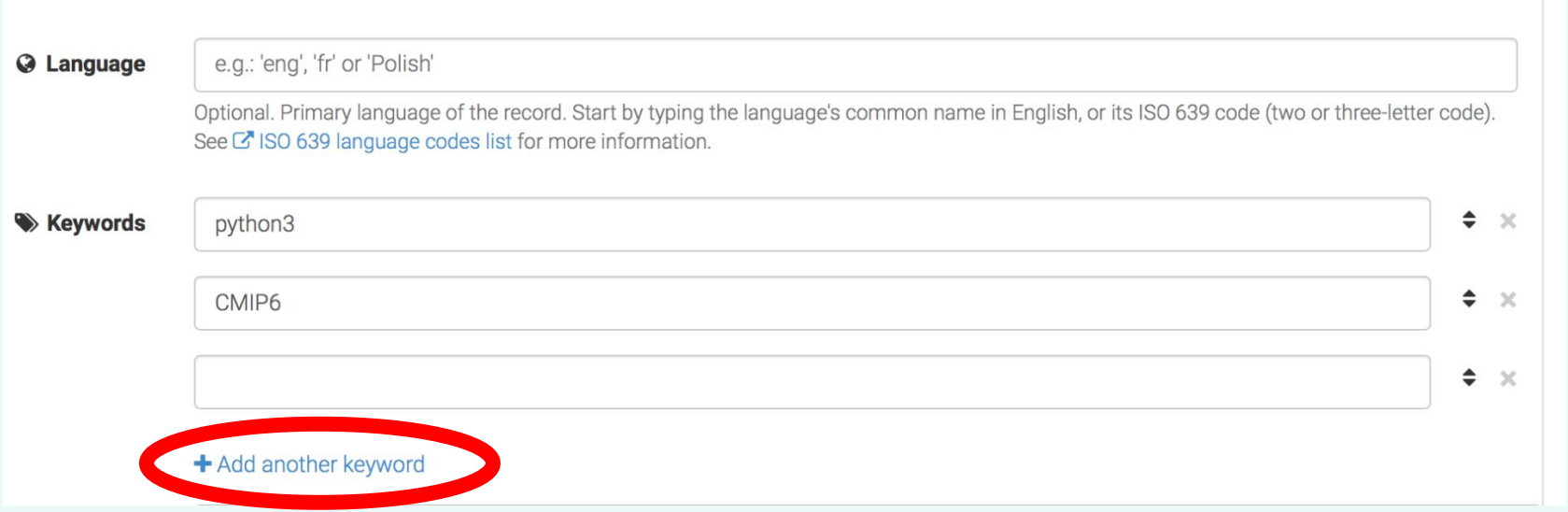
MAJOR.MINOR.PATCH as in v1.3.0

1. MAJOR changes when the updates will break previous behaviour, use "0" to indicate a code still under development
2. MINOR changes when adding new functionality in a backwards compatible manner
3. PATCH number when you make backwards compatible bug fixes.

In the example above we went from 1.2.3 to 1.3.0 as we added a new functionality without breaking existing behaviour.

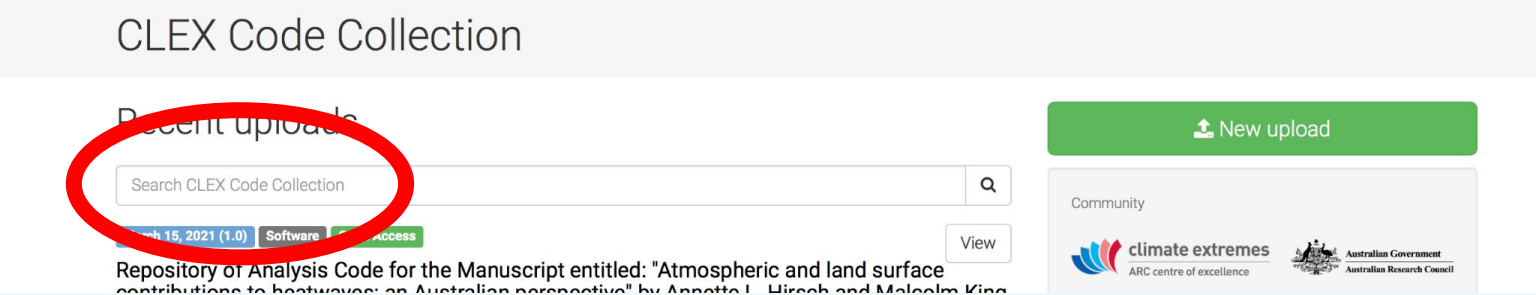
If you are using GitHub, produce a release before publishing.

Most of all it is important to be consistent and versions should always progress from lower to higher.



Language indicates the language used to write the record, not the code language, you can use eng for English

Keywords can be added by clicking on the **+Add another keyword** button. We have a controlled vocabulary for **Keywords**, this is to encourage using the same terms across the collection. (continue on next slide)

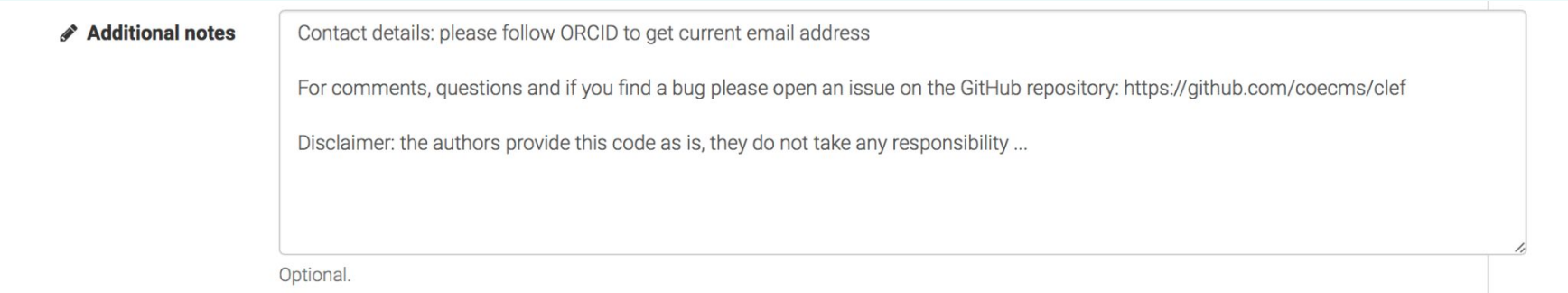


If you execute a query from the main CCC page then the query will include only the collection records, and we want the keywords in this space to be consistent.

You should always include the Programming language, other categories to use wherever is applicable are:

- data type the code apply to (ie. observations, model outputs). If the code applies to a specific dataset you should use the official acronym
- discipline
- geographic domain
- time frequency
- code purpose (ie. model, post-processing.)

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.



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

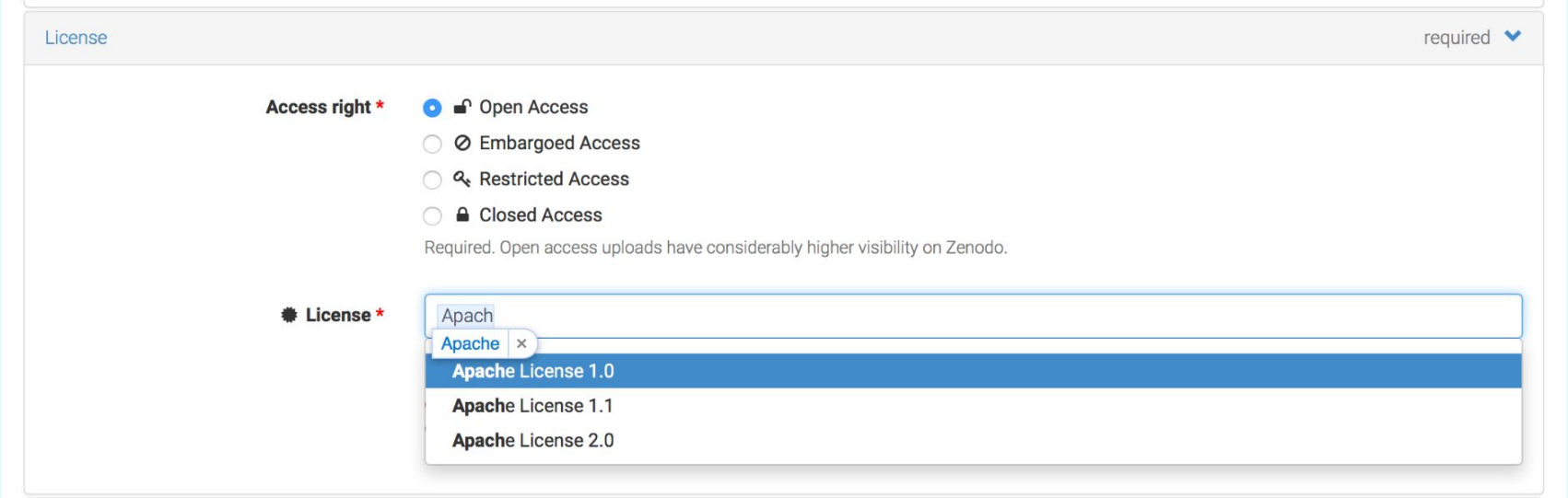
In this example it highlights the way the author prefers to be contacted in case of issues and the contact details.

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 at least include here the contact details or how to retrieve them, for example from the ORCID, as illustrated above.

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

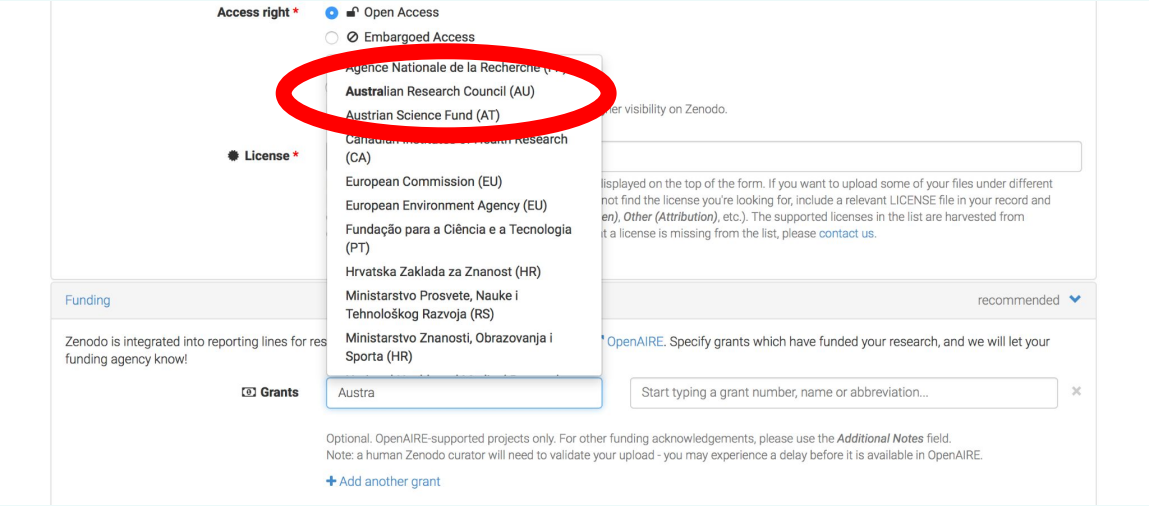
We are covering all the options to handle them in a separate slide.



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 Apache 2.0, or you can choose another software license. Creative Common licenses are common (and the default in Zenodo) but they are not suited to software. Start typing the name of the license to see more options.



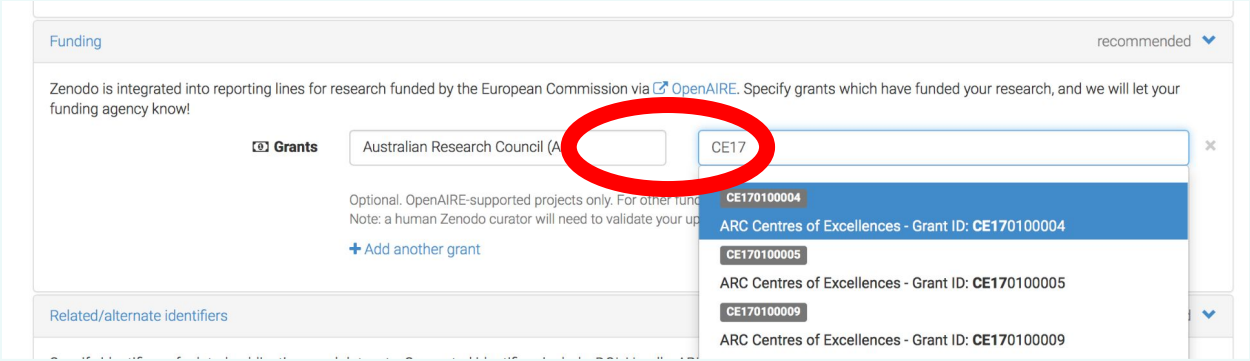
If your code was funded by a grant you can list the funder and the grant code in **Grants.**

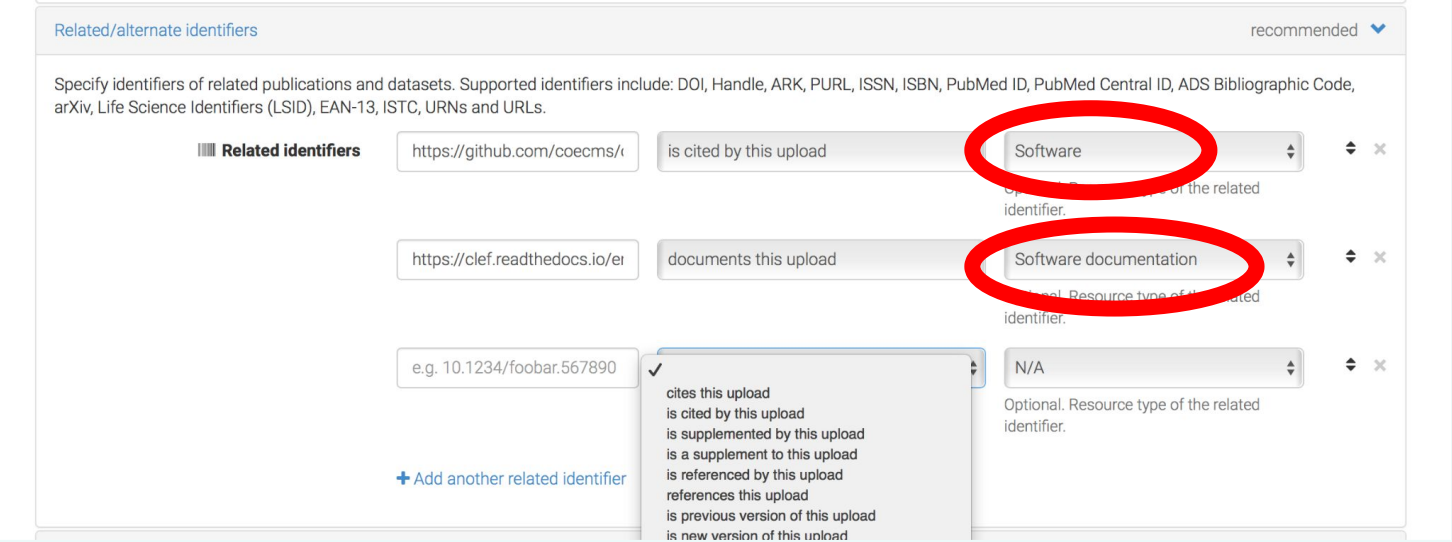
There are funders listed from all over the world. In this case, as I want to list the CLEX grant, I choose the ARC as funder.

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)



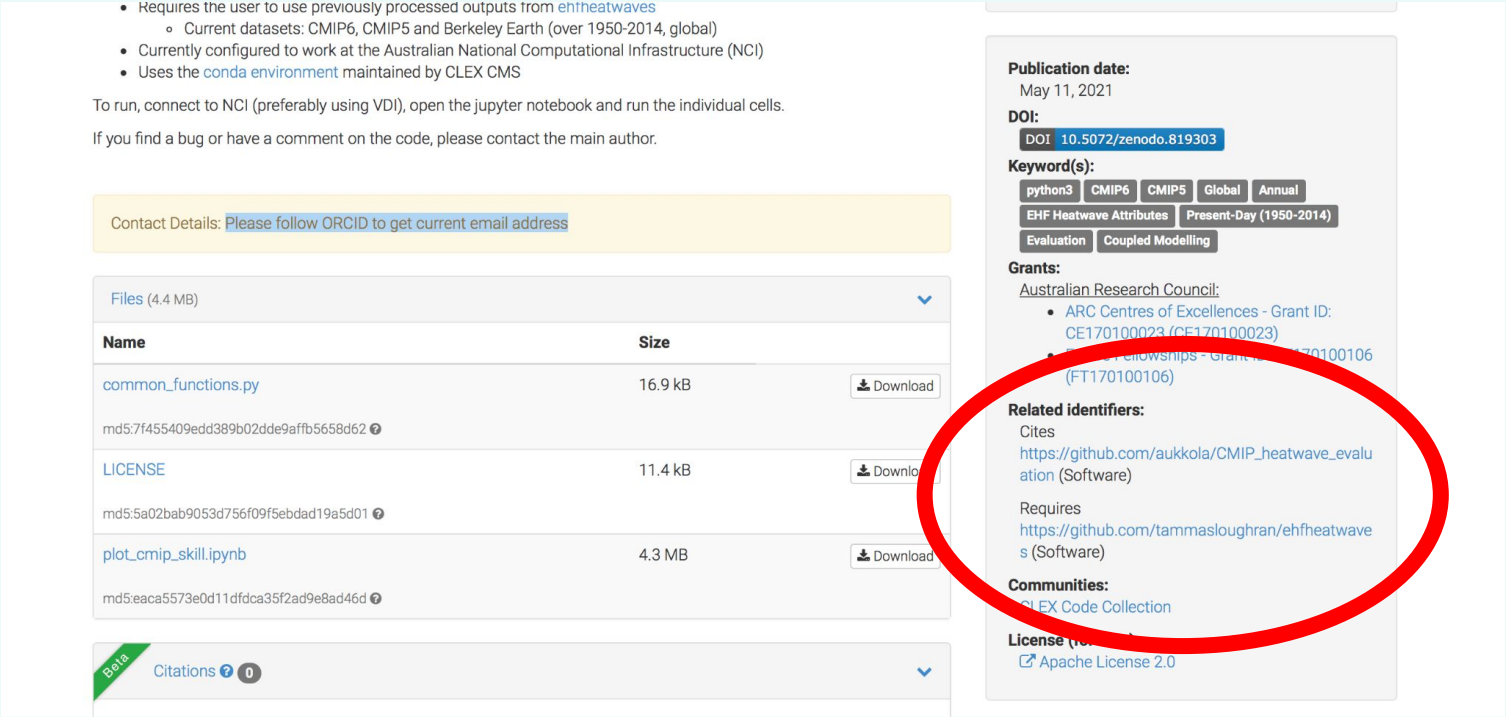


Related/alternate identifiers are used to add links to related resources. There are 3 columns:

- the first is to list the **resource url**
- the second is for the r**elationship** 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,** examples of relevant types are “Software” and “Software documentation” as visible above

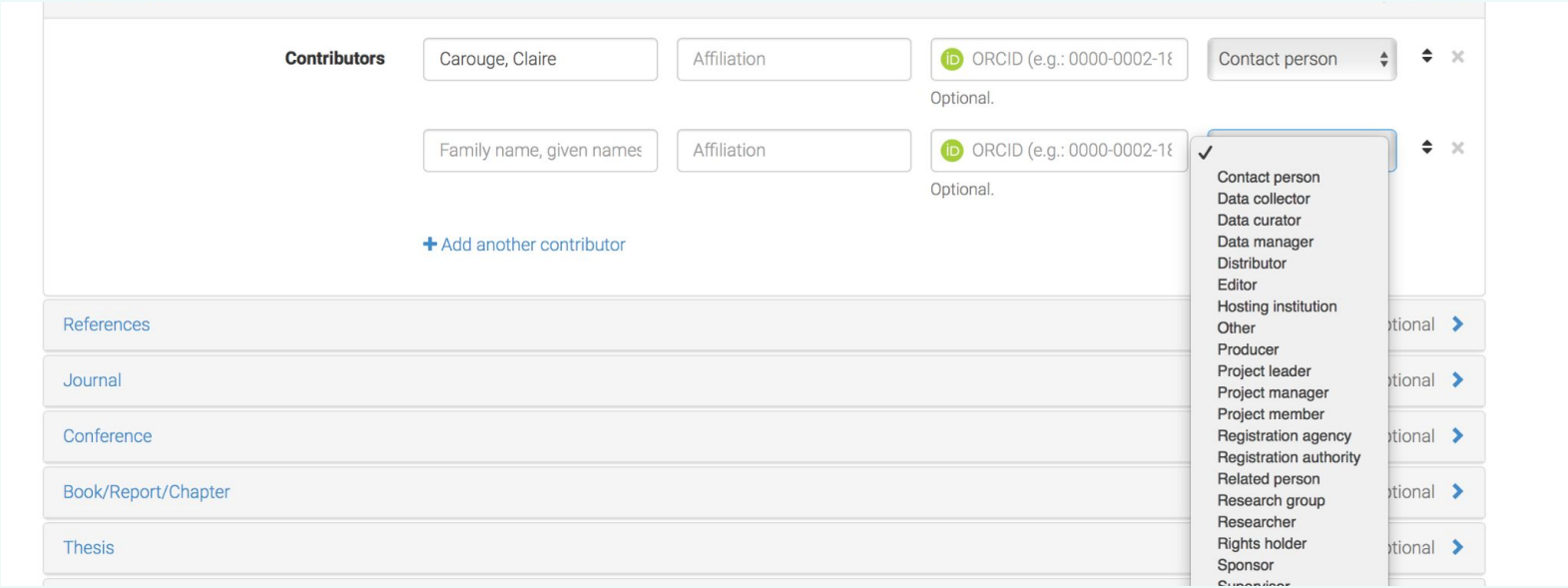


Related identifiers will appear in a prominent position on a right side panel, as shown in the image. They should be used wherever applicable, examples with a suggested relationship are shown in the next slide.

It is possible to update the record at any time to add additional Related identifiers as those might be created after the publication of the code.

Resource	Suggested relationship
code repository	is cited by this upload
Online documentation or training	documents this upload
official metadata record, if the code is already published in another repository	is alternate identifier of this upload
technical reports, paper describing the code	describes this upload documents this upload
published data produced with the code	has this upload as its source

Resource	Suggested relationship
Another code or data required by the code	is required by this upload
A code used as base for this	is the source this upload is derived from
A previous/new version of the same code (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
Paper referencing/citing the code	references this upload cites this upload



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.

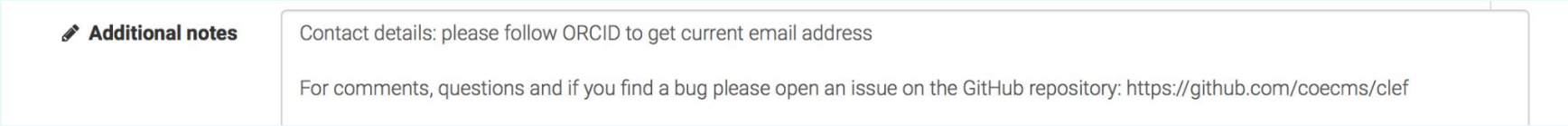
You can refer to the CLEX Code Collection authorship policy to see when to use one or the other.

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:

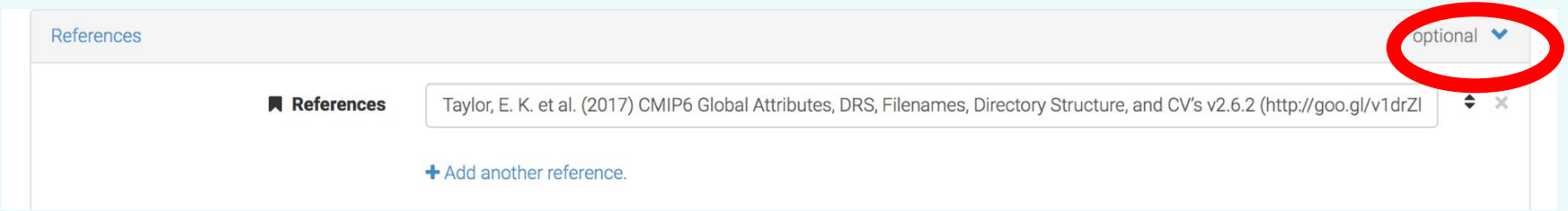
   “ Contact details: <email> “

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).



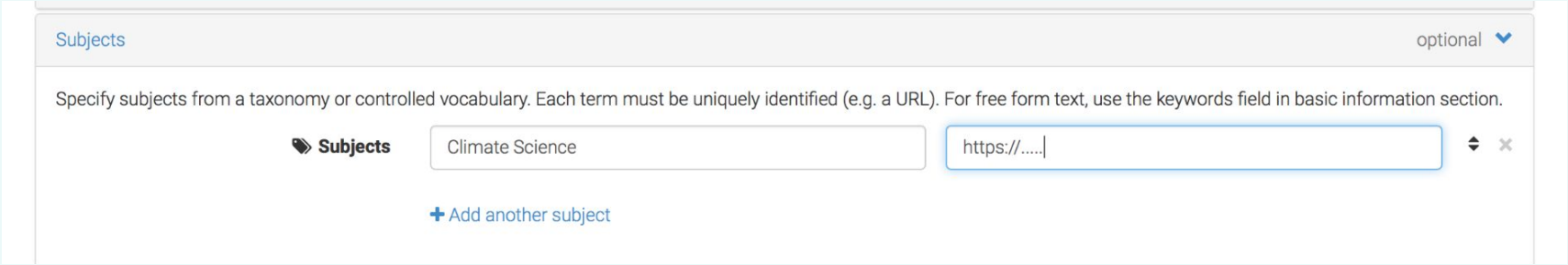
There are a few more optional fields, potentially the most relevant are **References** and **Subjects.**

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

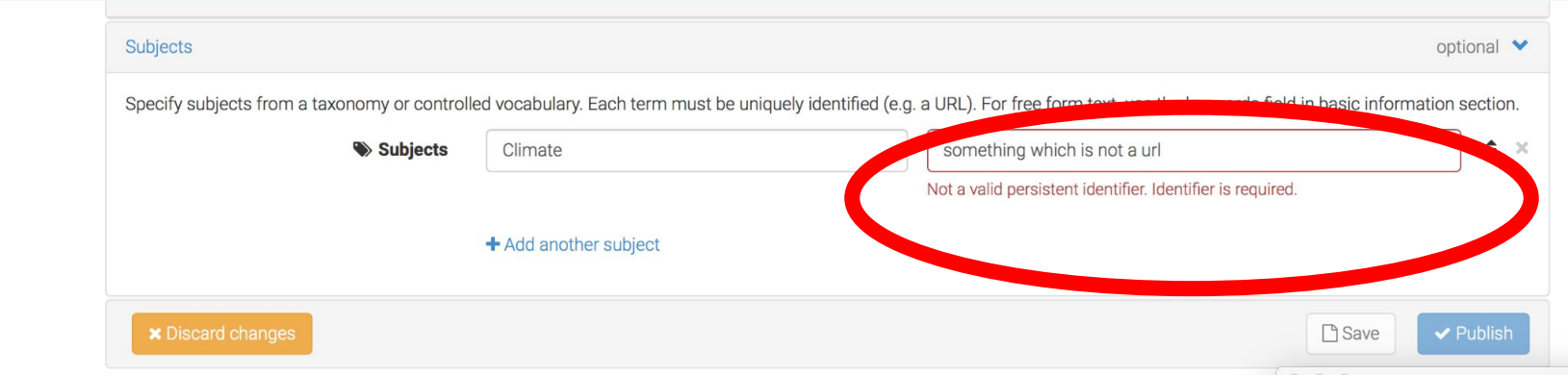


References can be used to add references as you would for a paper.

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.



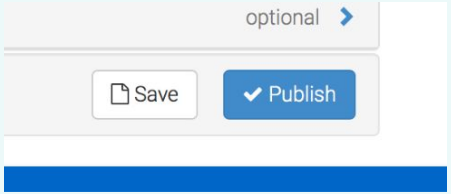
At the bottom of the web form you will find 3 buttons **Discard Changes**, **Save** and **Publish.**



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.

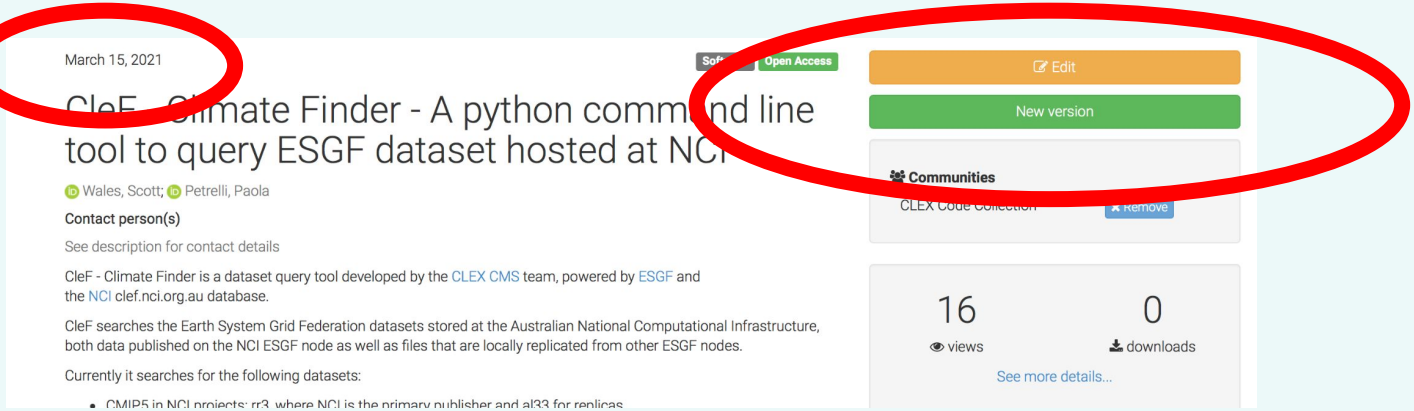
The **Publish** button is disabled (transparent as in picture above) until:

- 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 Code 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 of 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.

climate extremes

ARC centre of excellence

For feedback and further questions please contact

clexcodecollection@gmail.com