Trying out Zenodo data repository

Ghent University Data Stewards

License: [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

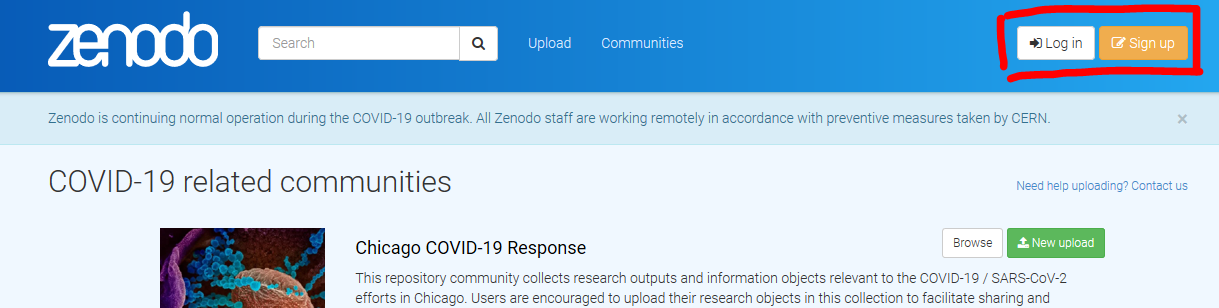
# Outline

The goal of this assignment is to get familiarized with data repositories, the process of uploading a file(set) and providing metadata to describe the content of your research output. We will create a test record using the training environment of the repository (Zenodo Sandbox). Your test dataset will be published in the sandbox environment of Zenodo, which is cleaned from time to time.

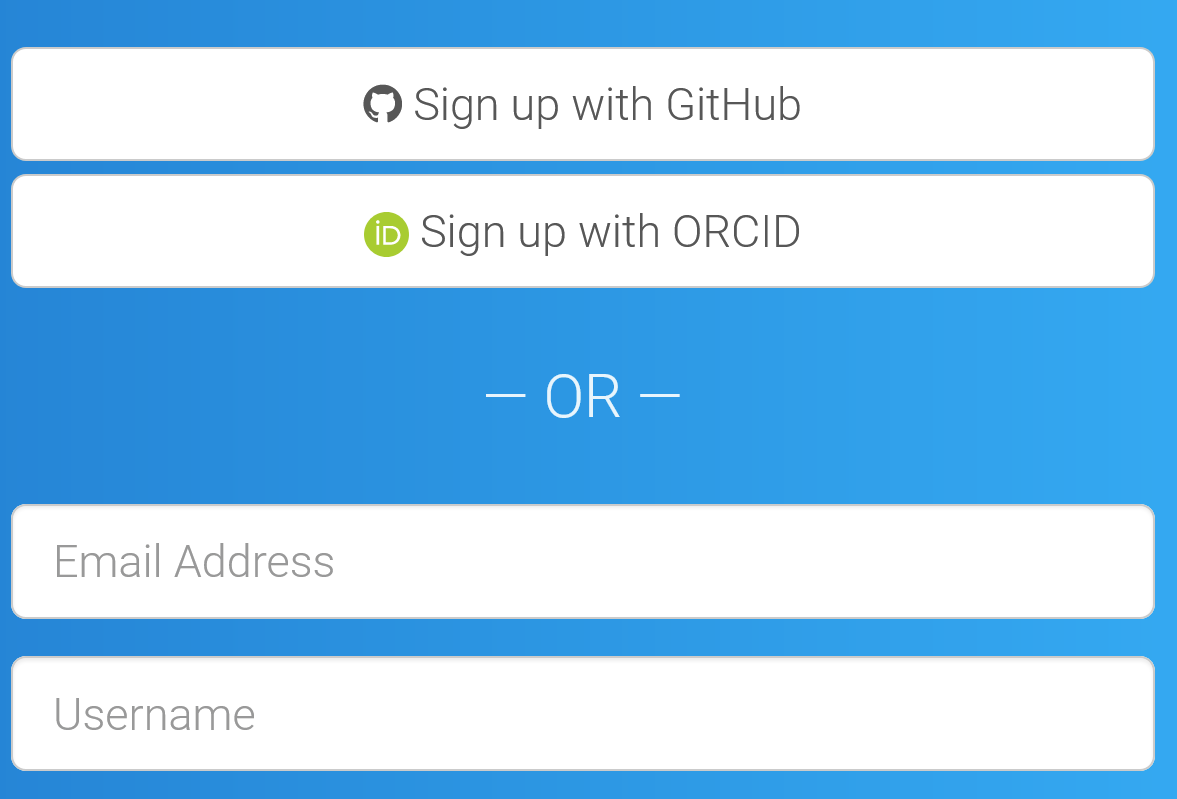
**Do not use actual data or confidential information for this exercise!**

# Preparation – Register and login

1. Navigate to Zenodo Sandbox: <https://sandbox.zenodo.org/>. If you already have an account, log in. If not, click on “Sign up”, in the upper right corner of the page.



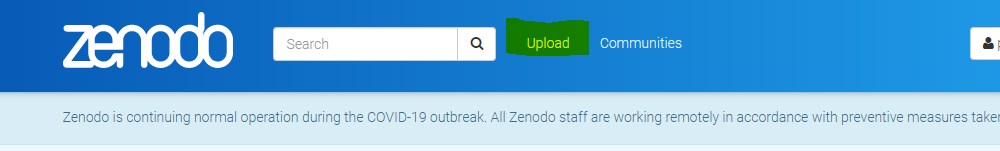
1. Select Sign up with Github, using your account on github.com; or with ORCID, but note this is sandbox.orcid.org - not your regular ORCID; or sign up to sandbox.zenodo.org itself.



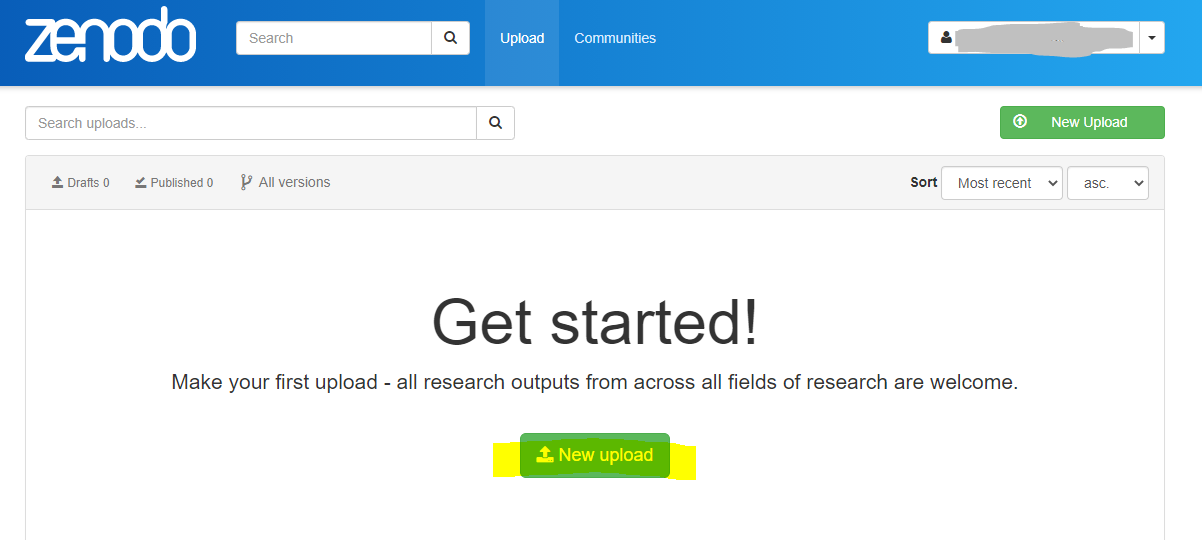
1. Log in to Zenodo once your account is created and confirmed.

# Exercise instructions

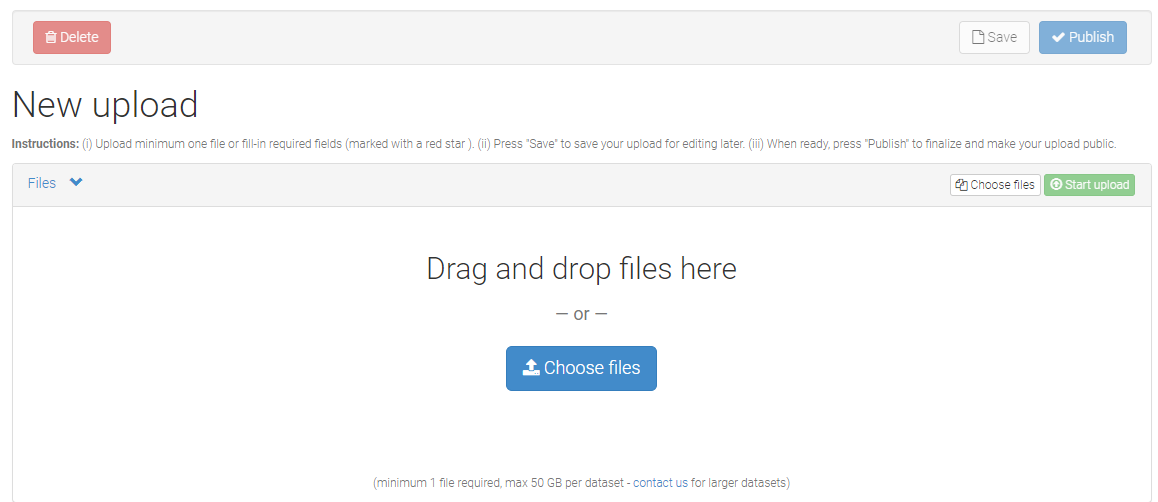
1. Navigate to Zenodo Sandbox and make sure to be logged in: <https://sandbox.zenodo.org/>
2. Click on upload:



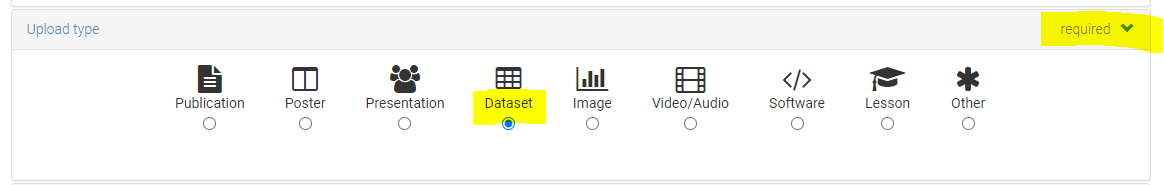
1. You will now be guided through a series of steps to upload a file(set) and provide metadata. Click on New upload.



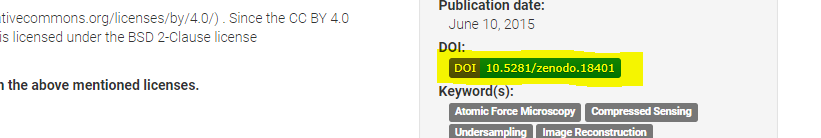
1. The first step is to upload a file or a set of files. In real life, you should upload your data files(s), together with any relevant documentation necessary to understand and reuse your data (e.g. variable dictionary, interview template, etc).



1. Do not forget to click “Start upload” after selecting or dropping the files.
2. By selecting a “Community”, your data will be added to a community, providing more visibility in search results. For example, type “ugent” and select the “Ghent University” community. (optional)
3. Move to “Upload **type**”. Select a type, e.g. dataset. On the right side of each step (or metadata element), you can see whether this is a recommended or obligatory field.



1. Leave the Digital Object Identifier field empty and keep publication date as it is.
2. Provide a **title** for your dataset. In real life, you should choose a good title for your dataset. The title is one of the most important fields for data discovery, and it should be descriptive, meaningful but also concise. It should allow users to know what the data is about: what, where, when, how… These are some examples found in other thematic data repositories:
   1. “[Chemical and mineral compositions of sediments from ODP Site 127-797](https://doi.pangaea.de/10.1594/PANGAEA.726855)”.
   2. “[Well-being, life conditions and quality of life in Denmark during the crisis, 1982-1994](https://datacatalogue.cessda.eu/detail?q=%22DNA__oai:kuha.sa.dk:11521%22)”
   3. “[Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index](https://doi.org/10.4225/13/50BBFD7E6727A)”
   4. “[Finnish National Election Study 2011: Telephone Interviews among Finnish-speaking Voters](http://urn.fi/urn:nbn:fi:fsd:T-FSD2673)”
3. Add yourself as an **author**. Make sure to add your ORCID.
4. Provide a **description** (abstract). You don’t need to provide an elaborate abstract for the moment. However, keep in mind that, together with the title, the abstract is one of the fields that a potential user will check in order to decide whether the dataset is useful for him/her or not. Have a look at the checklist in [this page](https://www.aalto.fi/en/services/describing-datasets-in-the-data-repositories) to get a feeling of what a good abstract should include.
5. Choose a language and add some **keywords**. You can use free text for the purpose of this exercise. In real life (and as much as possible), use controlled terms, classification codes or thesauri in your discipline to provide a list of keywords. Some examples of thesauri and controlled vocabularies and resources in which you can find terminology standards:
   1. [IEEE Thesaurus (Institute of Electrical and Electronics Engineers)](http://www.ieee.org/publications_standards/publications/services/thesaurus2.html)
   2. [EcoPortal](http://ecoportal.lifewatchitaly.eu/)
   3. [Food Ontology](https://foodon.org/)
   4. [Agronomy Ontology](https://bigdata.cgiar.org/resources/agronomy-ontology/)
   5. [AGROVOC](http://aims.fao.org/vest-registry/vocabularies/agrovoc)
   6. [GeoNames](http://www.geonames.org/)
   7. [Research Organisation Registry](https://ror.org/search)
   8. [Ontology Lookup Service](https://www.ebi.ac.uk/ols/index)
   9. [FAIRsharing](https://fairsharing.org/standards/)
6. Choose between one of the **access rights** options, as well as a **license** for reuse.
7. Explore the remaining metadata elements that can be provided. As you can see, it is possible to provide detailed contextual information:
   1. Reference to the grant under which the research data was funded.
   2. Other contributors and their roles (e.g. data collector, supervisor…).
   3. References to related publications or related datasets.
   4. Etc
8. Once you are satisfied, publish your dataset into the Zenodo Sandbox.
9. FYI, you can find a dummy DOI on the right panel at the page of your record.



Alternative: B2SHARE data repository

# Preparation – Register and login

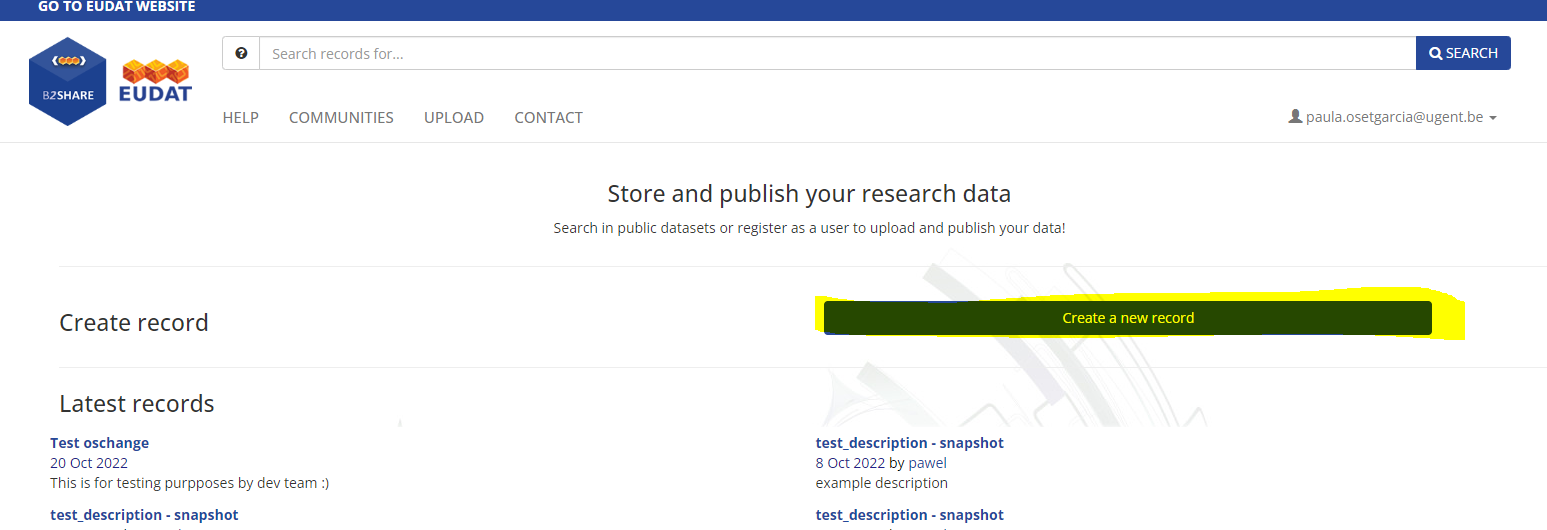
1. Navigate to the training environment of B2SHARE: <https://trng-b2share.eudat.eu/>. If you already have an account, log in. If not, click on “Login”, in the upper right corner of the page.



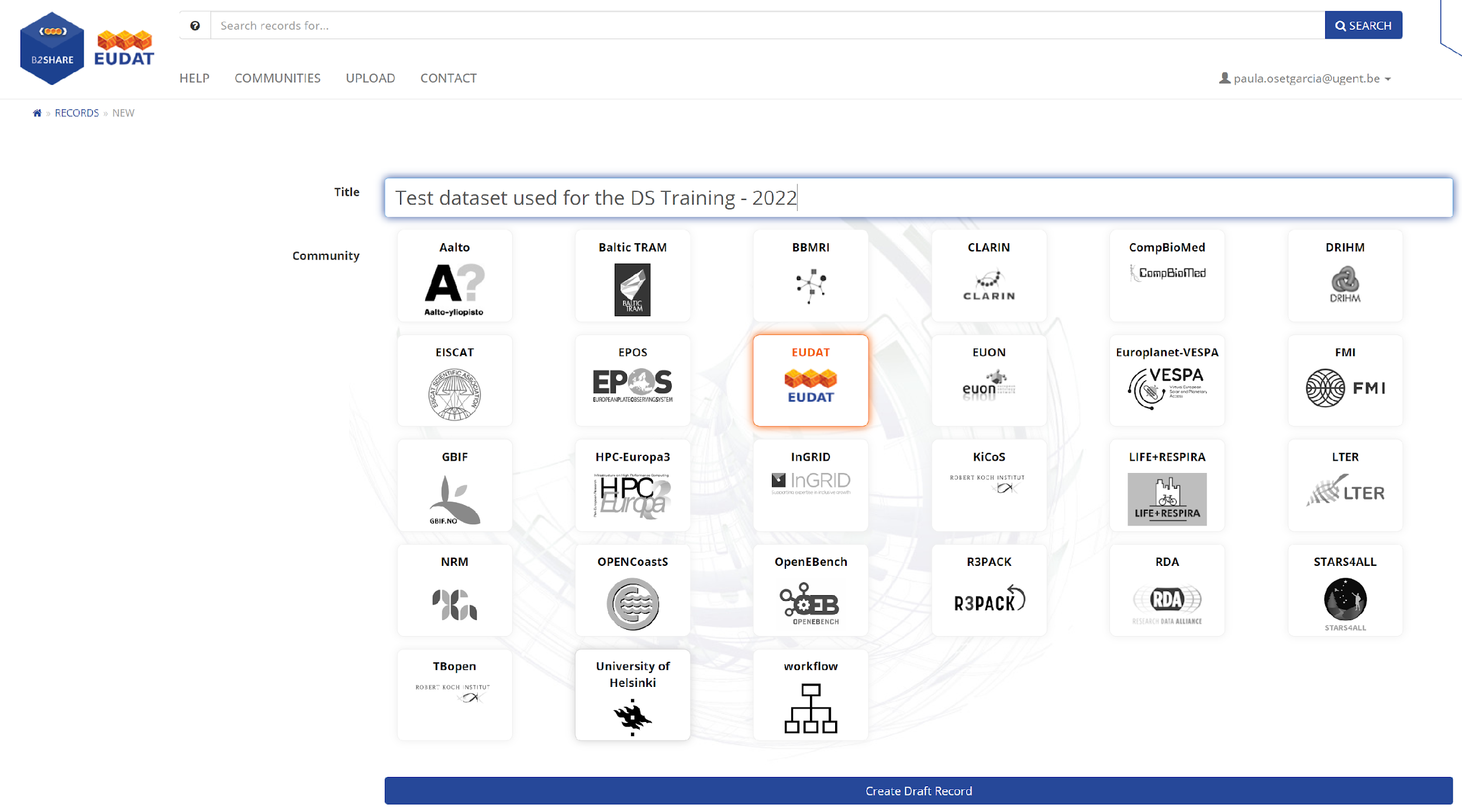
1. Select Sign up with **ORCID**, so that your B2SHARE publications are linked with your ORCID, contributing to your scholarly records. The application will ask you whether you want to grant permission to B2SHARE from your ORCID record. Click on allow.

# Exercise instructions

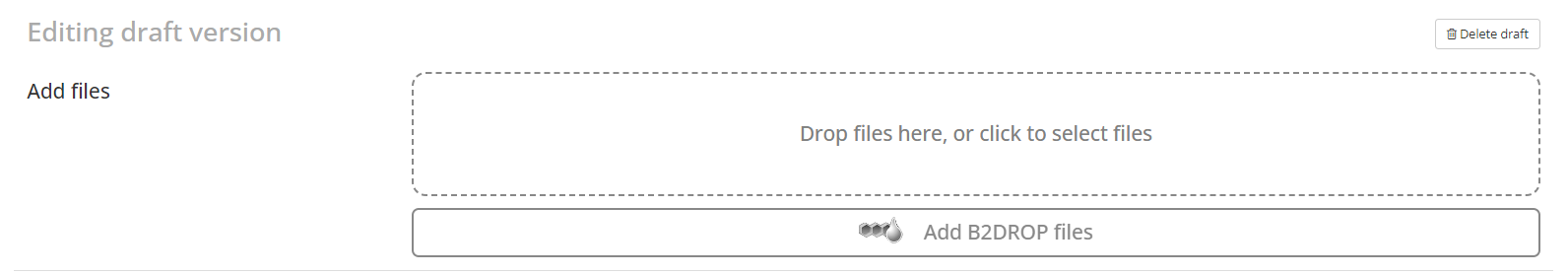
1. Once logged in, click on the Create a new record button.



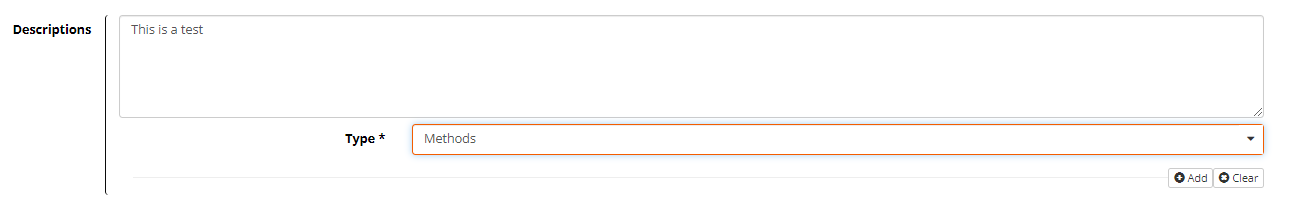
1. In the following step, B2SHARE will ask you to provide a title for your dataset and to choose from a “Community”. B2SHARE is a generic repository and, therefore, uses generic metadata. However, the generic metadata profile can be extended with discipline-specific fields by choosing a community. For example, BBMRI (biomedical research) provides additional metadata fields such as age interval, sex, material type; whereas EPOS adds fields related to the geographical coverage of the data. If you do want to stick with generic metadata, **choose the EUDAT community**. Then click on “Create draft record” at the bottom of the community tiles.



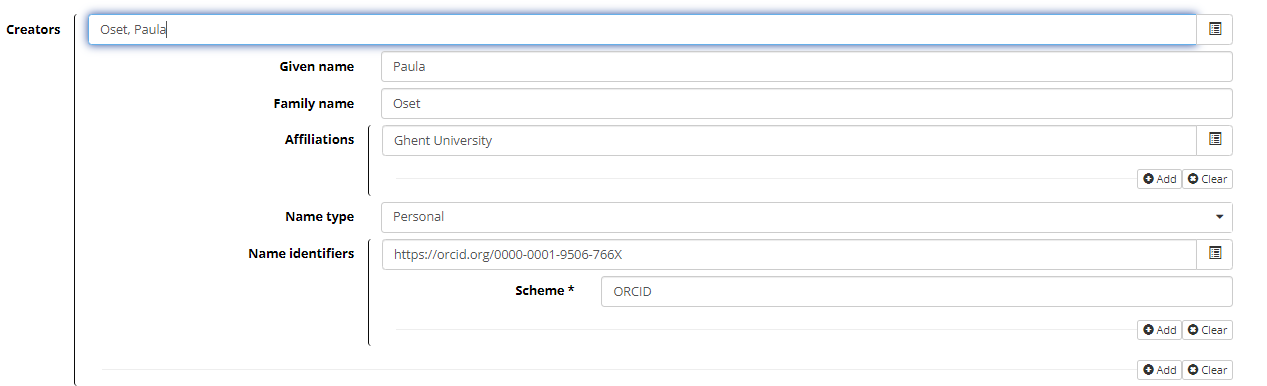
1. You will now be guided through a series of steps to upload a file(set) and provide metadata.
2. The first step is to upload a file or a set of files. In real life, you should upload your data files(s), together with any relevant documentation necessary to understand and reuse your data (e.g. variable dictionary, interview template, etc).



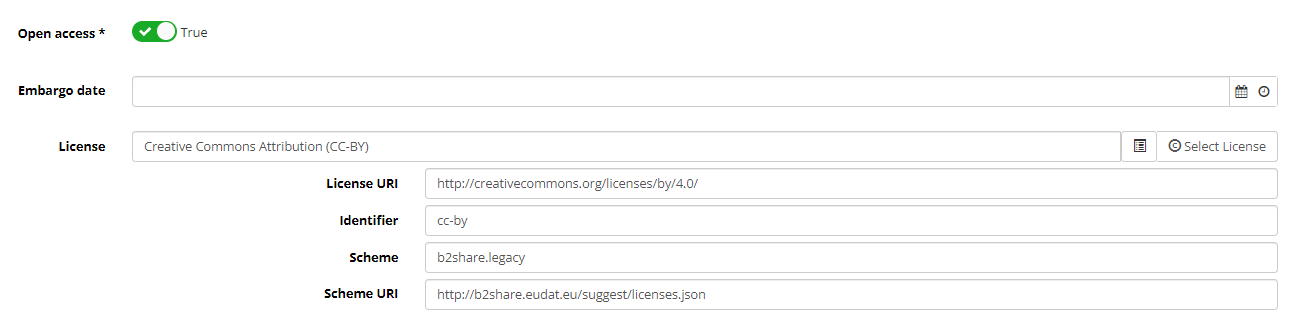
1. At this stage, you can still modify the **title** for your dataset. In real life, you should choose a good title for your dataset. The title is one of the most important fields for data discovery, and it should be descriptive, meaningful but also concise. It should allow users to know what the data is about: what, where, when, how… These are some examples found in other thematic data repositories:
   1. “[Chemical and mineral compositions of sediments from ODP Site 127-797](https://doi.pangaea.de/10.1594/PANGAEA.726855)”.
   2. “[Well-being, life conditions and quality of life in Denmark during the crisis, 1982-1994](https://datacatalogue.cessda.eu/detail?q=%22DNA__oai:kuha.sa.dk:11521%22)”
   3. “[Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index](https://doi.org/10.4225/13/50BBFD7E6727A)”
   4. “[Finnish National Election Study 2011: Telephone Interviews among Finnish-speaking Voters](http://urn.fi/urn:nbn:fi:fsd:T-FSD2673)”
2. Provide a **description(s)** (abstract). You don’t need to provide an elaborate abstract for the purposes of this exercise. However, keep in mind that, together with the title, the abstract is one of the fields that a potential user will check in order to decide whether the dataset is useful for him/her or not. Have a look at the checklist in [this page](https://www.aalto.fi/en/services/describing-datasets-in-the-data-repositories) to get a feeling of what a good abstract should include.
   1. You can add more than one description and indicate a description type. For example, you might want to add a generic abstract and a second “Methods” abstract, where you describe your methodology in more detail.



1. Add yourself as an **author**. If you have an ORCID, make sure to add it under the “Name identifier” field, and indicating ORCID under the “Scheme” field.



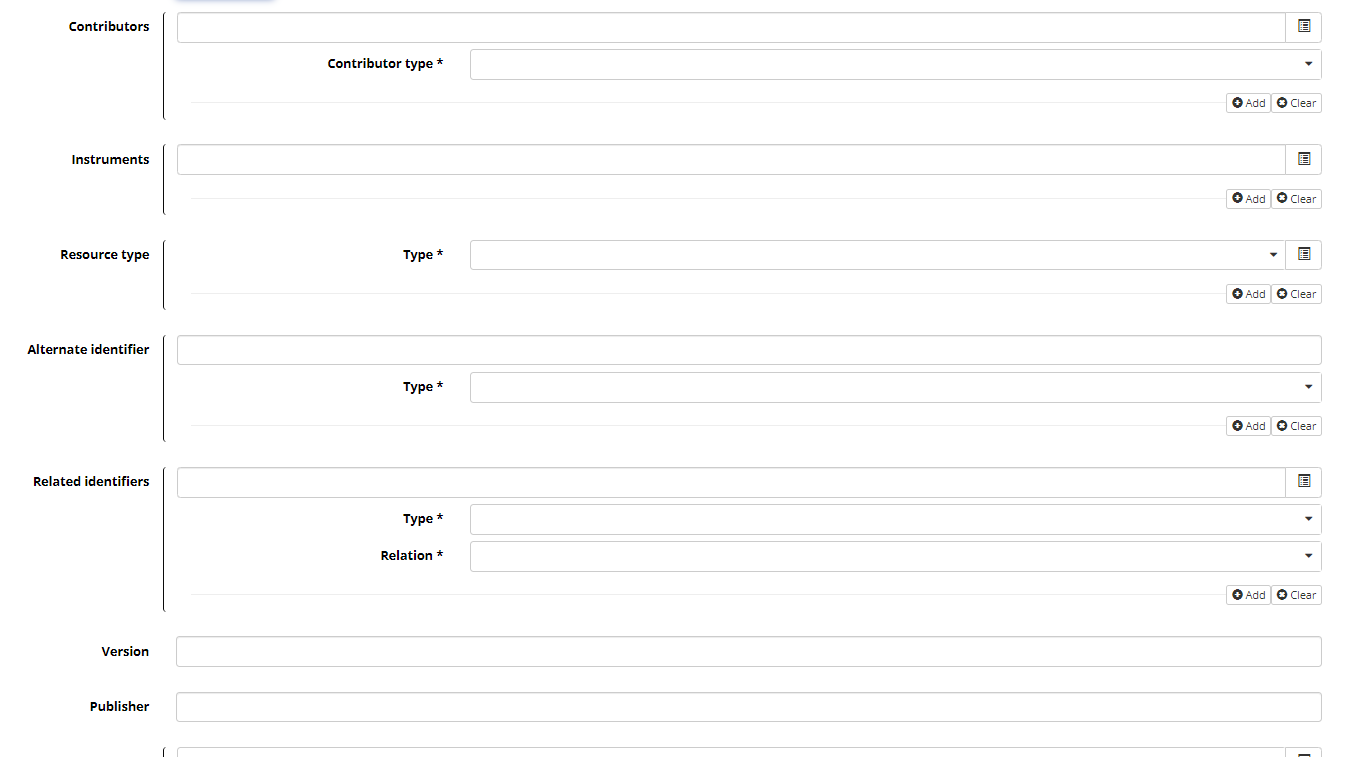
1. Fill in the metadata fields related to the access and reuse conditions. Use the select license button to select a standard license, such as Creative Commons Attribution (CC BY).



1. Choose a discipline from the dropdown list.
2. Choose a language and add some **keywords**. You can use free text for the purpose of this exercise. In real life (and as much as possible), use controlled terms, classification codes or thesauri in your discipline to provide a list of keywords. Some examples of thesauri and controlled vocabularies and resources in which you can find terminology standards:
   1. [IEEE Thesaurus (Institute of Electrical and Electronics Engineers)](http://www.ieee.org/publications_standards/publications/services/thesaurus2.html)
   2. [EcoPortal](http://ecoportal.lifewatchitaly.eu/)
   3. [Food Ontology](https://foodon.org/)
   4. [Agronomy Ontology](https://bigdata.cgiar.org/resources/agronomy-ontology/)
   5. [AGROVOC](http://aims.fao.org/vest-registry/vocabularies/agrovoc)
   6. [GeoNames](http://www.geonames.org/)
   7. [Research Organisation Registry](https://ror.org/search)
   8. [Ontology Lookup Service](https://www.ebi.ac.uk/ols/index)
   9. [FAIRsharing](https://fairsharing.org/standards/)
3. Indicate a contact email and add the publication date.



1. Explore the remaining metadata elements (expand by clicking on “Show more details”) that can be provided. As you can see, it is possible to provide detailed contextual information:
   1. References to related publications or related datasets (related identifiers)
   2. Other contributors and their roles (e.g. data collector, supervisor…).
   3. Instruments, language(s), spatial and temporal coverage, etc…



1. Once you are satisfied, publish your draft dataset.
2. FYI, you can find a dummy DOI on the left panel at the page of your record.

