# CESAB Thesauform annotation manual

The CESAB Thesauform is a tool that allows the creation of a simple thesaurus for terms, synonyms, and their meanings, enabling reliable machine-discovery of data. The present use is to create a *Thesaurus for Ecological Observations*. Several terms have been uploaded and arranged in a loose suite of relationships. These need to be checked and defined. This manual shows you how to annotate or modify the terms already present, propose any re-organisation of their occurrence, and add any terms that you think are missing.

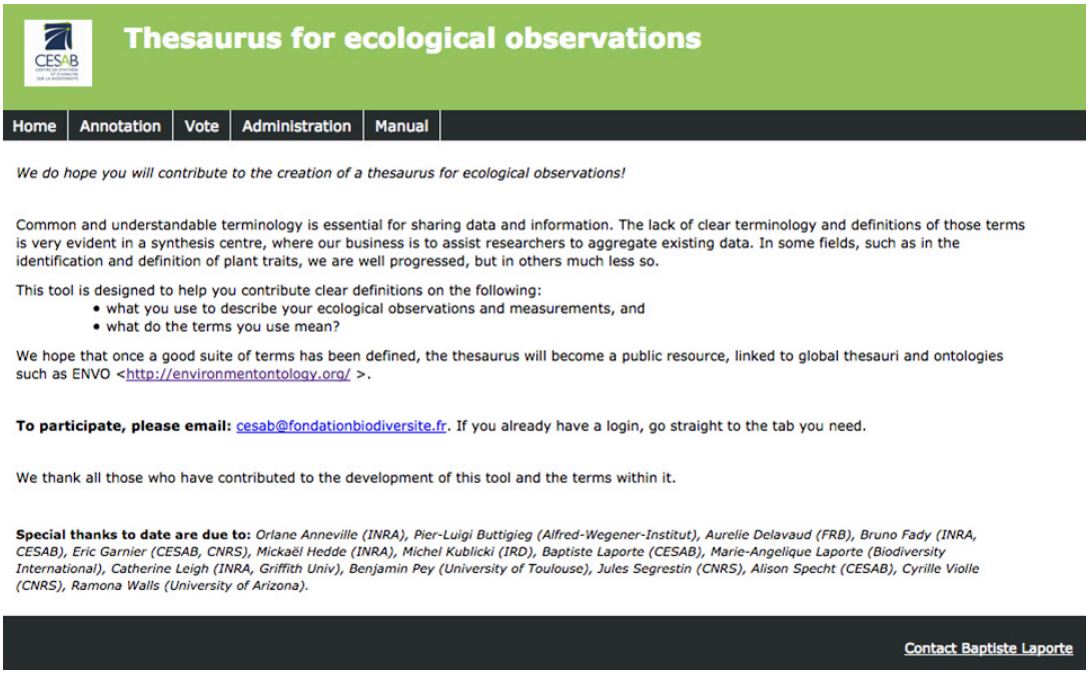
Once submissions have been made, the Thesauform also provides the opportunity for you to 'vote' on the suitability of the terms, definitions and sources proposed. These are then considered by an expert committee and the most favourable term or terms, with annotations, will be submitted to a relevant global ontology (e.g. ENVO) for ratification and the source, CESAB, will be acknowledged.

**What is an annotation?** Simply, an explanation of something. Each term in the Thesaurus needs to have a **definition** (one term can mean different things to different people). We would like this definition to be 'discoverable' (the provenance of the definition), for example from a book, journal article, or from a published ontology or other web site. Each term may have synonyms or units, for which we also provide an entry.

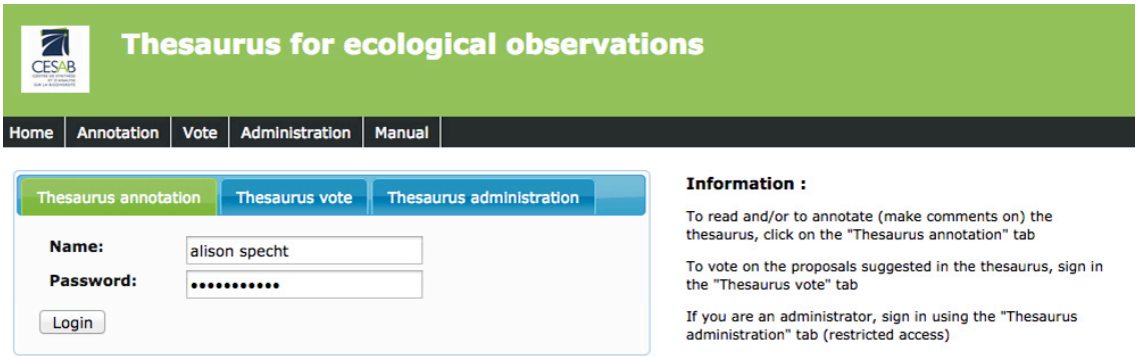
*If you are interested in using this tool for your own project, please contact Baptiste Laporte (email below).*

## To start your journey

On the 'home' page you will see the following welcome. This Manual is downloadable from the appropriately-named 'Manual' tab on the menu bar. If you need a login, contact CESAB as instructed.

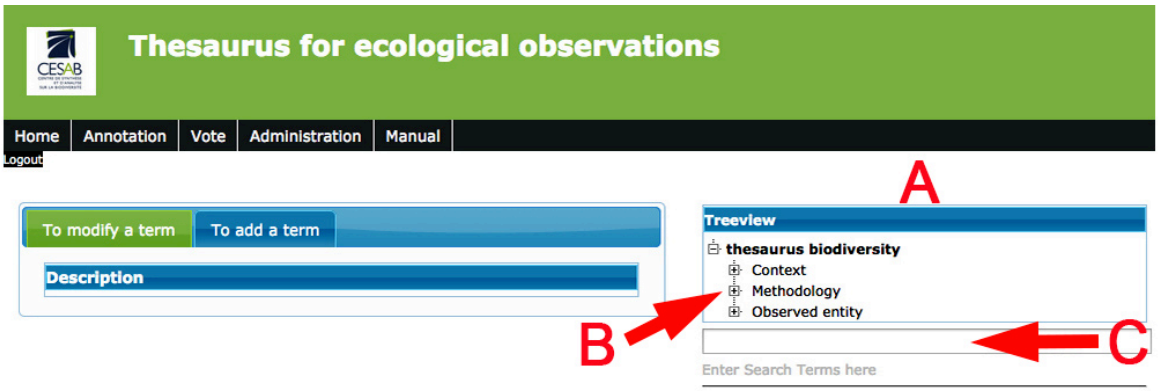



Once you have your login information, your first entry will be to the 'Annotation' tab. On later occasions you may enter using the 'Vote' tab. Only administrators can enter using the 'Administration' tab.



If your login information is accepted, you will see the interface below. On the right of the screen you will see the main component of the Thesauform, the visualization tree (A: Treeview). All terms have been aggregated under three headings: Context, Methodology, and Observed Entity.

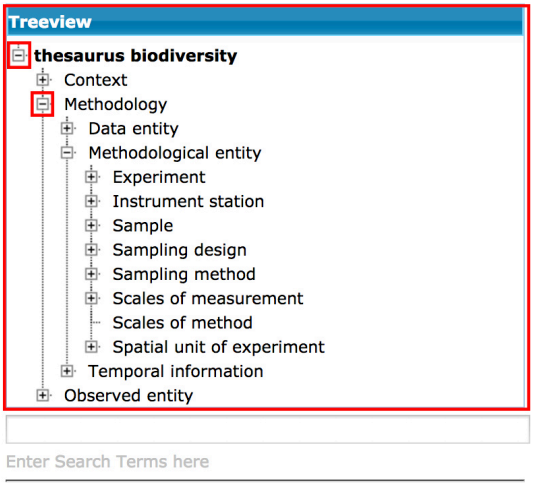
All the terms currently in the thesaurus can be discovered by clicking on the '+' symbol (B) or you can search for them directly via the text box (C).




Clicking on the '+' symbol opens the related layers of the tree as shown on the left screen capture below. By continuing to scroll and click you can search for the term in which you are interested.

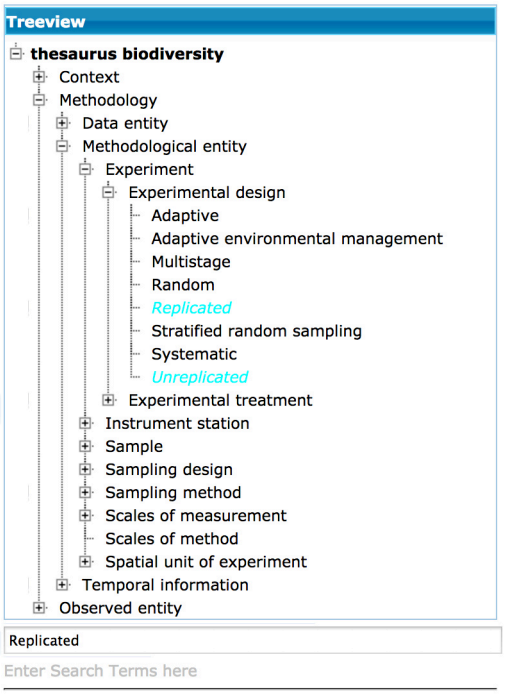
As another option, entering a term into the search box (C above) will open the tree, highlighting the wanted term and those with similar roots (below right).

Using the '+' link



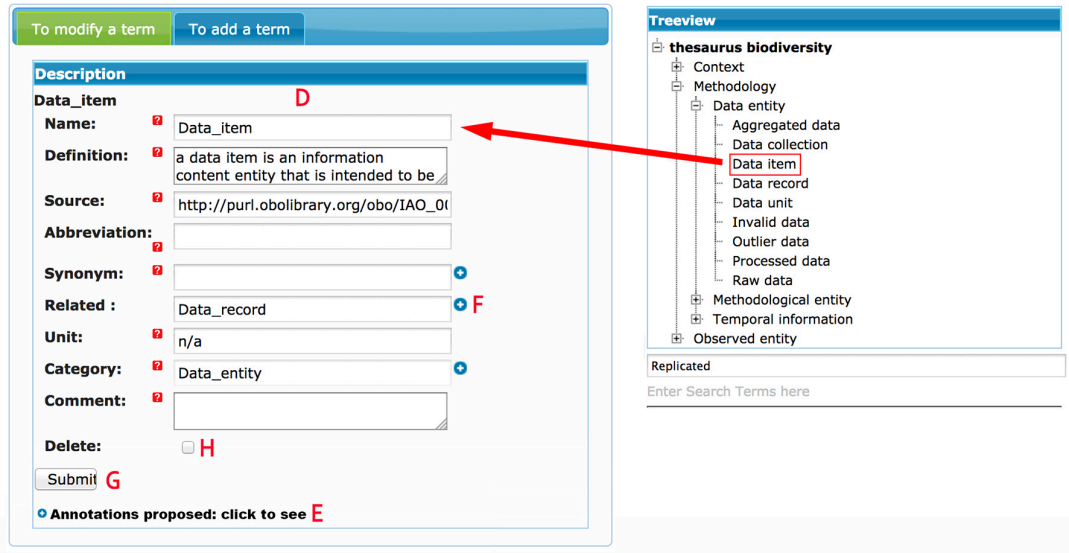

In the example on the right, the search term was 'replicated' and the search returned 'replicated' and 'unreplicated', both highlighted in aqua.

Using the search interface



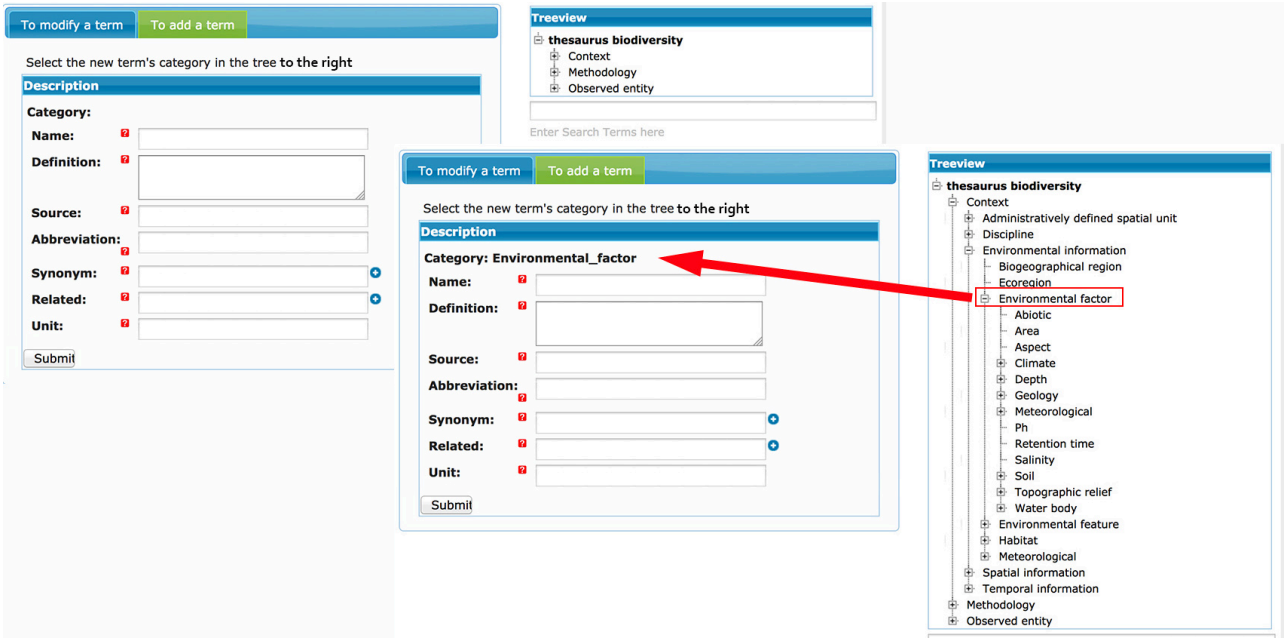
To see the details about a term or concept, you need to click on its image in the tree. The left side of the screen will then open to show the information already added for that term (D in the screen shot below). For most attributes there is a button '?' which, when clicked, will show a brief explanation of the item. After reading each, click on the text to close.

You can enter new information for each field or attribute. If you are the first to enter information (a definition, reference, abbreviation and so on), your information will appear directly under the term when you display it again. If you were not the first, you will not replace any previous entries, and your new information will appear as an annotation below the table and others can click on a link to see what was proposed (E). If you wish to add more than one synonym, related term or category, you can click the '+' (F) sign to open up a new entry cell. If you wish to store it as a choice for others to consider, you need to click 'submit' (G). Please check that all is as you want it before clicking the submit button.




A definition should be accompanied by a source: a reference to a book, journal article or a url. For guidance on how to write a good definition, please see the Appendix of this Manual. You can delete a term or concept by clicking the 'delete' checkbox (H).

To add a new term click on the 'Add a term' tab. A similar window to that shown above will appear and allow you to add the information wanted (below left). The category under which you would like the term to appear needs to be nominated. To do this find the category by expanding the tree and click on the desired one. It will appear on the left and then enter the details of your new term. Be careful to not overlay an existing term! Check first using the search interface (B or C). Then click on the submit button.



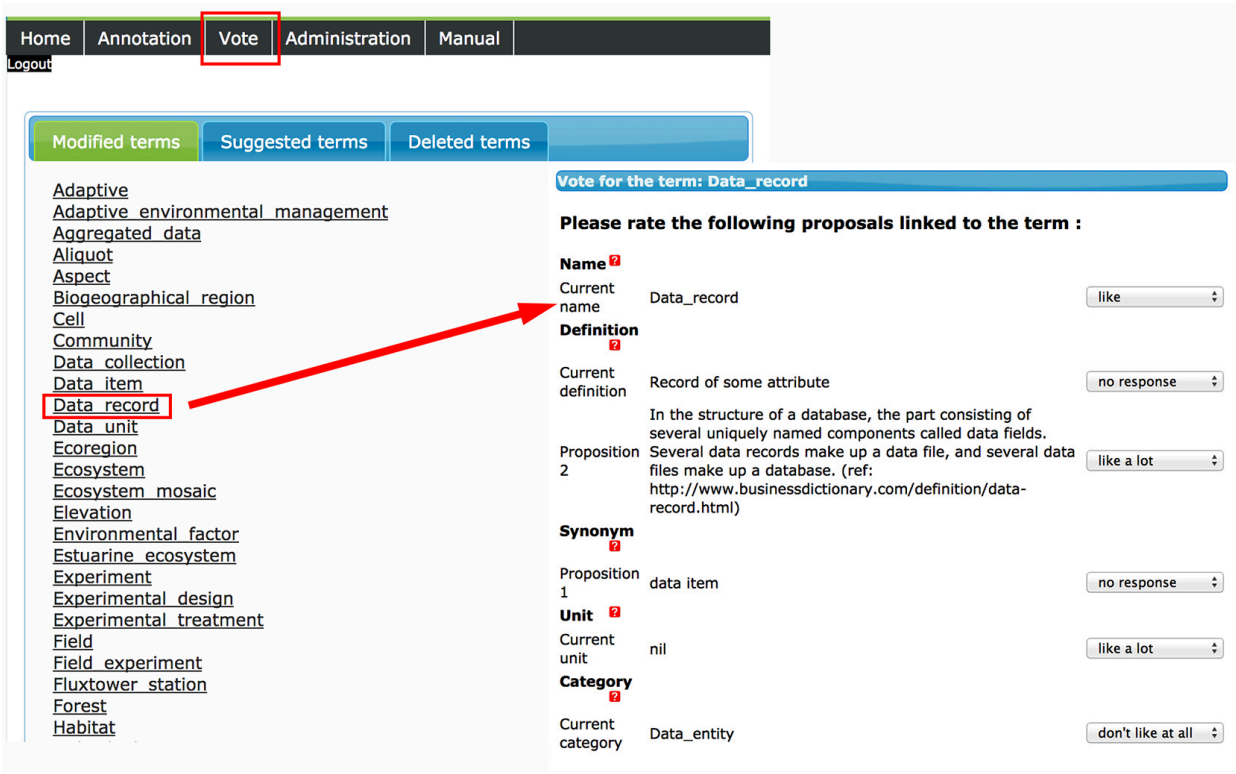
All annotations and new terms will appear in the Voting panel.

You can make your comments about what you think of any proposals made through the 'voting' tab. By clicking on this three sub-tabs appear:

- one for modified (or annotated) terms,
- one for proposed new terms, and
- the third for any terms proposed for deletion.

Clicking on 'modified terms' gives you a list of all terms that have had their details modified or annotated in some way (see below left). If you click on one of these terms—in this example 'data record'—the options entered for that term will appear and alongside each a pick list is available. There is a choice of five options on the picklist from 'don't like at all' to 'like a lot' with 'ambivalent' in the middle.

In this example for 'data record' two definitions have been proposed, and I have voted for the second (like a lot) and not responded to the first. I have selected 'don't like at all' for the current category 'data entity'. You will note the source of the definition is incorporated into the definition display.



**How to write a good definition**

Following realist principles [1], terms in the CESAB thesaurus should be used to classify entities that exist, including material, processual, qualitative, and informational entities (e.g. rocks, photosynthesis, colours, or data, respectively). Therefore, definitions of CESAB thesaurus terms should unambiguously specify the criteria for determining whether an entity belongs to the class of entities covered by the term. This is opposed to "concepts", which are ideas, notions, or units of thought and are subjective [2]. Although concepts are useful for cataloging and discovery, they have limited utility for rigorously integrating scientific data.

**Ontologies**, which are concerned with grouping entities into classes (also known as types), often use Aristotlean or genus-differentia definitions for specifying the conditions of membership in a class. Such definitions first define the broader class to which all members of a more specific class belong (the genus) and then specify the unique characteristics shared by all members of the more specific class (the differentiae). This method requires first understanding the nature of the term, so that the genus can be correctly specified. For example, if you want to define the term "observation", you must first determine if you are referring to the process of observing, the characteristics being observed, or the data that is generated by an observing process. Differentiae may fall along any number of axes, including physical or behavioural characteristics, temporal or spatial limitations, parenthood, and developmental relations.

Here are some examples of genus-differentia definitions from ontologies:

**ecological community** (http://purl.obolibrary.org/obo/PCO_0000002): A community of at least two different species, living in a particular area. Must have at least two populations of different species as members.

In this case, the parent term, **community** (http://purl.obolibrary.org/obo/PCO_0000028), is defined as 'A collection of organisms connected by social or biological relations (biotic interactions).'

**embryo apical cell** (http://purl.obolibrary.org/obo/PO_0025284): A sporophyte meristematic apical cell that is part of a plant embryo and is the uppermost cell formed after the first division of a zygote.

**beaver pond** (http://purl.obolibrary.org/obo/ENVO_00000270): A pond that has formed as a consequence of the activities of beavers, building a beaver dam.

While such explicit language is important for precisely defining terms in ontologies, it can be awkward and unfamiliar for experts in other domains. Therefore, alternative definitions that use more 'every day' language can also be used, as long as they convey the meaning of the more formal definition. For the CESAB thesaurus, we encourage you to use wording that is appropriate for your discipline, and to be as rigorous and unambiguous as possible in defining terms. It is more important to state clearly what you mean than to follow a strict 'An X is a Y that Z' format. Clarifying comments are encouraged.

1. Smith B., Ceusters W. (2010) Ontological realism: a methodology for coordinated evolution of scientific ontologies. *Appl. Ontol.* 5: 139-88.

2. https://www.w3.org/TR/skos-reference/#L1437