

EcoTaxa



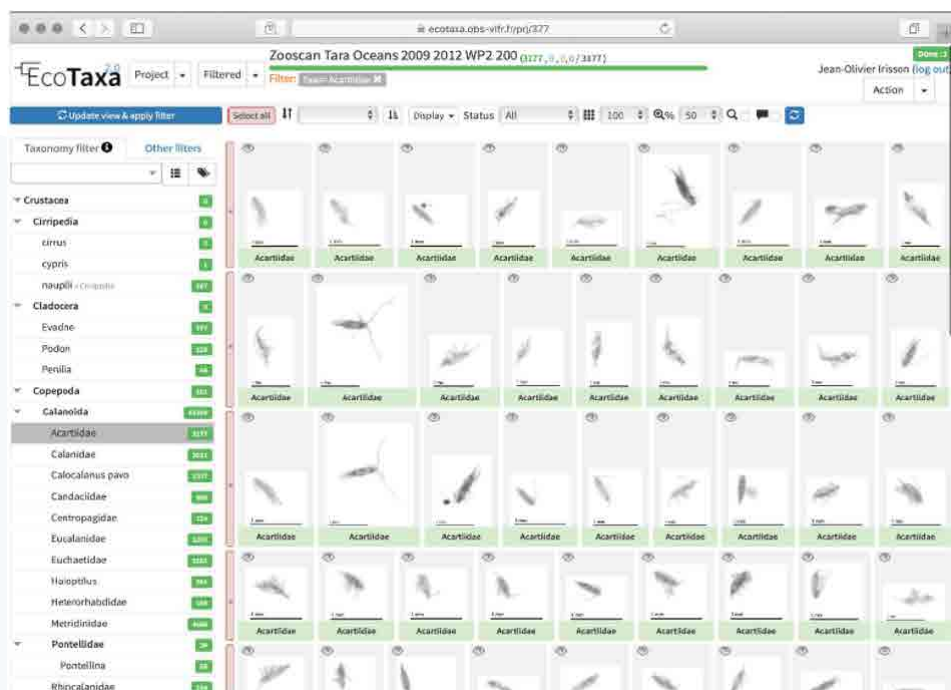
About

EcoTaxa is a web application dedicated to the visual exploration and the taxonomic identification of images of plankton. EcoTaxa was born from the experience developed at Laboratoire d'Océanographie de Villefranche (LOV) regarding the quantitative, high-throughput imaging of plankton and of the Oceanomics project which exploited the data collected during the Tara Oceans cruise, including quantitative imaging. It is now developed mainly through the WWWPIC project funded by the Belmont Forum and as part of the Blue-Cloud project.

The aim of EcoTaxa is to centralise images of plankton, to allow their collaborative classification along a universal taxonomy and to accelerate this process through machine learning. It produces ecological data in the form of concentration and biovolume of organisms in a given taxon, at a given point in latitude, longitude, depth and time. Visitors have free access to the specimens that have been already identified by taxonomist experts. They can explore the database by navigating along a taxonomic tree. Then images can be filtered according to several sampling criteria (location, time, etc.). For operators, easy-to-use tools are provided to suggest a classification for each image over large datasets, through supervised machine learning.

Type & number of data sets

Currently, EcoTaxa contains circa 150 million images of which circa 63 million have been annotated in over 2000 datasets. Of these, circa 30 million images concern living organisms. The growth rate is circa 4 million new images per month. Not all of these datasets will be accessible for Blue-Cloud, because this depends on the data policy of data providers, who come from circa 350 organisations worldwide. Some of those datasets (in particular the Tara Oceans ones, needed by the demonstrators) will be sent from EcoTaxa to EMODnet Biology and be harvested by the Blue-Cloud data discovery service from there. Then, Blue Cloud users will be able to browse them in finer details through the EcoTaxa API.

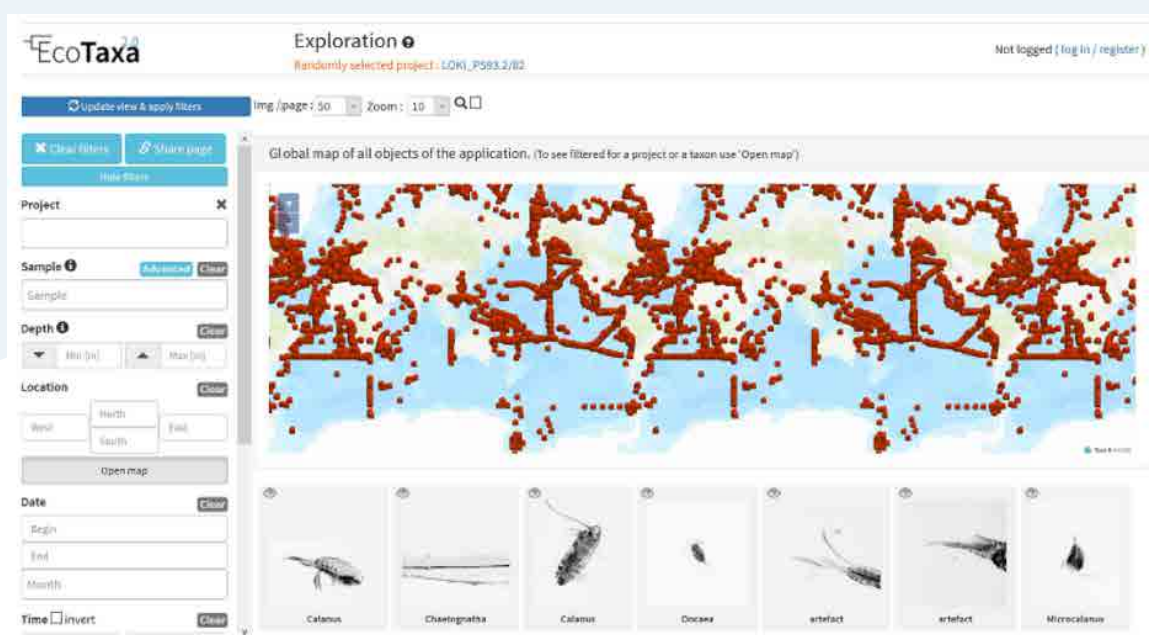


EcoTaxa dashboard for annotation

Core Services

The core of EcoTaxa is the rapid identification of large numbers of images by a combination of machine learning and human validation. The resulting data can be browsed and downloaded once access to the dataset is granted by the dataset owner.

Data discovery used to be manual. Validated data from all public datasets can be browsed at <https://ecotaxa.obs-vlfr.fr/explore/>. As part of the WWWPIC and Blue-Cloud projects an Application Programming Interface was built to allow programmatic browsing of public datasets that were uploaded to EMODnet Biology.



EcoTaxa interface for exploring images

Function in BlueCloud

Thanks to the collaboration within Blue-Cloud, EcoTaxa can accelerate data processing and sharing for the “Zoo & Phytoplankton EOVS Products” and “Plankton Genomics” Demonstrators. Within the goal of Blue-Cloud to federate information from different databases at EU level, EcoTaxa contributes data to EMODnet Biology, which is then aggregated in Blue-Cloud’s data discovery service, and also provides finer scale data as a second level of query.

URL: ecotaxa.obs-vlfr.fr

Partners involved: Sorbonne University