

ECOTAXA, a web based application for collaborating on large plankton image datasets

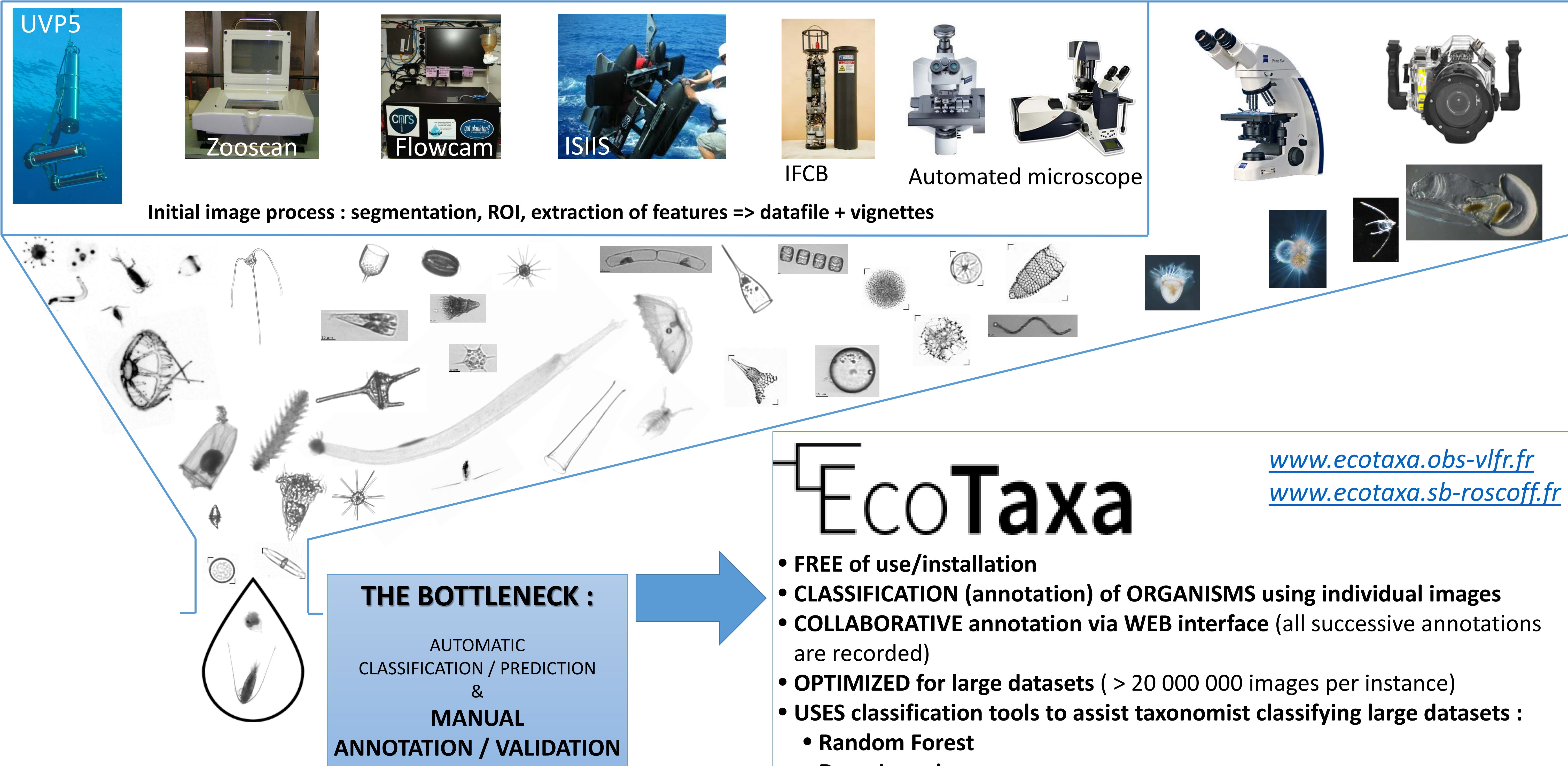
Marc Picheral¹, Sebastien Colin², Jean Olivier Irisson¹, Amanda Elineau¹, Colomban de Vargas² & Lars Stemmann¹

Over the last 25 years, our team developed laboratory and in situ imaging instruments to study marine particles and zooplankton. The Zooscan (Gorsky, 2010) and the Underwater Vision Profiler (Picheral, 2010) are now used worldwide (102 Zooscan and 19 UVPs). We also developed a software tool-chain, based on ImageJ, to analyze the images acquired (Zooprocess). This suite can now process many kinds of images and has been used with other instruments such as ISIS, Flowcam, etc. It can pilot the imaging instrument, subtract the background of the image and segment regions of interest, measure ~50 properties on each object and create labelled vignettes. These properties and vignettes were used by the Plankton Identifier application to automatically classify objects using machine learning. This tool assisted the users for the creation of the learning sets and the validation of the identification proposed by the classification algorithm; it achieved good success using Random Forests.

Plankton Identifier paved the way for a new web based application enabling a network approach to plankton identification on images. Ecotaxa (<http://ecotaxa.obs-vlfr.fr/> & <http://ecotaxa.sb-roscoff.fr/>) was specified in 2015 in collaboration with the CNRS EPEP team in Roscoff in the frame of the Tara Oceans expeditions and released in 2016. As its predecessor, Ecotaxa handles images of individual organisms, proposes identifications using machine learning, and keeps all metadata associated with each vignette, from the acquisition to the final identification. However, the founding principles of Ecotaxa are that:

- The identification of organisms is collaborative, through the internet;
- Every change is explicit and recorded in a robust relational database (including the simple confirmation of a correct identification);
- Identifications are based on a universal taxonomy (<http://unieuk.org/>) that allows to link the morphology of organisms with genomic information.
- ECOTAXA can easily import image datasets from any instrument

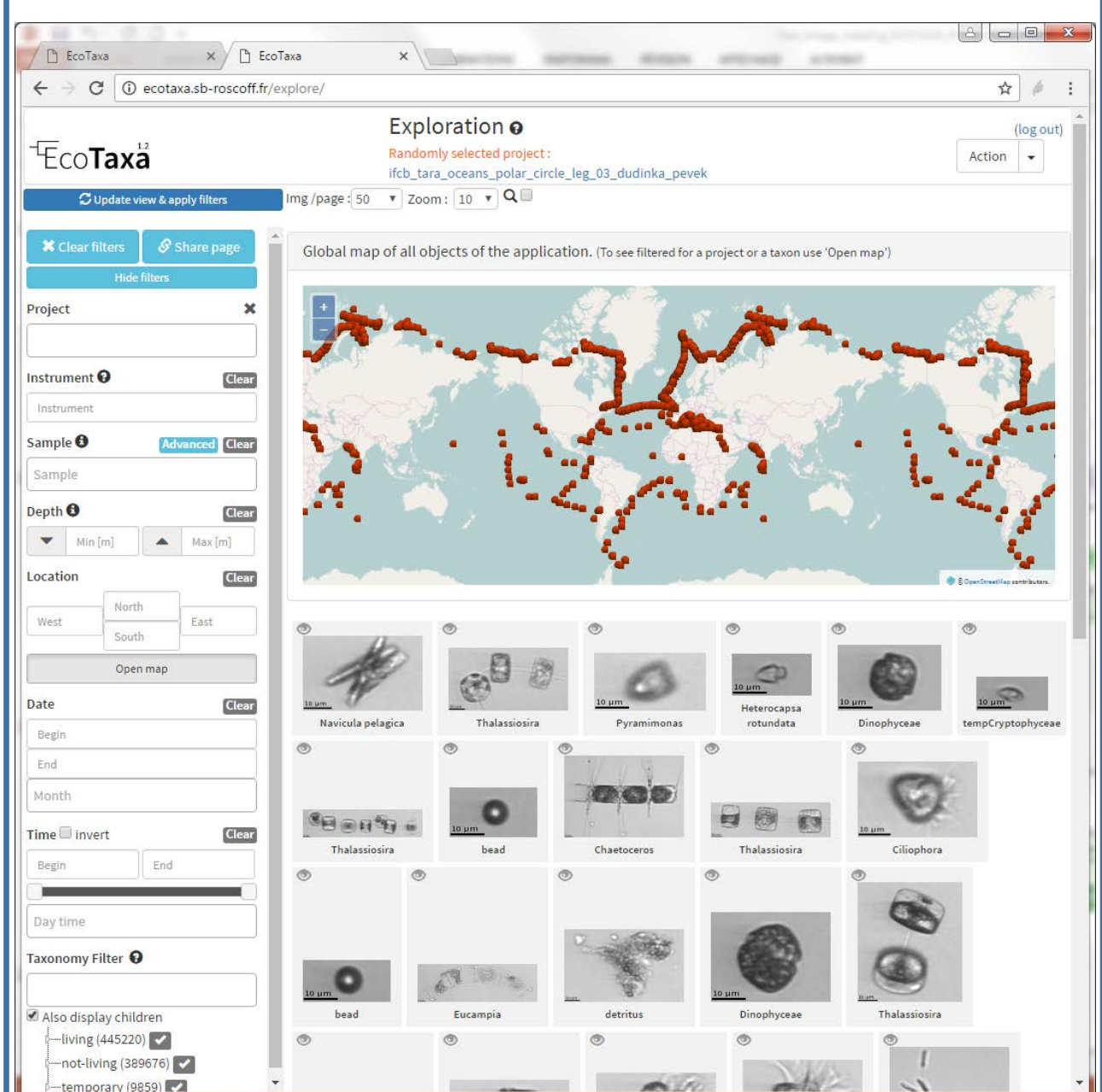
The application already hosts over 30 million images of plankton, about 30% which have been verified by experts. We are entering a round of updates that will allow communication between Ecotaxa instances, ease the subscription of new users, ease the selection of a subset of objects and offer new deep learning algorithms for automatic classification. We are also publishing instrument-specific taxonomic guides to homogenize the sorting of the images by different experts worldwide.



www.ecotaxa.obs-vlfr.fr
www.ecotaxa.sb-roscoff.fr

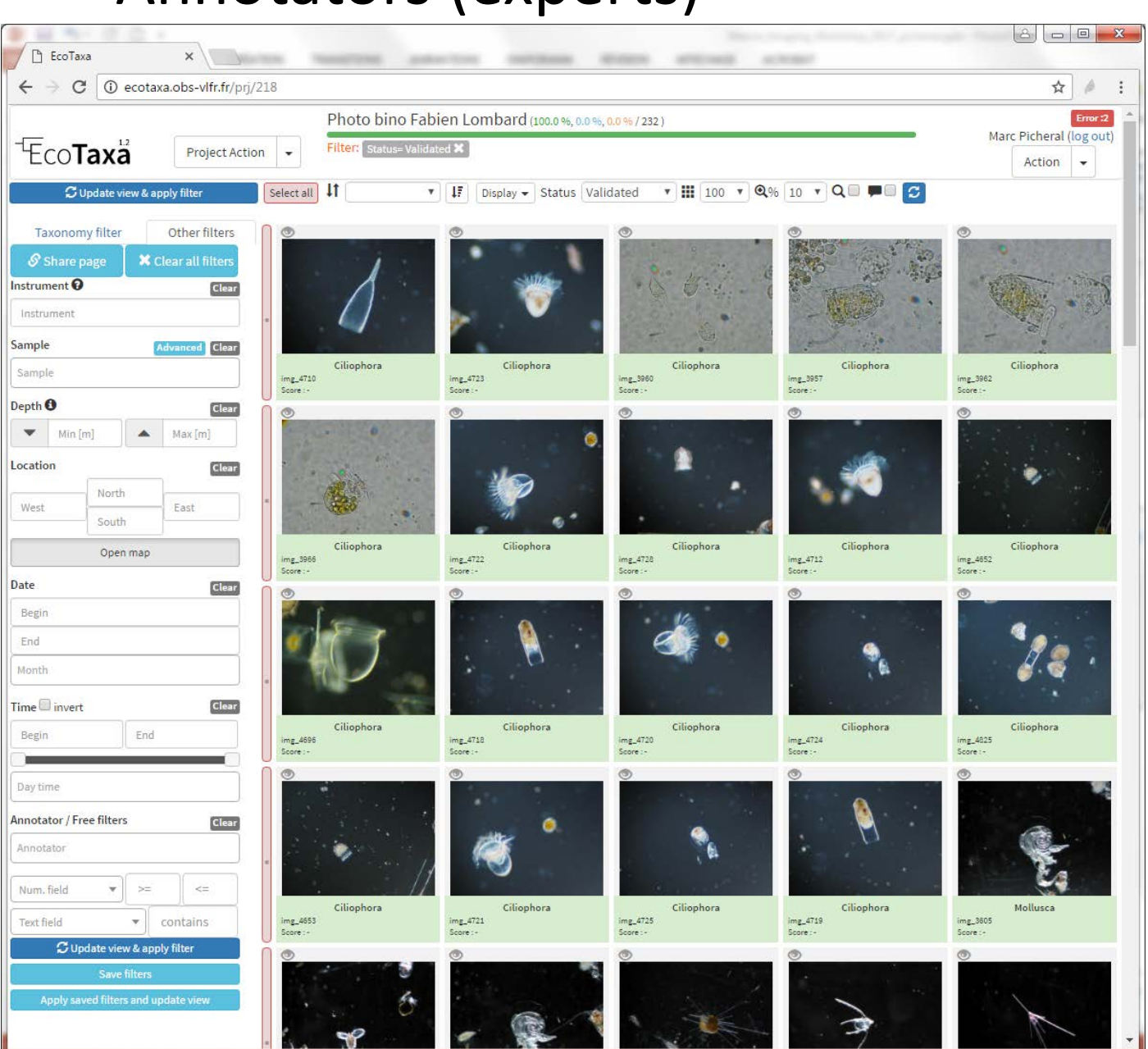
EcoTaxa today :

Public exploration of validated images across oceans



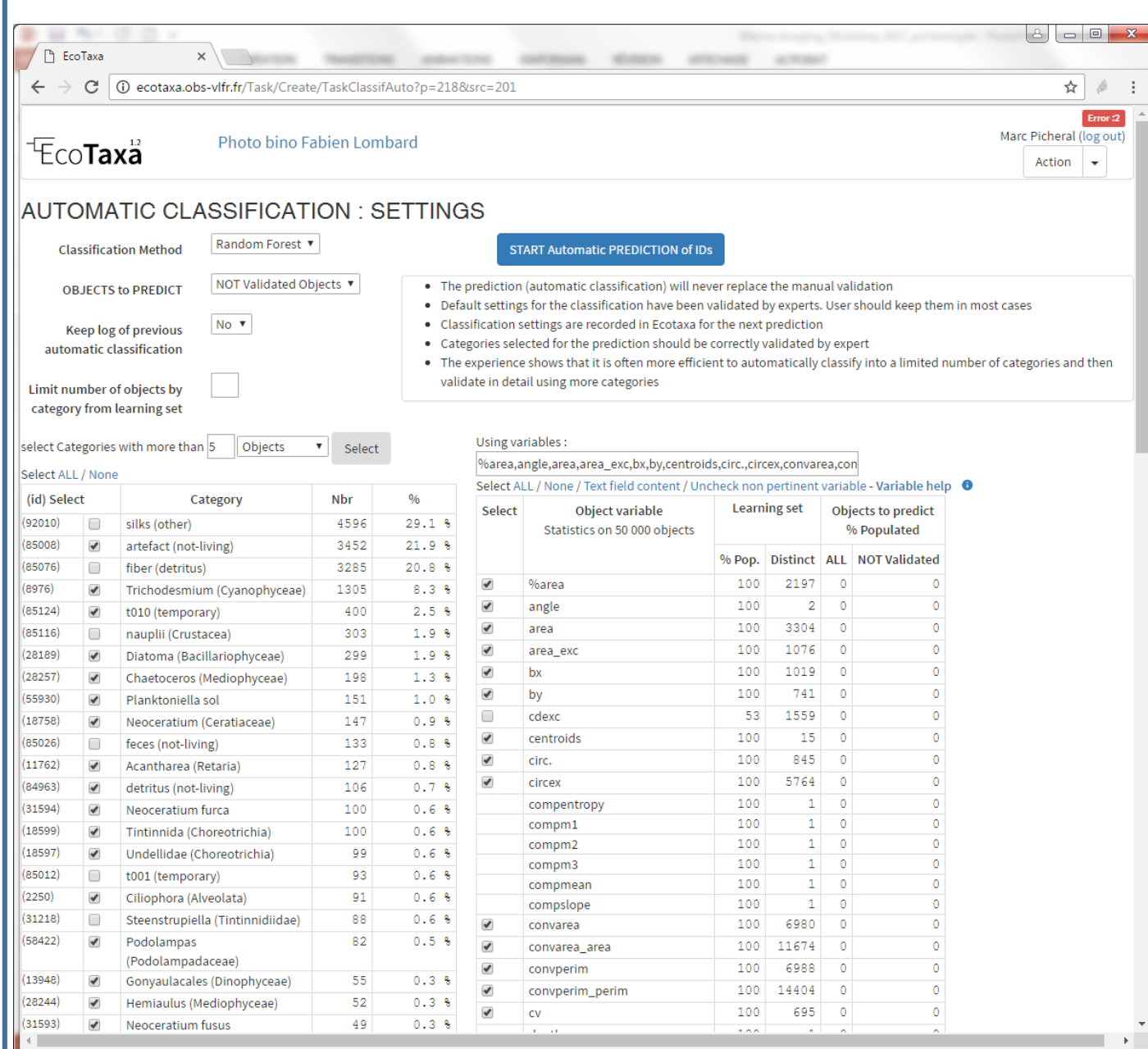
Powerfull filters:

- Taxonomy
- Date/time/month/Depth/sample...
- Annotators (experts)



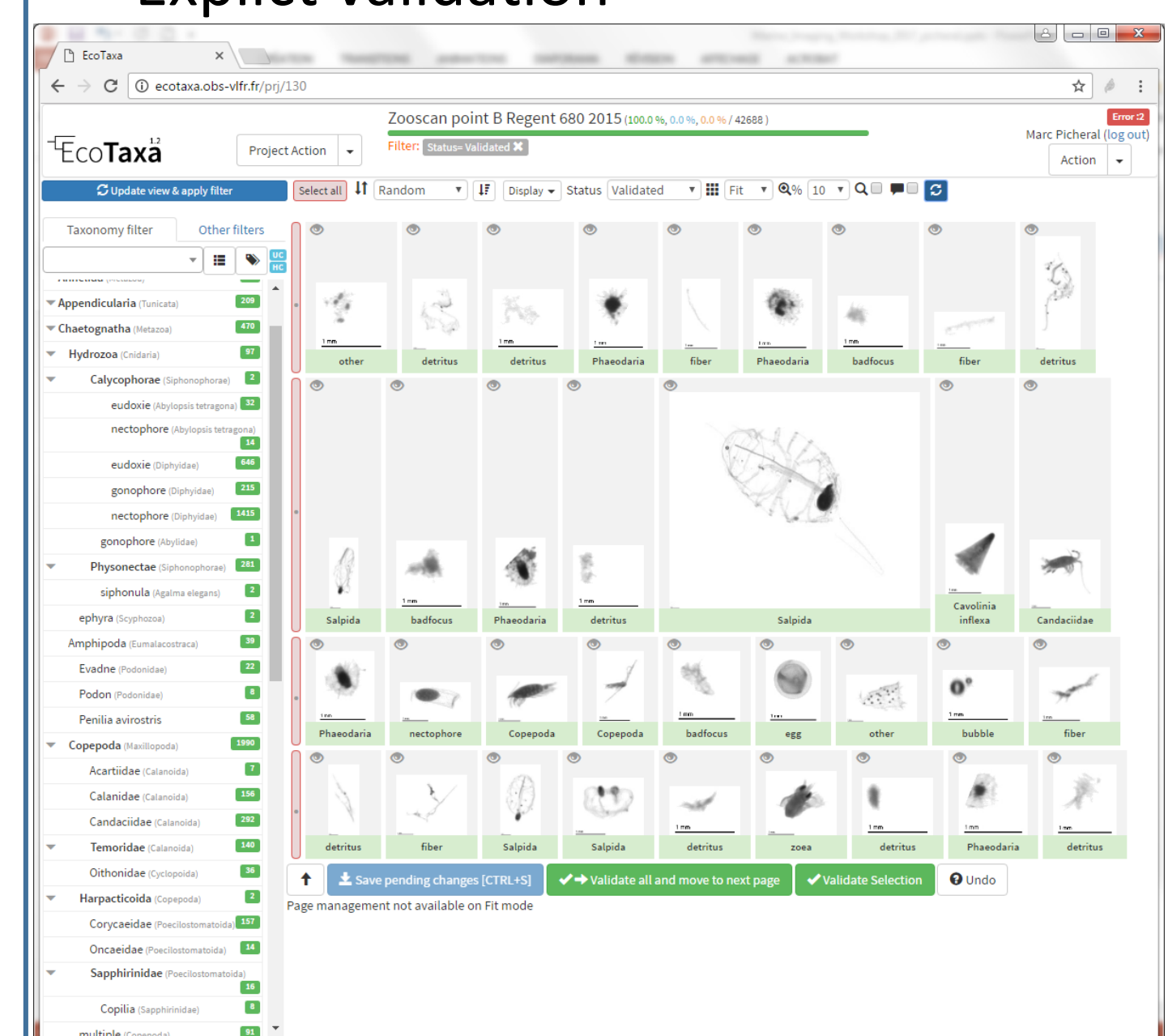
Automatic classification

- Random Forest
- Deep Learning (soon)



POWERFUL manual annotation

- > 20 000 images / day !!!
- All operations recorded
- Explicit validation



Instruments datasets : IFCB, FlowCam, Zooscan, ISIS, Camera, UVP5, HCS1, Zoocam...

> 30 10⁶ images hosted / > 10 10⁶ annotated (prediction + validation by experts)

➤ **Today, the MOST IMPORTANT worldwide dataset of annotated plankton images !**