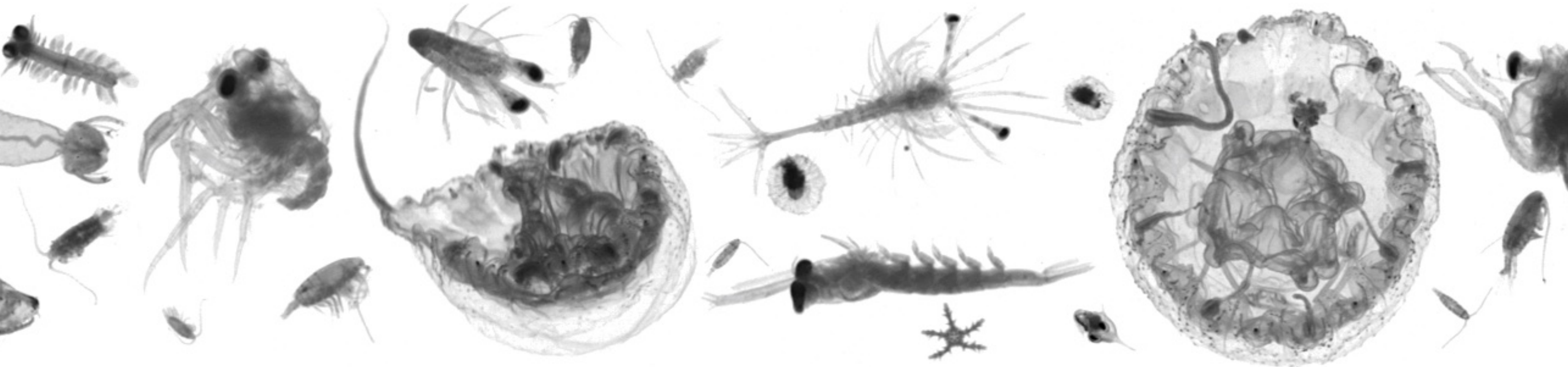


EcoTaxa

a tool to support the taxonomic classification of large image datasets through supervised machine learning



Why EcoTaxa?

ZooScan

1.5M objects/y

FlowCam

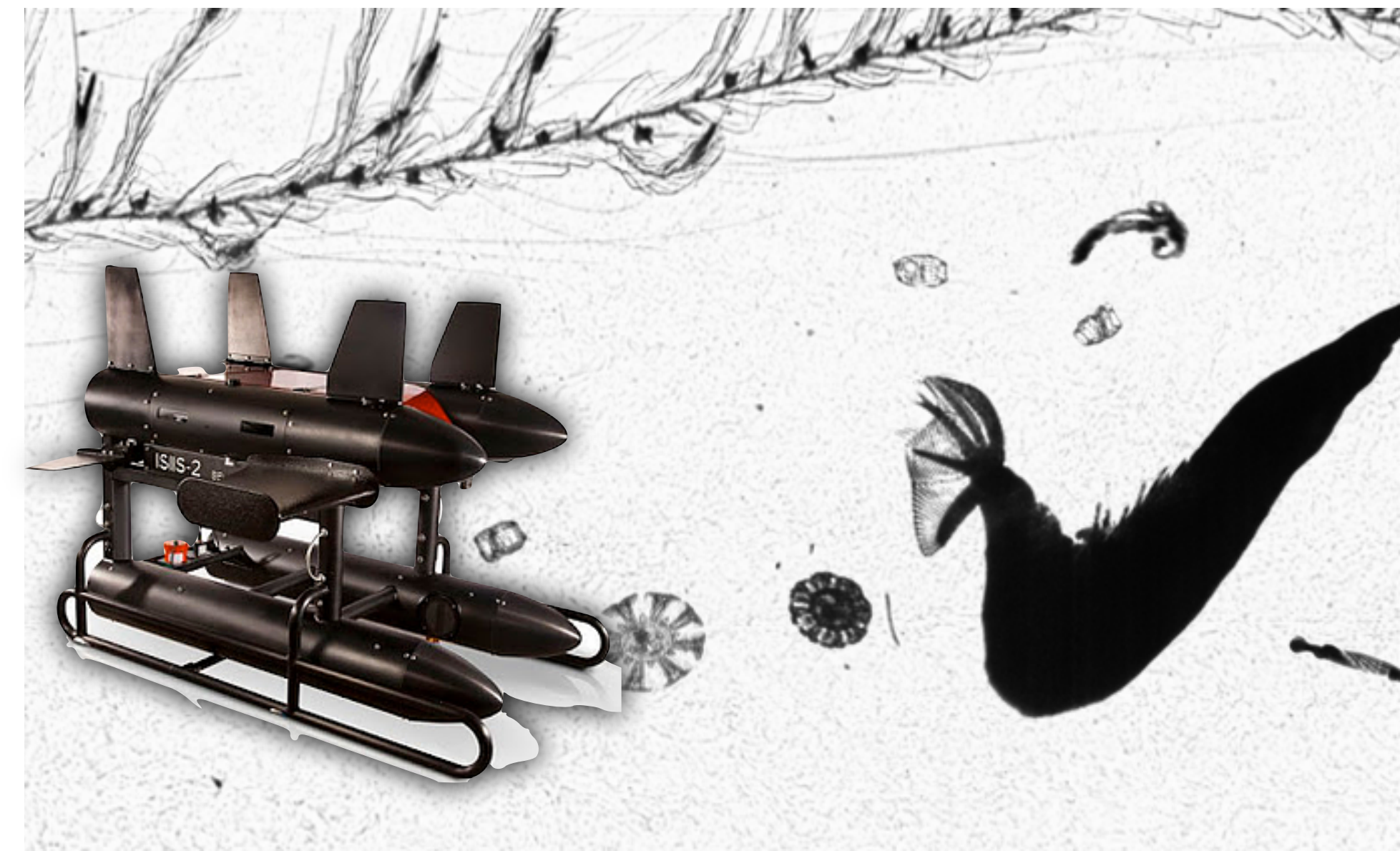
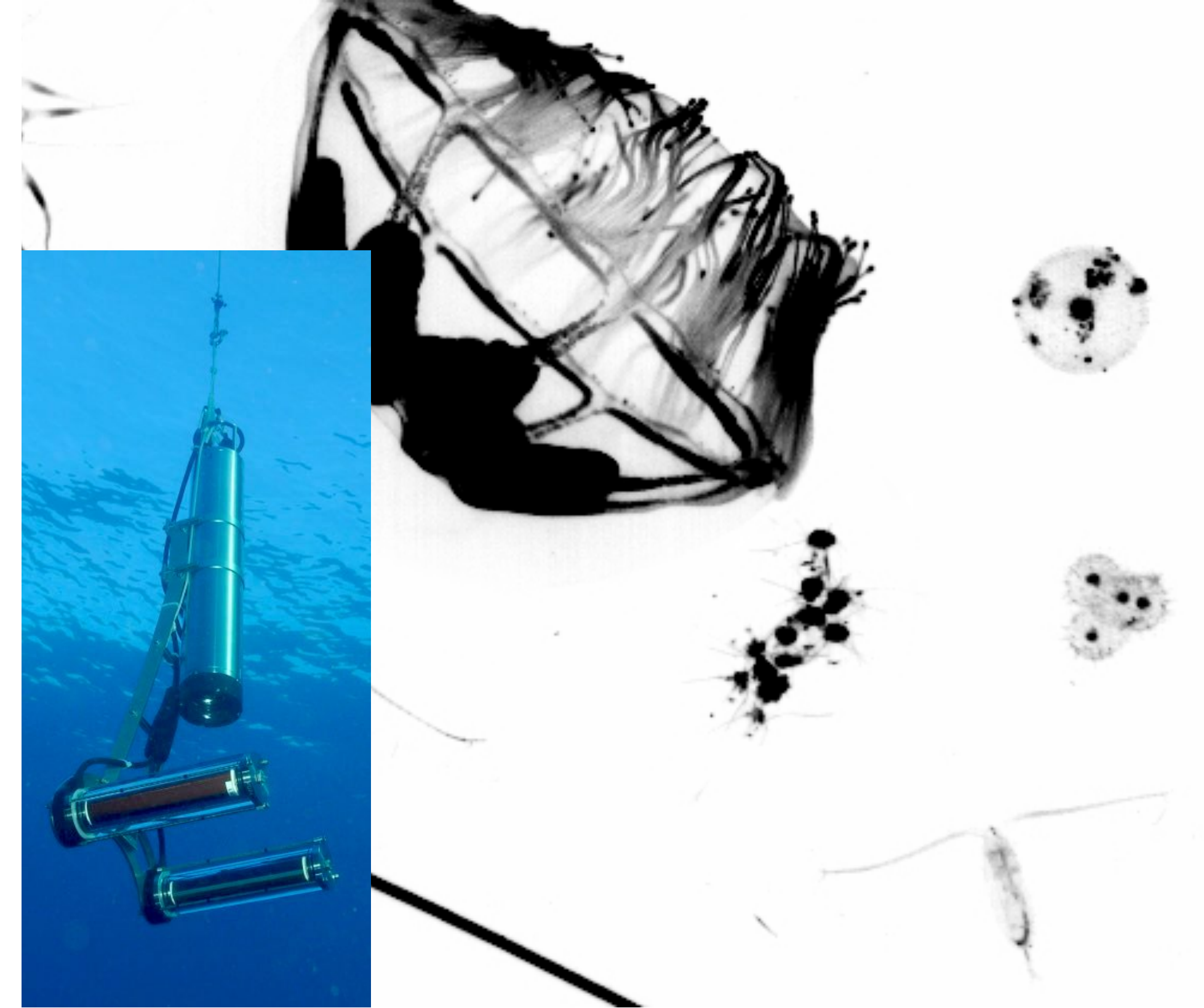
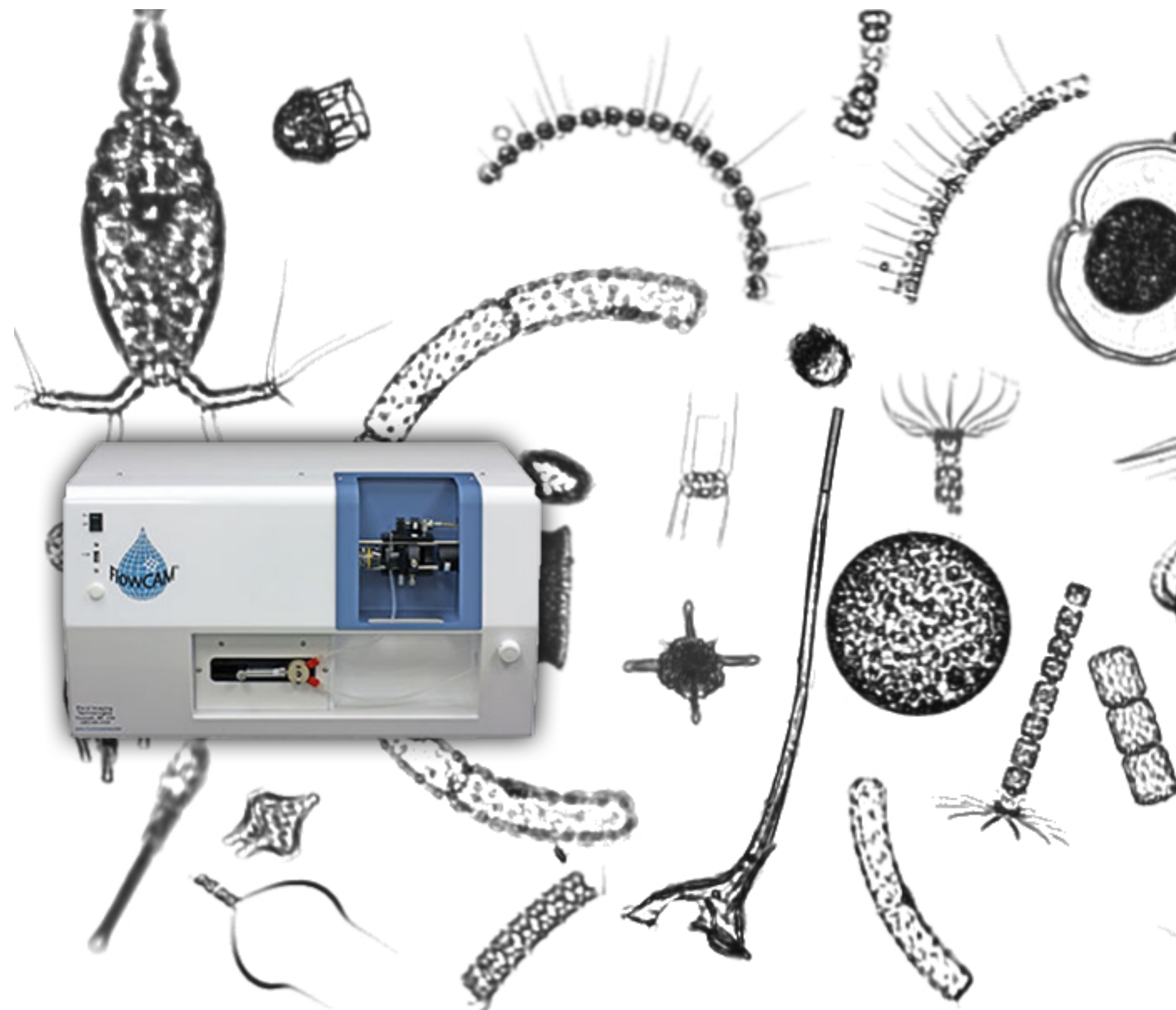
~5M objects/y

UVP

~10M objects/y

ISIIS

100M objects/y



How?

Step 1: Import images (and metadata) into a project

Demo Zooscan JO low taxo res (241, 1156, 0, 0 / 1397)

Done :2

Close

Open in a separate window (right click to copy link)

Project: [Demo Zooscan JO low taxo res](#) (managed by : [Jean-Olivier Irisson](#))

To report a mistake, contact [Jean-Olivier Irisson](#)

Classification :

fiber<detritus

fiber < detritus < not-living (id=85076)

validated by Jean-Olivier Irisson (irisson@obs-vlfr.fr) on 2022-11-21 20:37:38

Complementary information ([edit](#)):

Image list : 1



1 mm

Set a new classification :

Save as Validated

Save as dubious

Enable Editing

Close

Object details

Sample details

Acquisition details

Processing details


Classification change log

Map

Edit complementary informations

longitude	7.31567	latitude	43.68500	Date	2018-02-05	Time (daytime)	12:00:00 (Day)
Depth min	0.0	Depth max	75.0	Classif auto	Copepoda (0.762)	Classif auto when	2021-07-29 14:09:52.442648
Object #	83060411	Original Object ID	wp220180205_d2_1_537				
lat_end		lon_end		area	1918	mean	166
stddev	50.73	mode	243	min	78.00	max	248
x	63.21	y	61.49	xm	63.33	ym	61.95
perim.	384	bx	4189	by	11923	width	133
height	99.00	major	112	minor	21.80	angle	140
circ	0.16	feret	156	intden	318630	median	163

A note on the data architecture

PROJECT
title
settings

mapping

SAMPLE
lat/lon
date/time
<i>volume</i>
<i>operator</i>
<i>instrument</i>
<i>resolution</i>
...

SUBSAMP
<i>fraction</i>
<i>px size</i>
<i>frac.method</i>
<i>frac. mesh</i>
...

OBJECT
objID
taxoID
who
when
<i>area</i>
<i>major ax.</i>
<i>minor ax.</i>
<i>feature 4</i>
...
<i>feature n</i>


IMAGE
path
rank
w / h

TAXO
taxoID
name
parentID

USERS
name
email
permissions

mandatory
fixed format
internal
free

A note on the data architecture

PROJECT
title
settings

mapping

SAMPLE
lat/lon
date/time
<i>volume</i>
<i>operator</i>
<i>instrument</i>
<i>resolution</i>
...

SUBSAMP
<i>fraction</i>
<i>px size</i>
<i>frac.method</i>
<i>frac. mesh</i>
...

OBJECT
objID
taxoID
who
when
<i>area</i>
<i>major ax.</i>
<i>minor ax.</i>
<i>feature 4</i>
...
<i>feature n</i>

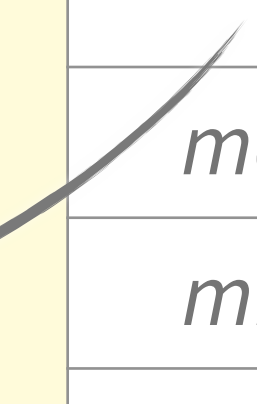
IMAGE
path
rank
w / h

TAXO
taxoID
name
parentID

USERS
name
email
permissions

mandatory
fixed format
internal
free

Allows to compute
concentrations
and **biovolumes**



How?

Step 2: Classify a few images

EcoTaxa 2.6 Project Filtered Demo Zooscan JO low taxo res (0, 1397, 0, 0 / 1397) Jean-Olivier Irisson (log out) Action

Update view & apply filter Select all Display Status All 1000 100% 4 Selected

Taxonomy filter Other filters

- Appendicularia 4
- Copepoda 3
- detritus 7
 - fiber < detritus 3
- t001 0 1397

Recalc. counts Hide empty categories

Image	Classification
1	detritus
2	detritus
3	fiber < detritus
4	Copepoda
5	t001
6	Appendicularia
7	t001
8	t001
9	t001
10	fiber < detritus
11	t001
12	t001
13	t001
14	detritus
15	detritus
16	detritus
17	detritus
18	t001
19	detritus
20	detritus
21	detritus
22	detritus
23	detritus
24	detritus
25	detritus
26	detritus
27	detritus





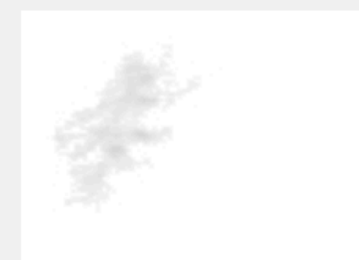









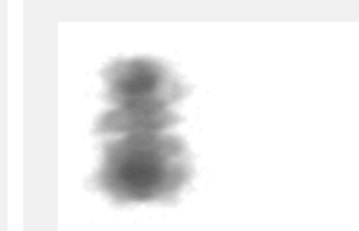


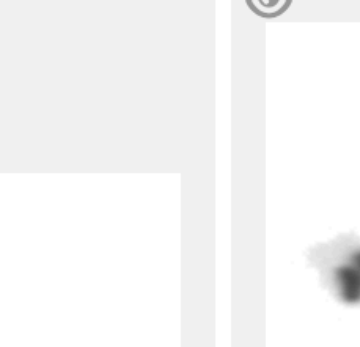

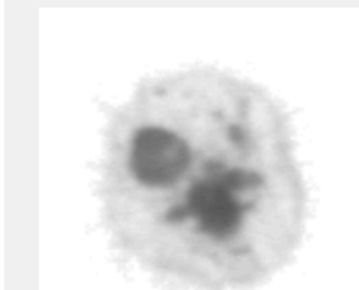





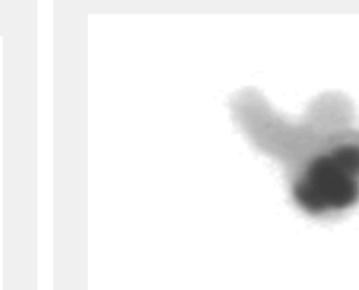

How?

Step 2: Classify a few images

EcoTaxa 2.6 Project Filtered Demo Zooscan JO low taxo res (0, 1397, 0, 0 / 1397) Jean-Olivier Irisson (log out) Action

Update view & apply filter Select all Display Status All 1000 100%

Taxonomy filter Other filters
detritus
Appendicularia 4
Copepoda 3
▼ detritus 12
fiber < detritus 3
t001 0 1397
Recalc. counts Hide empty categories

 1 mm detritus	 1 mm detritus	 1 mm fiber < detritus	 1 mm Copepoda	 1 mm t001	 1 mm Appendicularia	 1 mm t001	 1 mm t001	 1 mm t001
 1 mm fiber < detritus	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm detritus	 1 mm detritus	 1 mm detritus	 1 mm detritus	 1 mm t001
								

How?

Step 3: Train a classifier based on classified images and predict the others

The screenshot displays the EcoTaxa 2.6 web interface. At the top left, the logo 'EcoTaxa 2.6' is visible. The main header shows the project name 'Demo Zooscan JO low taxo res' with a progress indicator '(14, 1383, 0, 0 / 1397)'. On the right, the user 'Jean-Olivier Irisson' is logged in, with a 'log out' link and an 'Action' dropdown menu.

The interface features a left sidebar with a 'Taxonomy filter' section containing a list of categories: 'detritus', 'Appendicularia', 'Copepoda', 'detritus' (highlighted), 'fiber < detritus', and 't001'. Below this is a 'Recalc. counts' button and a 'Hide' button.

A central dropdown menu is open, listing the following actions: 'Train and Predict classifications', 'Import images and metadata', 'Fix category issues', 'Export', 'Edit project settings', 'Extract Subset', 'Merge another project in this project', 'Edit or erase annotations massively', 'Batch edit metadata', 'Reset status to Predicted', and 'Delete objects or project'.

The main area shows a grid of image thumbnails. Each thumbnail includes a '1 mm' scale bar and a classification label below it. The labels are: 'fiber < detritus', 'Copepoda', 'detritus', 'detritus', 't001', 't001', 't001', 't001', 't001', 't001', 'Appendicularia', 'fiber < detritus', 'detritus', and 'detritus'. The labels 'fiber < detritus', 'Copepoda', 'Appendicularia', and the second 'detritus' are highlighted in green, while the others are in light blue.

How?

Step 3: Train a classifier based on classified images and predict the others

PREDICTION: Choice of Learning Set data source

Next: Choose objects in selected projects

A Learning Set AKA [Training Data Set](#) is built using validated data from one or several projects.

Inside these projects, you will be able, in next pages, to specify:

- Which categories you're interested in predicting.
- How many objects you want to use as reference for the prediction.
- For these objects, which features are relevant to the prediction.

This Learning set will then be used for running the prediction task. ?

Filter on title : # Matching features \geq 10 Instrument : Zooscan x Filter

Project deep features model: LOKI_2022-05-17.

Model is **usable**. ?

Inst.	# - Title	# Validated	# Matching features	Deep features model
<input checked="" type="checkbox"/>	Zooscan #859 - Demo Zooscan JO low taxo res	0	69	LOKI_2022-05-17
<input type="checkbox"/>	Zooscan #2374 - North Inlet Zooplankton	1893472	69	
<input type="checkbox"/>	Zooscan #1040 - _Zooscan_Reference	1444921	69	zooscan
<input type="checkbox"/>	Zooscan #6212 - ZOOSCAN PELGAS_2004_2016_NGrandremy_PhD_thesis	1153507	69	zooscan
<input type="checkbox"/>	Zooscan #6729 - ZOOSCAN PELGAS Pour Laetitia	477713	69	zooscan
<input type="checkbox"/>	Zooscan #2291 - TRAFFIC_M153_Multinet_Maxi	465089	69	zooscan
<input type="checkbox"/>	Zooscan #5421 - Gradients_v2	428099	69	zooscan
<input type="checkbox"/>	Zooscan #714 - Zooscan Tara Oceans MTN 300 ALL NETS - NEW!	393382	69	zooscan
<input type="checkbox"/>	Zooscan #6285 - Zooplankton Helgoland - Normalnetz 150 2006	382243	69	zooscan
<input type="checkbox"/>	Zooscan #1125 - ICaWR	350447	69	
<input type="checkbox"/>	Zooscan #4565 - zooscan_cecilia_2020	323013	69	
<input type="checkbox"/>	Zooscan #5977 - Zooscan_SO21_MN_0-1000m	313351	69	zooscan
<input type="checkbox"/>	Zooscan #611 - Projet DVM Lacs Boreaux - Sabrina Gignac Brassard 2016-2017	304151	69	
<input type="checkbox"/>	Zooscan #3657 - Zooscan Triatlas 2019 MTN 200	285458	69	zooscan
<input type="checkbox"/>	Zooscan #6111 - Triatlas 2019 MTN 200 NEW	285458	69	zooscan
<input type="checkbox"/>	Zooscan #418 - Zooscan Tara Oceans 2009 2013 MTN 300 sn033	284105	69	zooscan

How?

Step 3: Train a classifier based on classified images and predict the others

PREDICTION: Choice of Learning Set categories and size

Next: Choose features in selected objects

From data source, which is : #859 - Demo Zooscan JO low taxo res, only objects validated in below chosen categories will be present in the Learning Set being built.

Optionally, each category can **appear** as another category, generally a parent one, to the machine learning algorithm. ?

- The experience shows that it is often more efficient to predict into a limited number of categories and then validate in detail using more categories.

Learn from max. objects per category. ?
Total is currently 14 objects.

Make categories appear like in:

▼

(id)	☐	Source (validated) category	# source	% source	# learning set
(84963)	☑	detritus	6	42.9 %	6
(25828)	☑	Copepoda	3	21.4 %	3
(85076)	☑	fiber < detritus	3	21.4 %	3
(85123)	☑	Appendicularia	2	14.3 %	2

Appear as category

▼

▼

▼

▼

How?

Step 3: Train a classifier based on classified images and predict the others

PREDICTION: Choice of features and settings

Start prediction task

Add deep features  

You have chosen 14 reference objects to build the Learning Set. In this last step, you can choose which features to associate with each of these objects, and start a prediction task using the Learning Set.

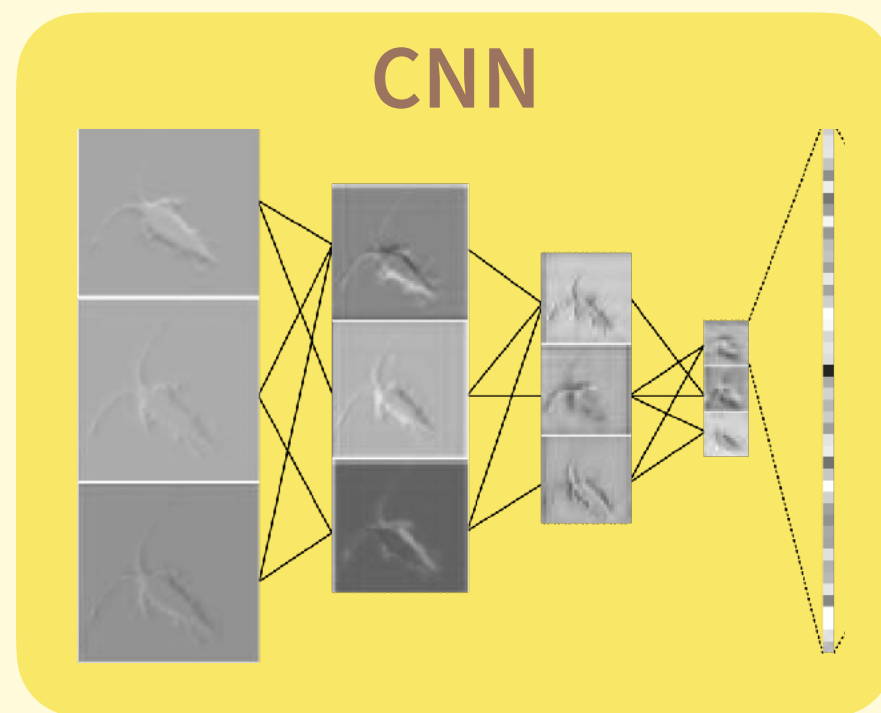
- Prediction will be better if you exclude features which are not related to the classification, e.g. coordinates in the raw image.
- Features with a single, constant value, or too many missing values, are useless for prediction and are automatically excluded. Some of them are listed here as a reminder.
- Missing values will be replaced by the median value for this feature from the reference objects.
- Prediction settings are recorded in EcoTaxa for the next prediction.



<input checked="" type="checkbox"/> %area	<input checked="" type="checkbox"/> angle	<input checked="" type="checkbox"/> area	<input checked="" type="checkbox"/> area_exc	<input checked="" type="checkbox"/> bx	<input checked="" type="checkbox"/> by	<input checked="" type="checkbox"/> cdexc	<input checked="" type="checkbox"/> centroids	<input checked="" type="checkbox"/> circ.	<input checked="" type="checkbox"/> circex	<input type="checkbox"/> compentropy	<input type="checkbox"/> compm1
<input type="checkbox"/> compm2	<input type="checkbox"/> compm3	<input type="checkbox"/> compmean	<input type="checkbox"/> compslope	<input checked="" type="checkbox"/> convarea	<input checked="" type="checkbox"/> convperim	<input checked="" type="checkbox"/> cv	<input type="checkbox"/> depth_max	<input type="checkbox"/> depth_min	<input checked="" type="checkbox"/> elongation	<input checked="" type="checkbox"/> esd	<input checked="" type="checkbox"/> fcons
<input checked="" type="checkbox"/> feret	<input checked="" type="checkbox"/> feretareaexc	<input checked="" type="checkbox"/> fractal	<input checked="" type="checkbox"/> height	<input checked="" type="checkbox"/> histcum1	<input checked="" type="checkbox"/> histcum2	<input checked="" type="checkbox"/> histcum3	<input checked="" type="checkbox"/> intden	<input checked="" type="checkbox"/> kurt	<input type="checkbox"/> lat_end	<input type="checkbox"/> lon_end	<input checked="" type="checkbox"/> major
<input checked="" type="checkbox"/> max	<input checked="" type="checkbox"/> mean	<input checked="" type="checkbox"/> meanpos	<input checked="" type="checkbox"/> median	<input checked="" type="checkbox"/> min	<input checked="" type="checkbox"/> minor	<input checked="" type="checkbox"/> mode	<input checked="" type="checkbox"/> nb1	<input checked="" type="checkbox"/> nb2	<input type="checkbox"/> nb3	<input checked="" type="checkbox"/> perim.	<input checked="" type="checkbox"/> perimareaexc
<input checked="" type="checkbox"/> perimferet	<input checked="" type="checkbox"/> perimmajor	<input checked="" type="checkbox"/> range	<input checked="" type="checkbox"/> skelarea	<input checked="" type="checkbox"/> skew	<input checked="" type="checkbox"/> slope	<input checked="" type="checkbox"/> sr	<input checked="" type="checkbox"/> stddev	<input checked="" type="checkbox"/> symetrieh	<input checked="" type="checkbox"/> symetriehc	<input checked="" type="checkbox"/> symetriev	<input checked="" type="checkbox"/> symetrievc
<input type="checkbox"/> tag	<input checked="" type="checkbox"/> thickr	<input checked="" type="checkbox"/> width	<input checked="" type="checkbox"/> x	<input checked="" type="checkbox"/> xm	<input type="checkbox"/> xmg5	<input checked="" type="checkbox"/> xstart	<input checked="" type="checkbox"/> y	<input checked="" type="checkbox"/> ym	<input type="checkbox"/> ymg5	<input checked="" type="checkbox"/> ystart	

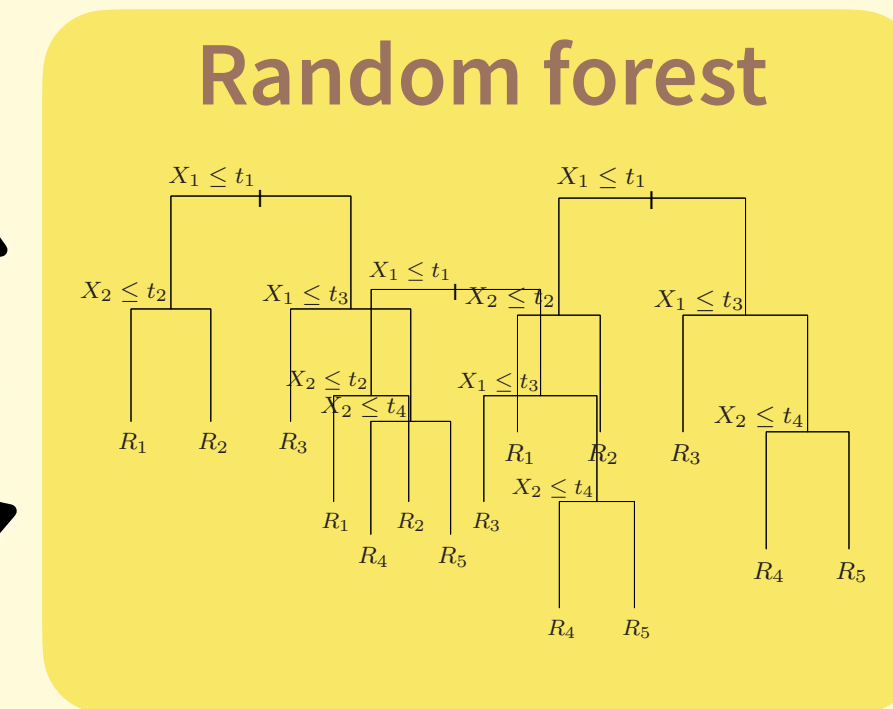
A note on the machine learning architecture

IMAGE
path
rank
w / h



OBJECT
objID
taxoID
area
major ax.
minor ax.
feature 4
...
feature n

CNN
CNN feature 1
...
CNN feature n



taxoID

score

How?

Step 3: Train a classifier based on classified images and predict the others

25% Retrieving objects to classify

ⓘ WARNING : this is an asynchronous process.

You can either:

- Open another tab or window to continue working in EcoTaxa (and monitor the progress of the task switching to this page from time to time)
- Leave this page to continue working in EcoTaxa (and click on the "Run" or "Done" buttons on the upper right of the Ecotaxa topper to monitor the progress of the task or check its completion)

Asynchronous tasks are utilized for processes that can be long. They can create temporary data (like export file) that can be manually removed at the end of the process even if EcoTaxa will do it for you on a regular schedule. Remaining tasks can be checked using the "Done" or "Error" top right buttons.

How?

Step 4: Sort predicted images by classification score

EcoTaxa 2.6 Project Filtered Filter: Taxo= Copepoda Done :1
Jean-Olivier Irisson (log out)

Update view & apply filter Select all Display Status All 1000 100%

Taxonomy filter Other filters

- Appendicularia 2 112
- Copepoda 3 209**
- detritus 6 904
 - fiber < detritus 3 158

Recalc. counts Hide empty categories

1 mm 1 mm 1 mm 1 mm 1 mm 1 mm

Copepoda Copepoda Copepoda Copepoda Copepoda Copepoda

How?

Step 4: Sort predicted images by classification score





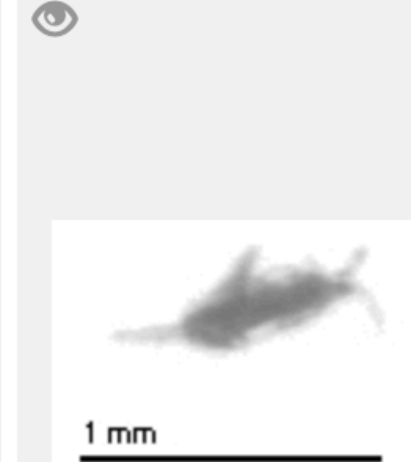


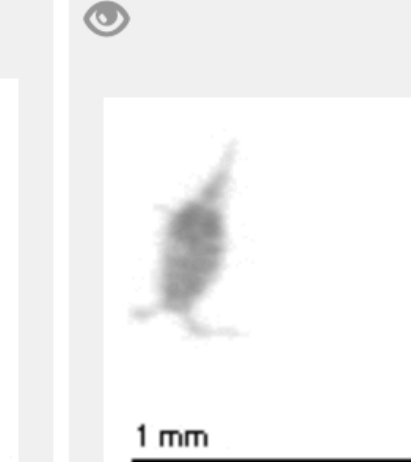

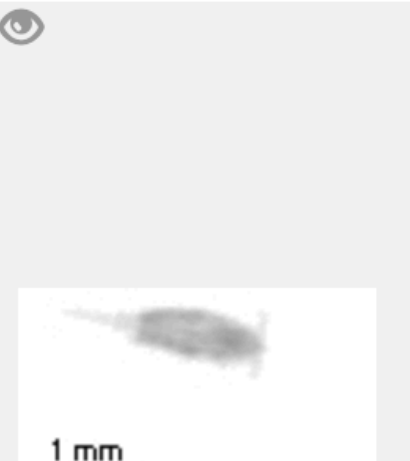

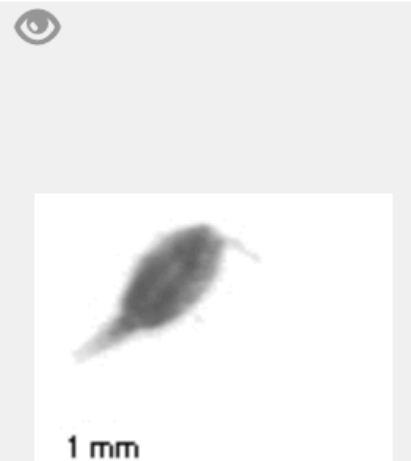
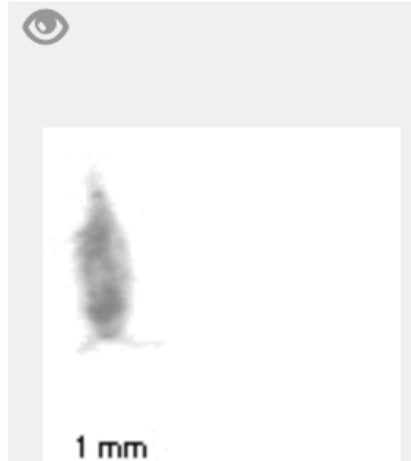


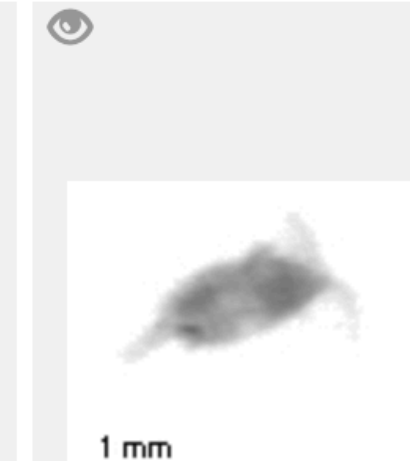




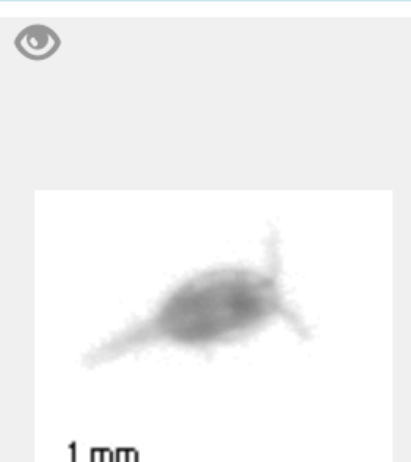
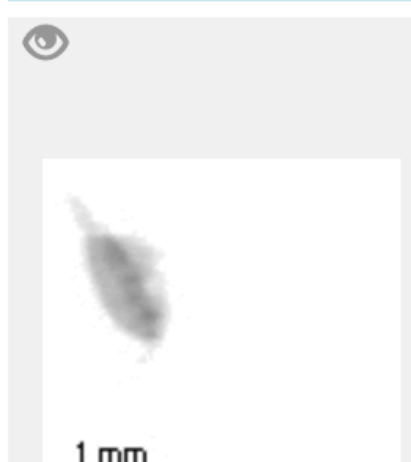
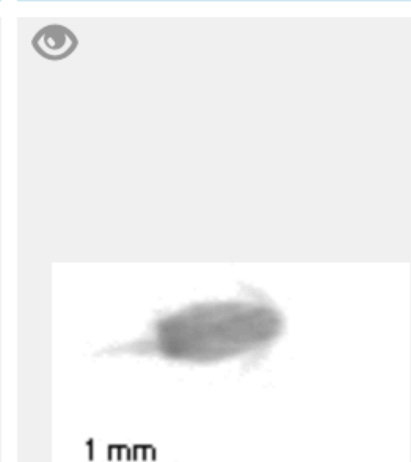
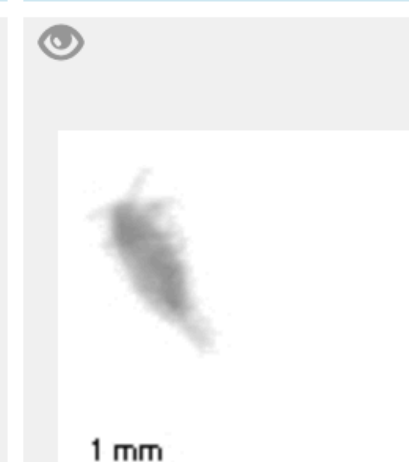
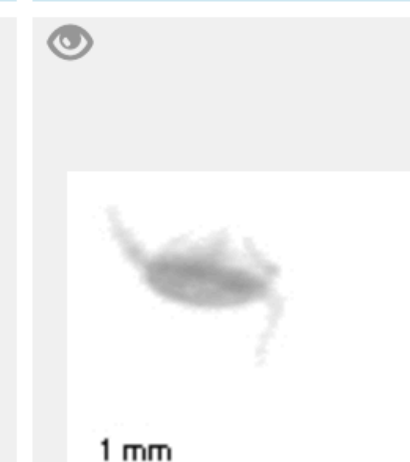


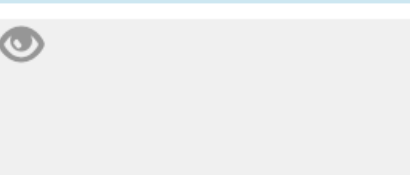
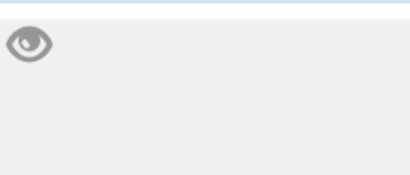
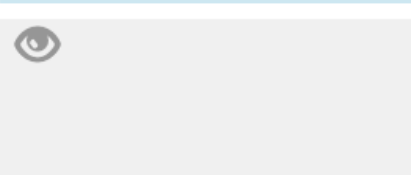

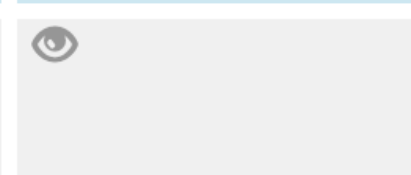
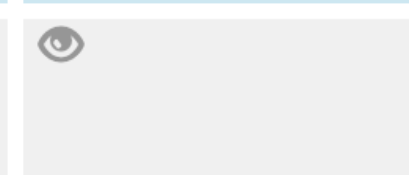
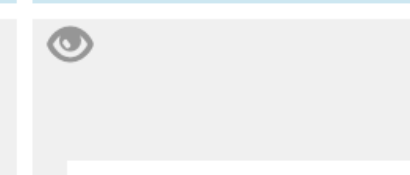
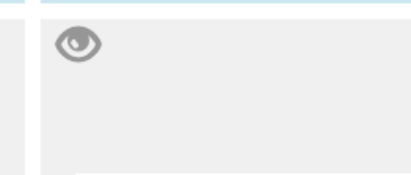
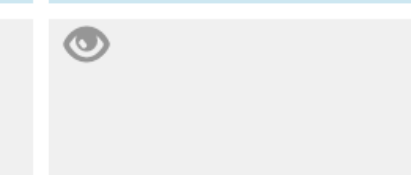
EcoTaxa 2.6 Project Filtered Filter: Taxo= Copepoda Done :1 Jean-Olivier Irsson (log out) Action

Update view & apply filter Select all Score Display Status All 1000 100% 100

Taxonomy filter **Other filters**

- Appendicularia 2 112
- Copepoda 3 209**
- detritus 6 904
- fiber < detritus 3 158

Recalc. counts Hide empty categories

 1 mm Copepoda Score : -	 1 mm Copepoda Score : -	 1 mm Copepoda Score : -	 1 mm Copepoda Score : 0.30	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29
 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29
 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28
								

How?

Step 5: Confirm or correct automatic classifications




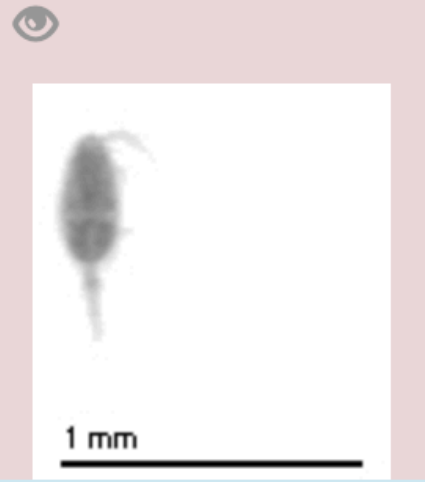



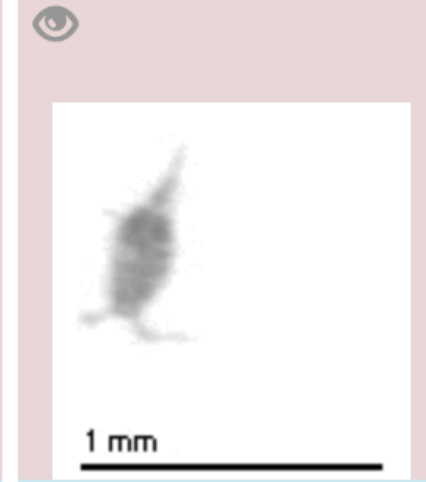

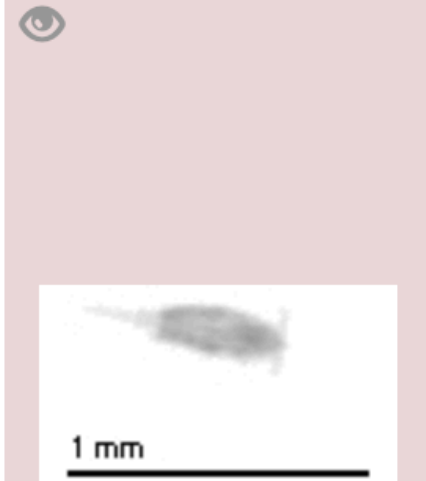

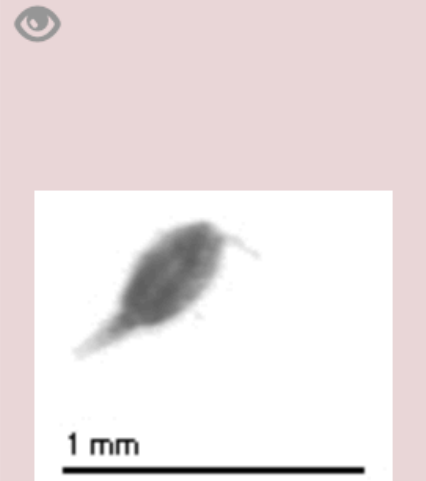









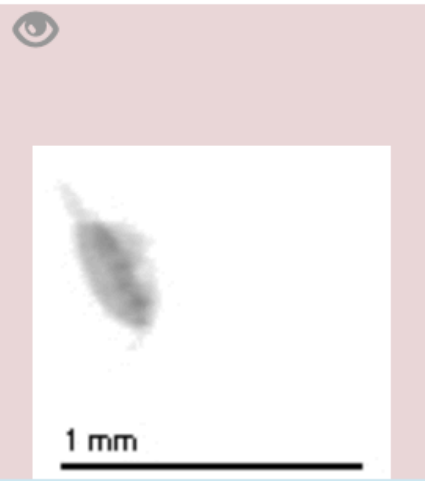





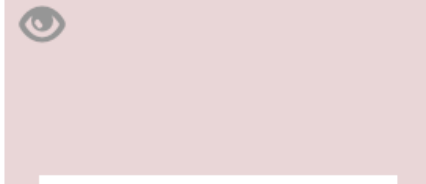

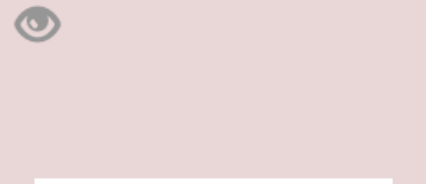






EcoTaxa 2.6 Project Filtered Filter: Taxo= Copepoda Done :1
Jean-Olivier Irisson (log out)

Update view & apply filter Select all Score Display Status All 1000 100% 100 109 Selected

Taxonomy filter Other filters

- Appendicularia 2 112
- Copepoda 3 209**
- detritus 6 904
- fiber < detritus 3 158

Recalc. counts Hide empty categories

 1 mm Copepoda Score : -	 1 mm Copepoda Score : -	 1 mm Copepoda Score : -	 1 mm Copepoda Score : 0.30	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29
 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.29
 1 mm Copepoda Score : 0.29	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28	 1 mm Copepoda Score : 0.28
								

How?

Step 5: Confirm or correct automatic classifications





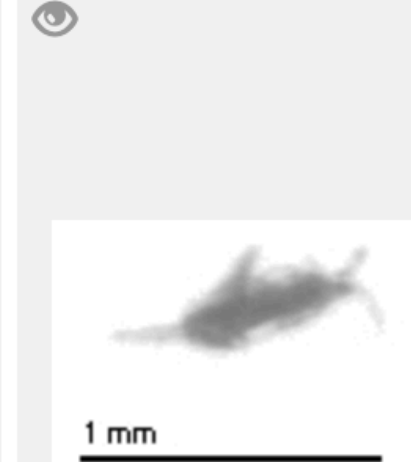


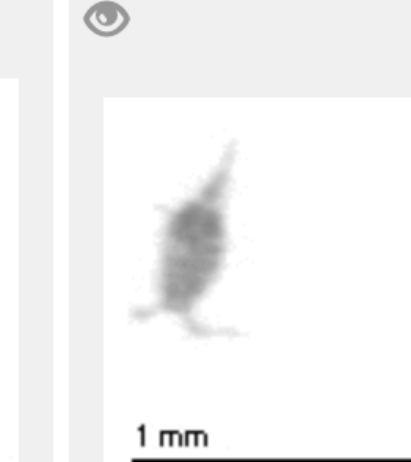

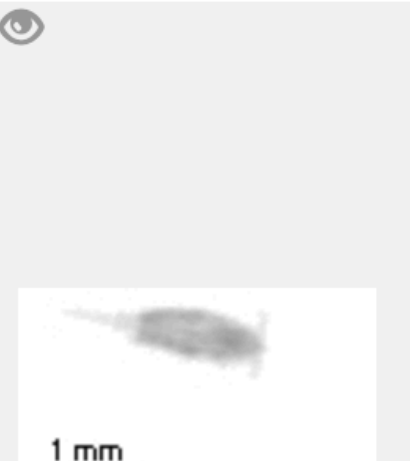

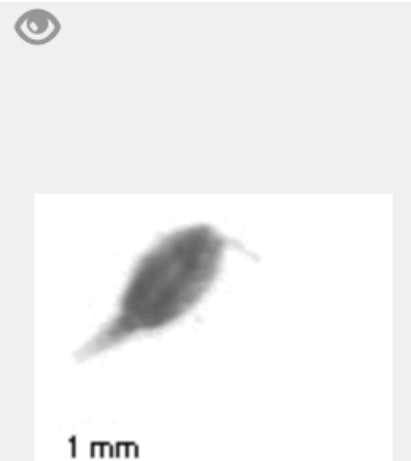
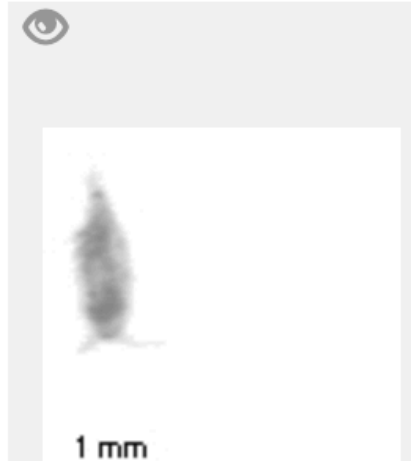


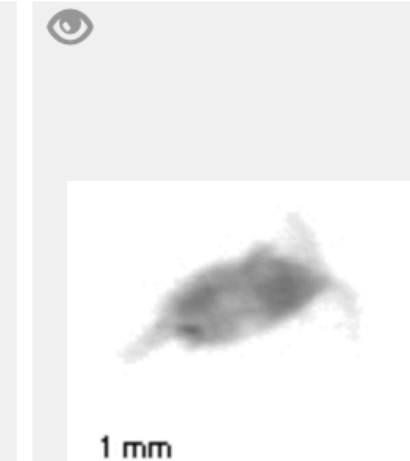




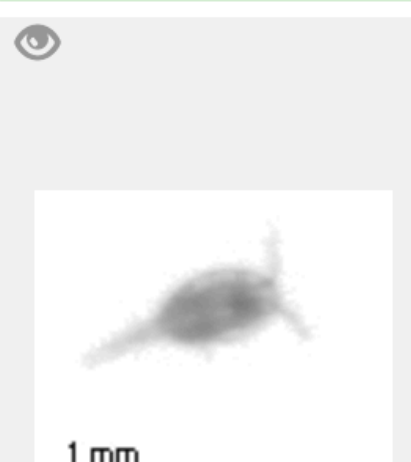
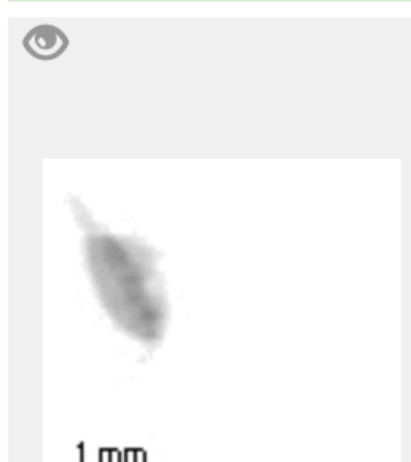
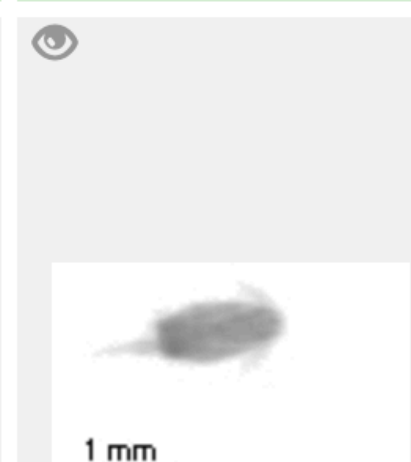
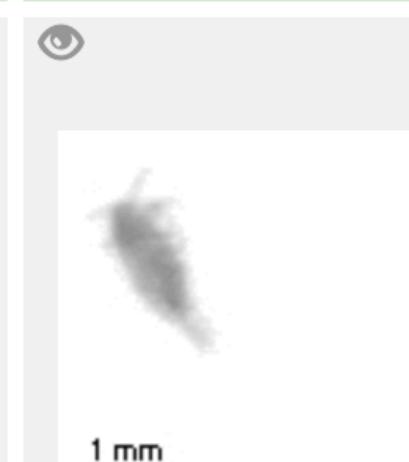
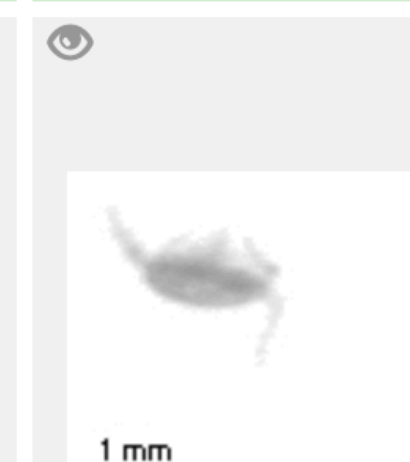


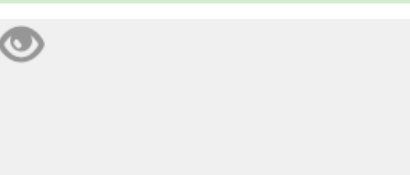
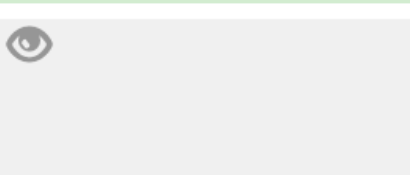
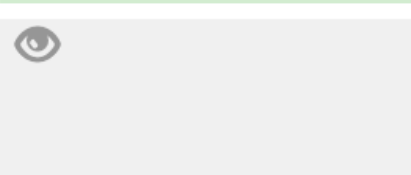
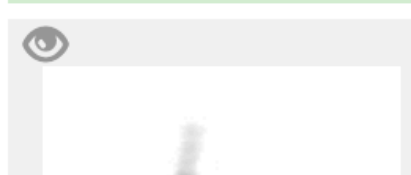
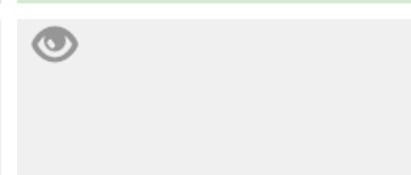
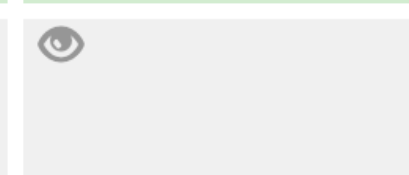
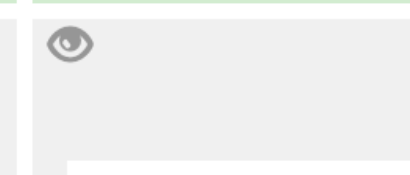
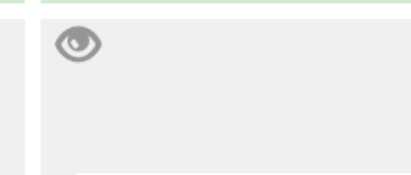
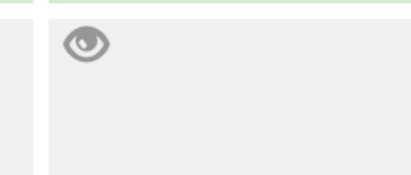
EcoTaxa 2.6 Project Filtered Filter: Taxo= Copepoda Done :1
Jean-Olivier Irisson (log out) Action

Update view & apply filter Select all Score Display Status All 1000 100% 116 Selected

Taxonomy filter Other filters

- Appendicularia 2 112
- Copepoda 119 93**
- detritus 6 904
- fiber < detritus 3 158

Recalc. counts Hide empty categories

 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-

How?

Step 5: Confirm or correct automatic classifications

EcoTaxa 2.6 Project Filtered Filter: Taxo= fiber X Done :1
Jean-Olivier Irsson (log out) Action

Update view & apply filter Select all Score Display Status All 1000 100%

Taxonomy filter Other filters
Copepoda
Appendicularia 18 94
Copepoda 122 93
detritus 64 846
fiber < detritus 158
Recalc. counts Hide empty categories

fiber < detritus Score: -
fiber < detritus Score: -
fiber < detritus Score: -
fiber < detritus Score: 0.30
fiber < detritus Score: 0.30
fiber < detritus Score: 0.30
fiber < detritus Score: 0.29

fiber < detritus Score: 0.29
fiber < detritus Score: 0.29
fiber < detritus Score: 0.29
fiber < detritus Score: 0.28
fiber < detritus Score: 0.28
fiber < detritus Score: 0.28
Copepoda Score: 0.28
fiber < detritus Score: 0.28

fiber < detritus Score: 0.28
fiber < detritus Score: 0.28
fiber < detritus Score: 0.28
fiber < detritus Score: 0.28
Copepoda Score: 0.28
fiber < detritus Score: 0.28

How?

Rinse and repeat!

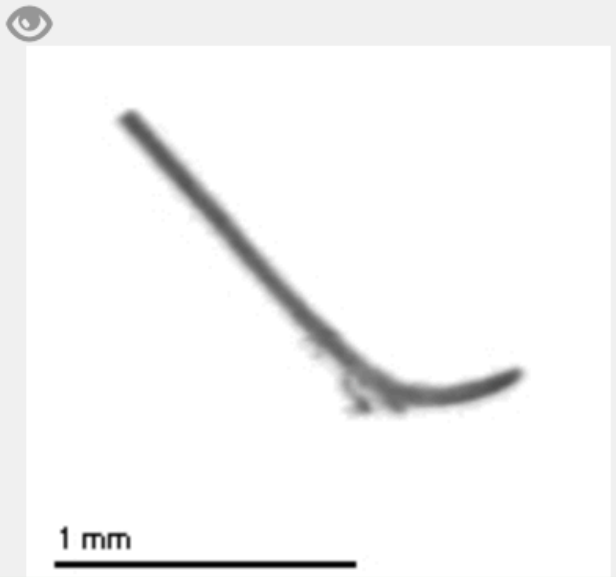

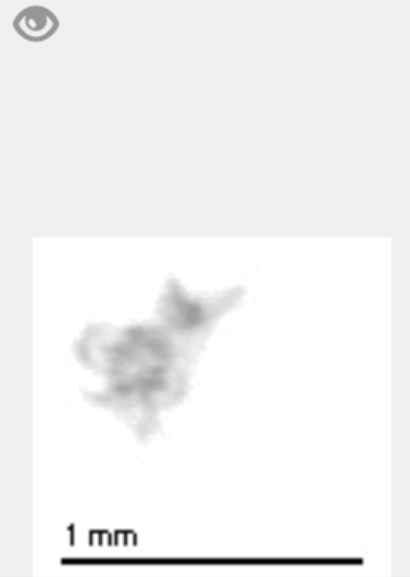
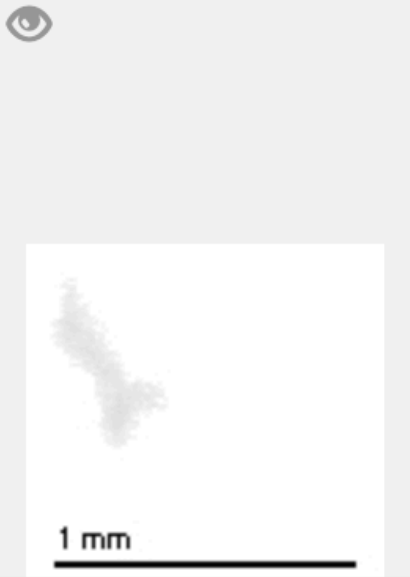
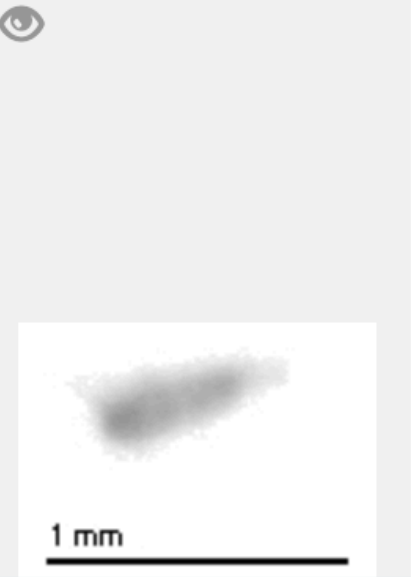
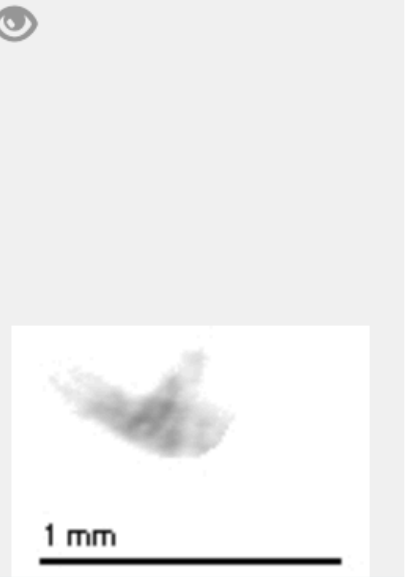
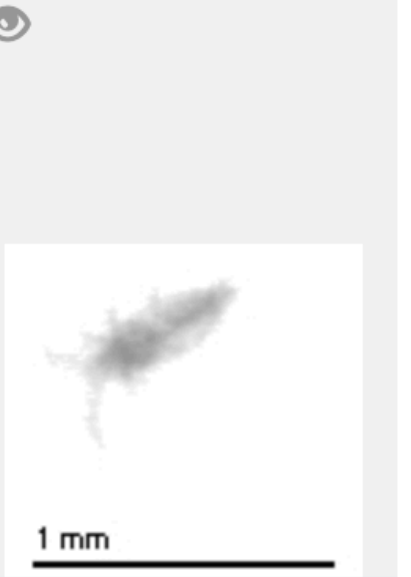
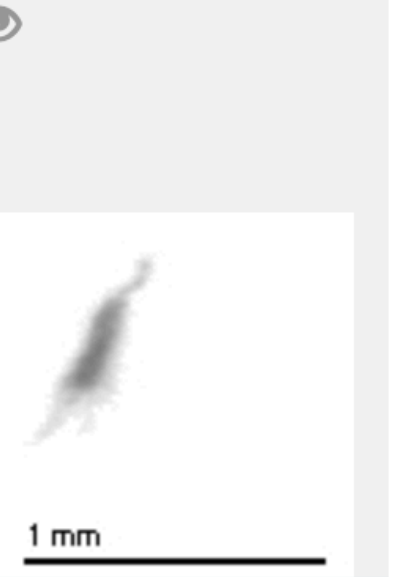
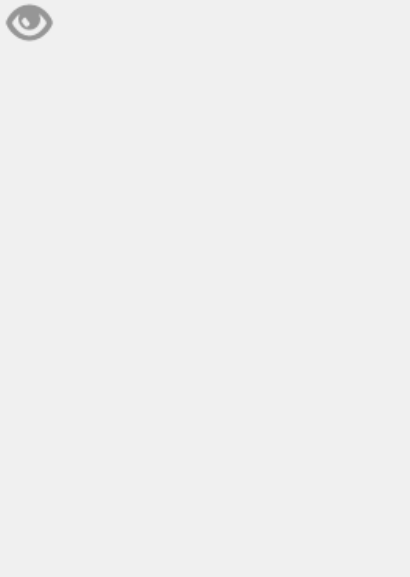
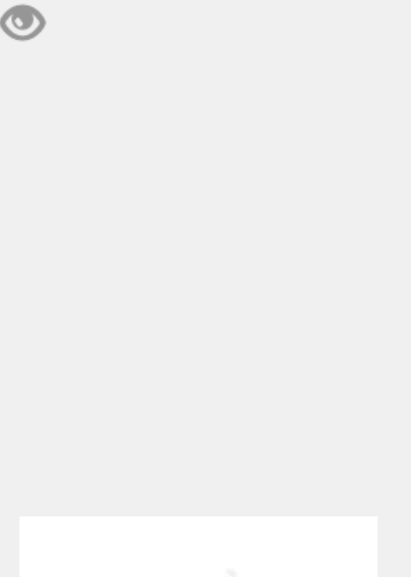
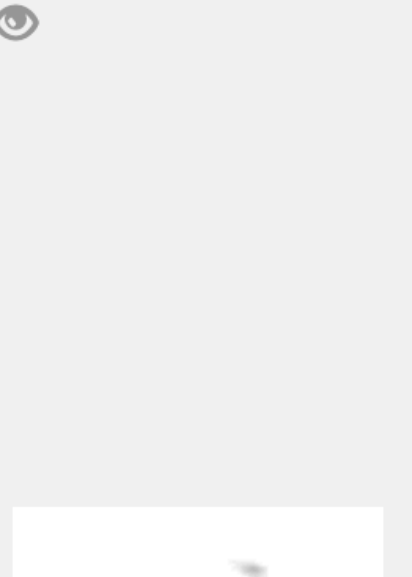
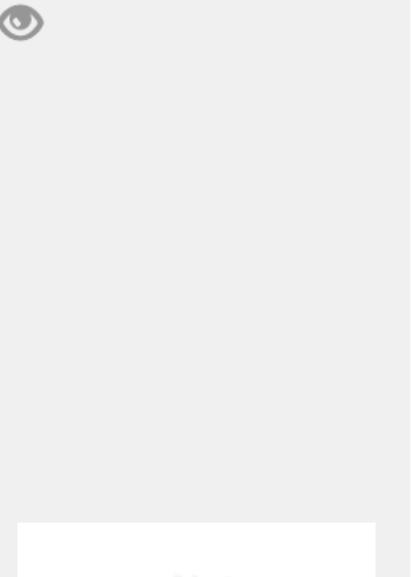

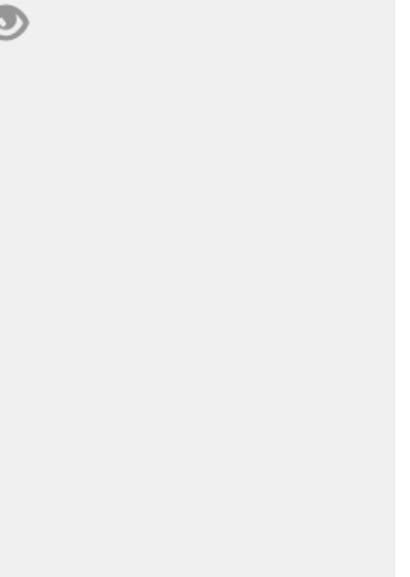
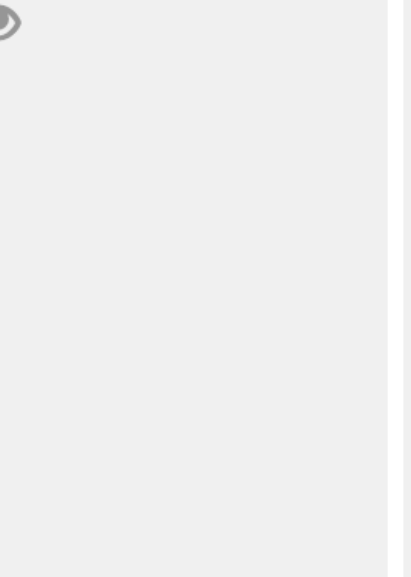
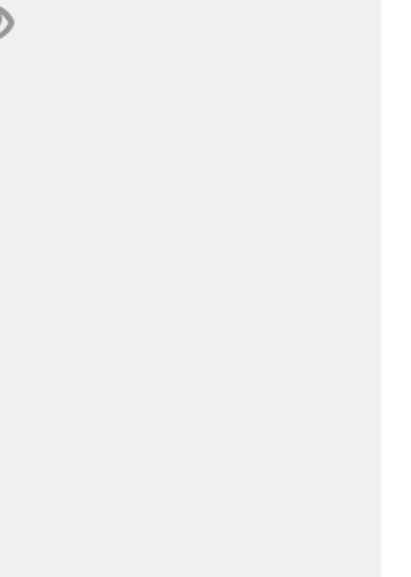
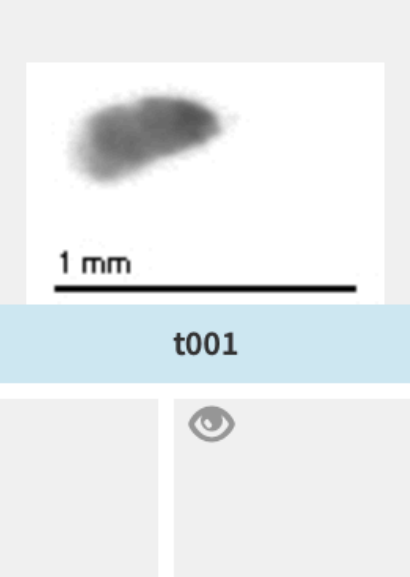
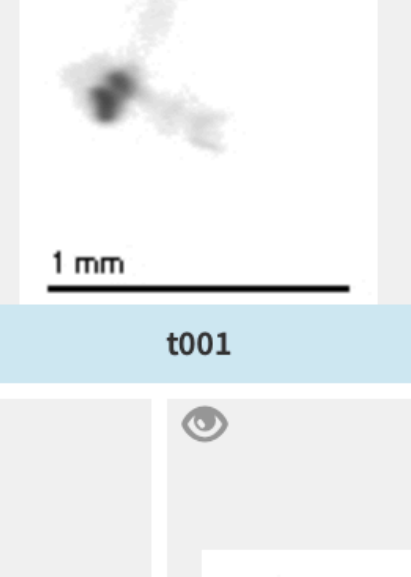
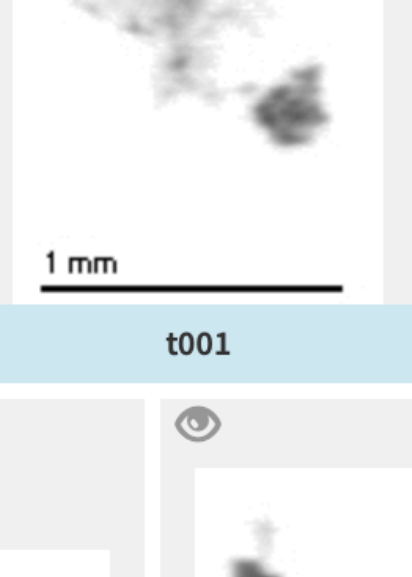
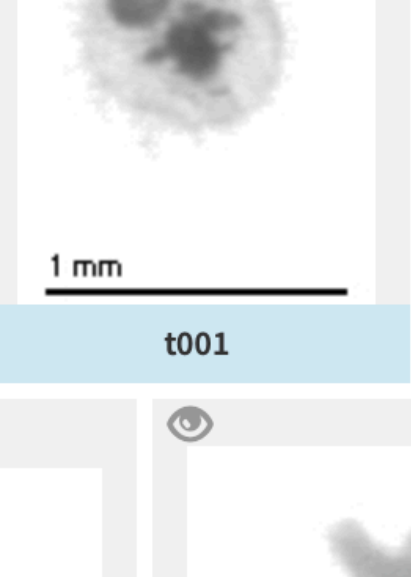
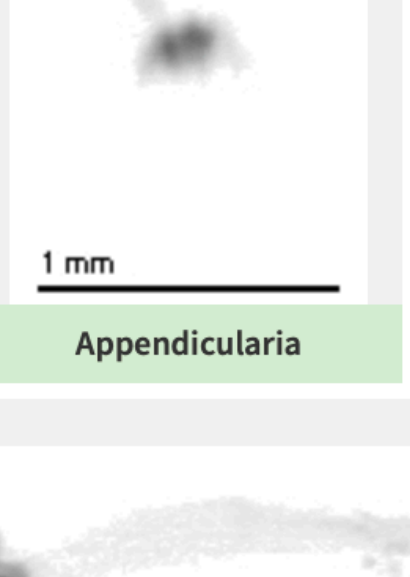
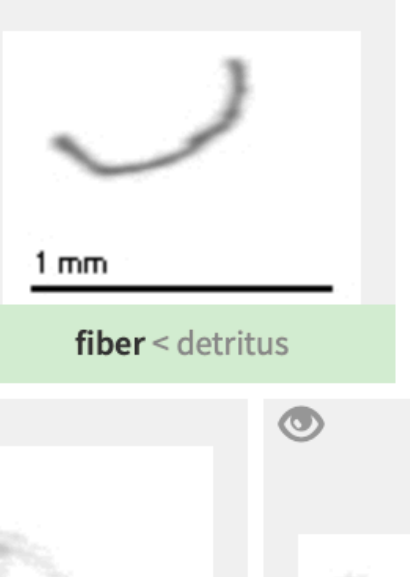
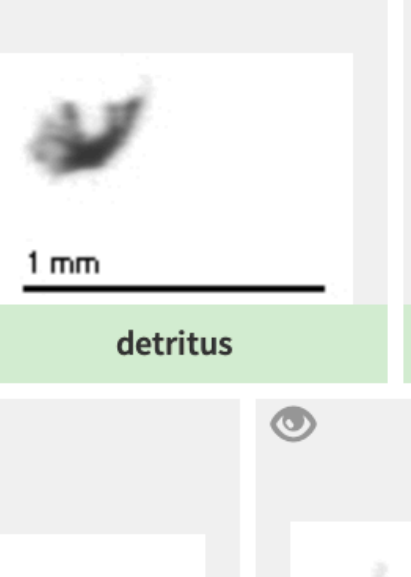
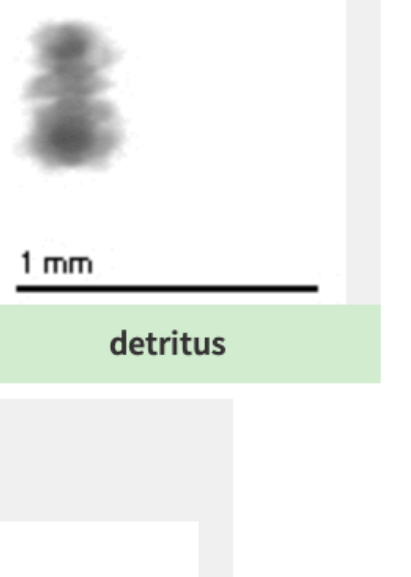
EcoTaxa 2.6 Demo Zooscan JO low taxo res (14, 1383, 0, 0 / 1397) Jean-Olivier Irisson (log out)

Project ▼ Filtered ▼ Action ▼

Update view & Display Status All 1000 100% 100 Q Q Q Q

Taxonomy filter ⓘ
detritus
Appendicularia
Copepoda
▼ **detritus**
fiber < detritus
t001
Recalc. counts Hide

- Train and Predict classifications
- Import images and metadata
- Fix category issues
- Export
- Edit project settings
- Extract Subset
- Merge another project in this project
- Edit or erase annotations massively
- Batch edit metadata
- Reset status to Predicted
- Delete objects or project

 1 mm fiber < detritus	 1 mm Copepoda	 1 mm detritus	 1 mm detritus	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001
 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm Appendicularia	 1 mm fiber < detritus	 1 mm detritus	 1 mm detritus
 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001	 1 mm t001

How?

Rinse and repeat!




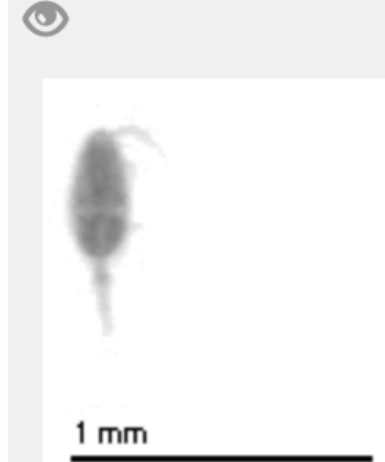
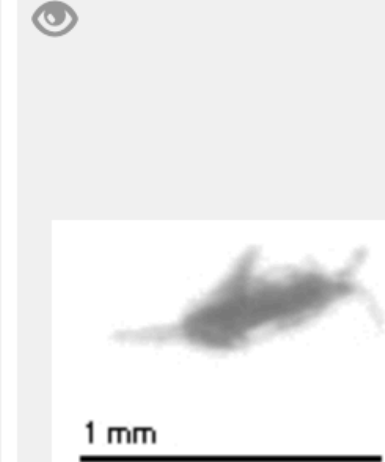


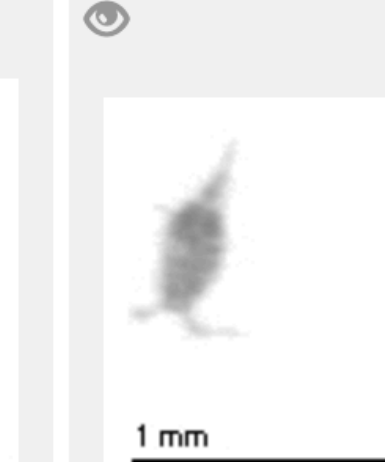

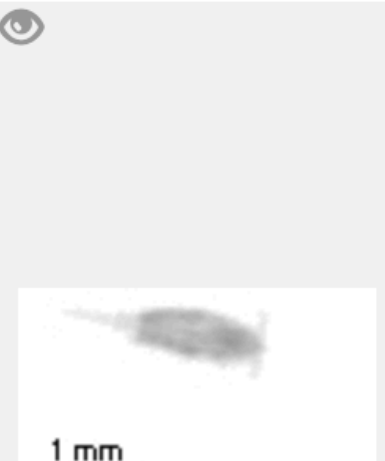

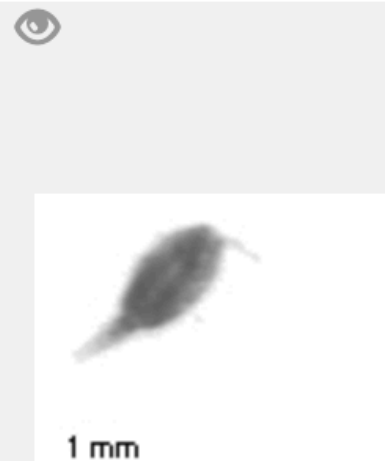
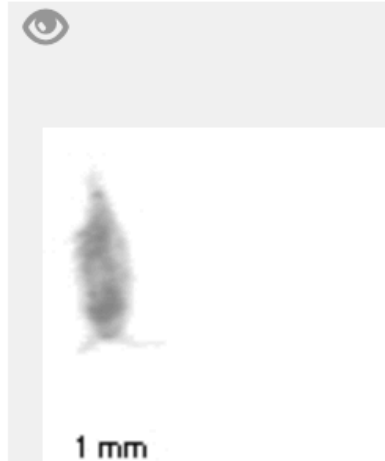
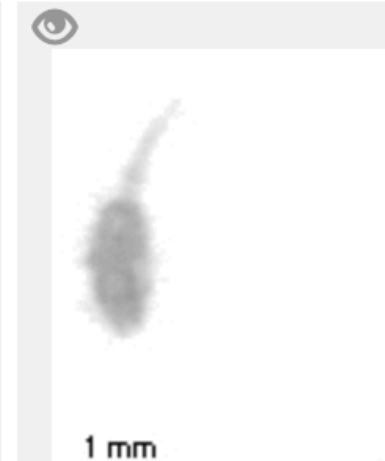

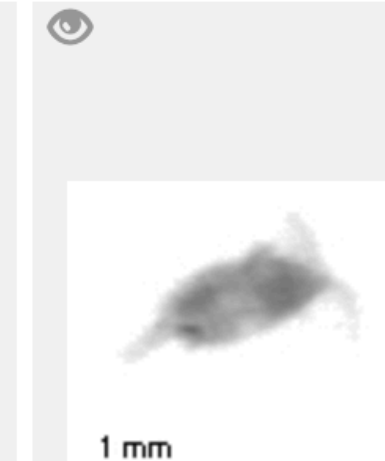



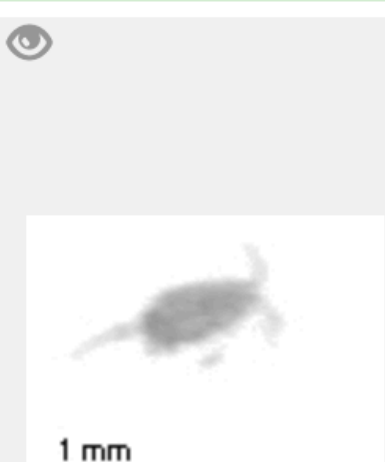

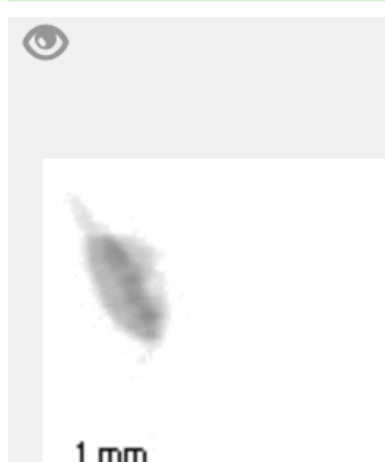
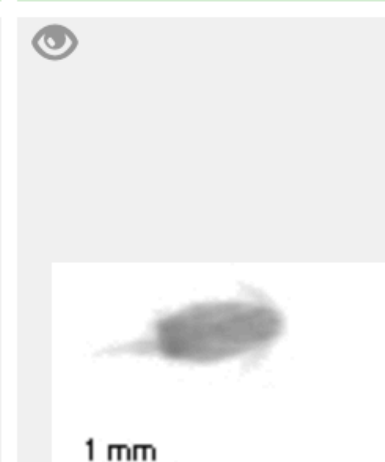
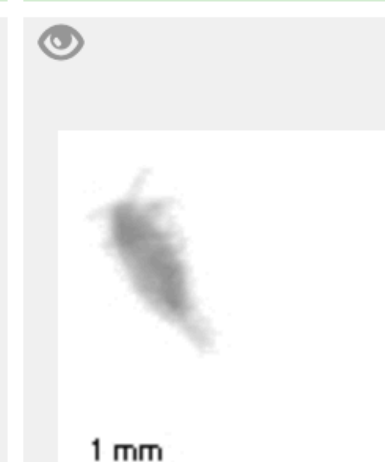
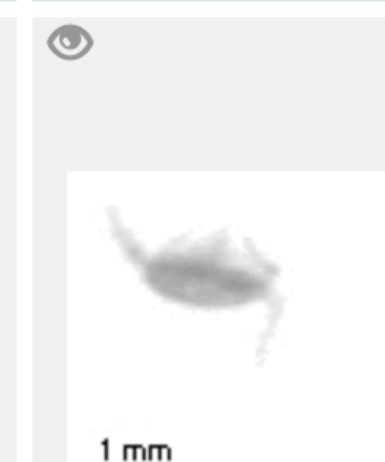


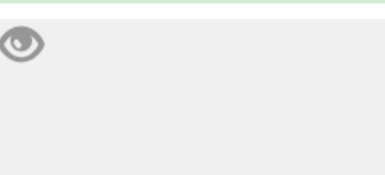
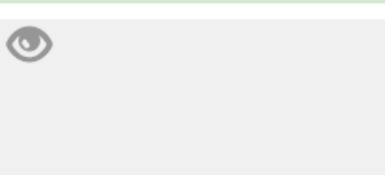
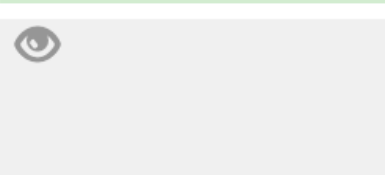

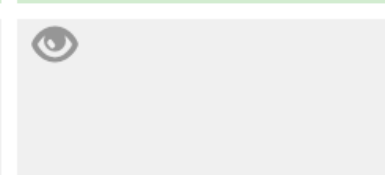
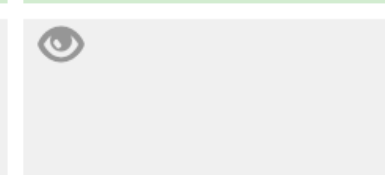
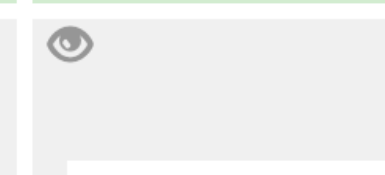
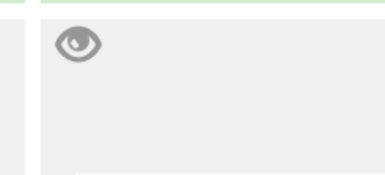
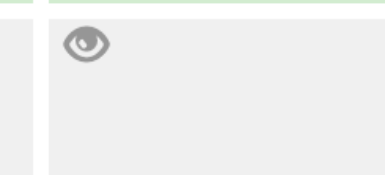
EcoTaxa 2.6 Project Filtered Filter: Taxo= Copepoda Done :1
Jean-Olivier Irisson (log out)

Update view & apply filter Select all Score Display Status All 1000 100% 116 Selected

Taxonomy filter Other filters

- Appendicularia 2 112
- Copepoda 119 93**
- detritus 6 904
 - fiber < detritus 3 158

Recalc. counts Hide empty categories

 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-
 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-	 1 mm Copepoda Score :-

EcoTaxa today

242M objects

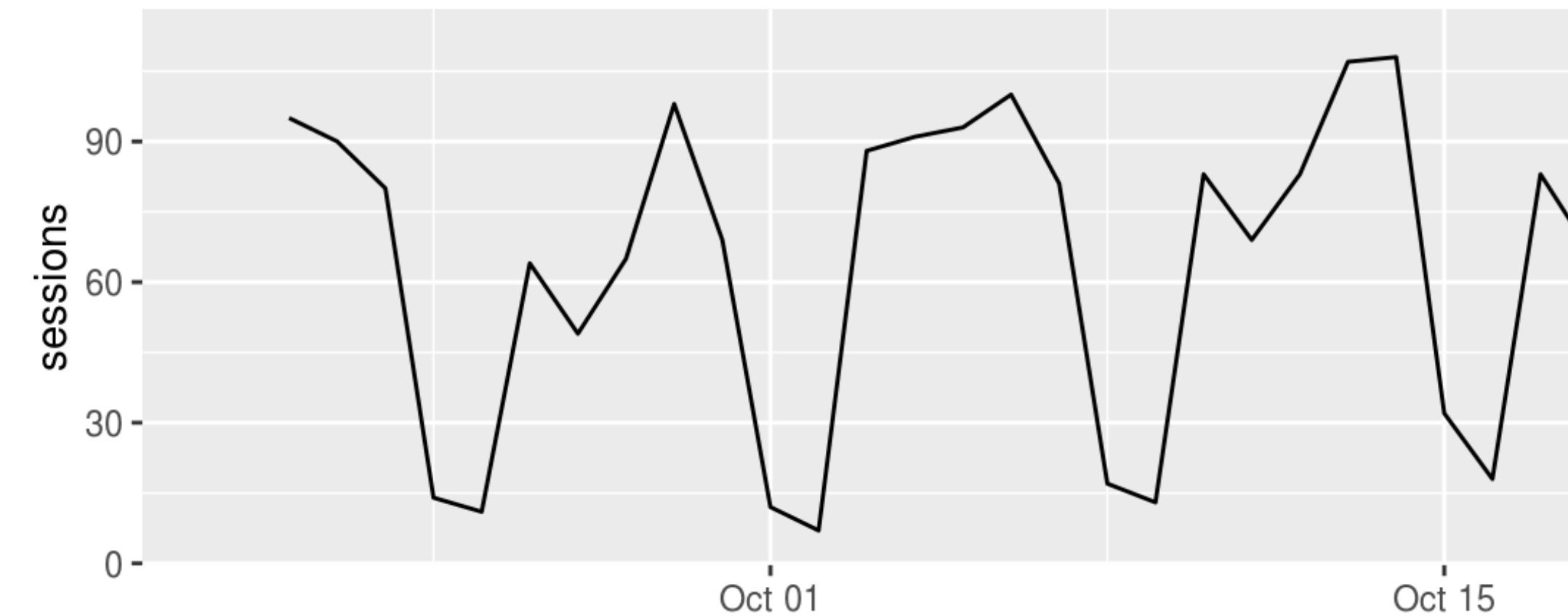
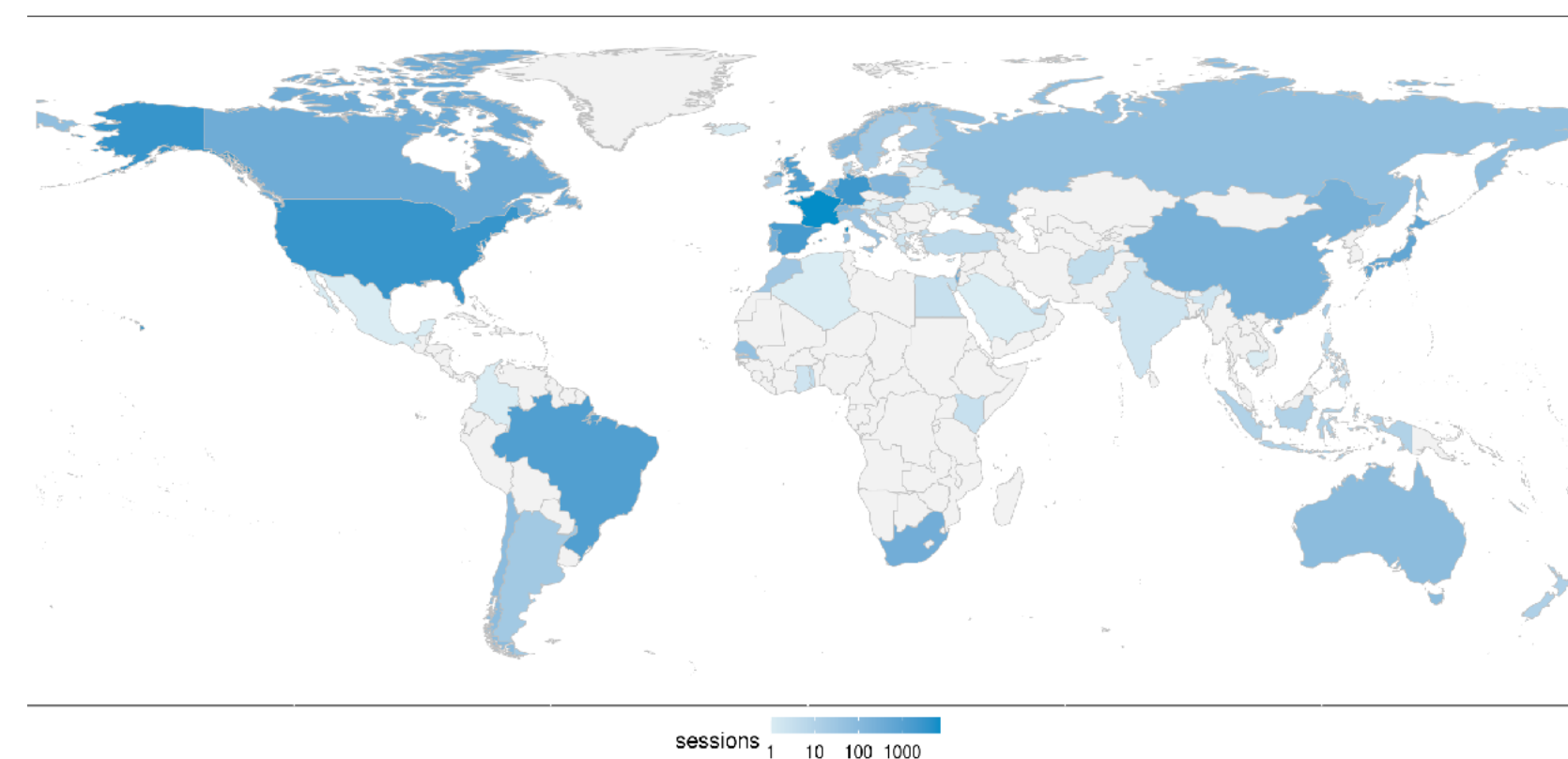
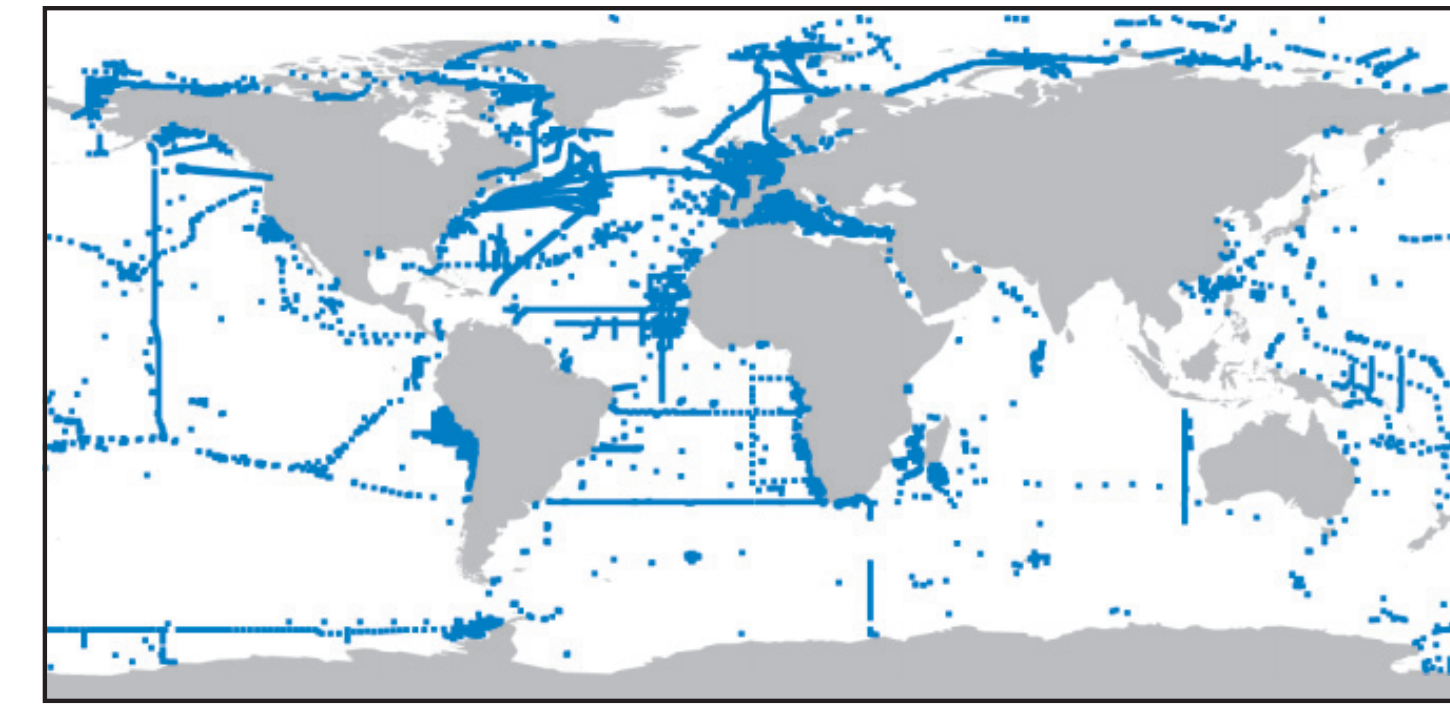
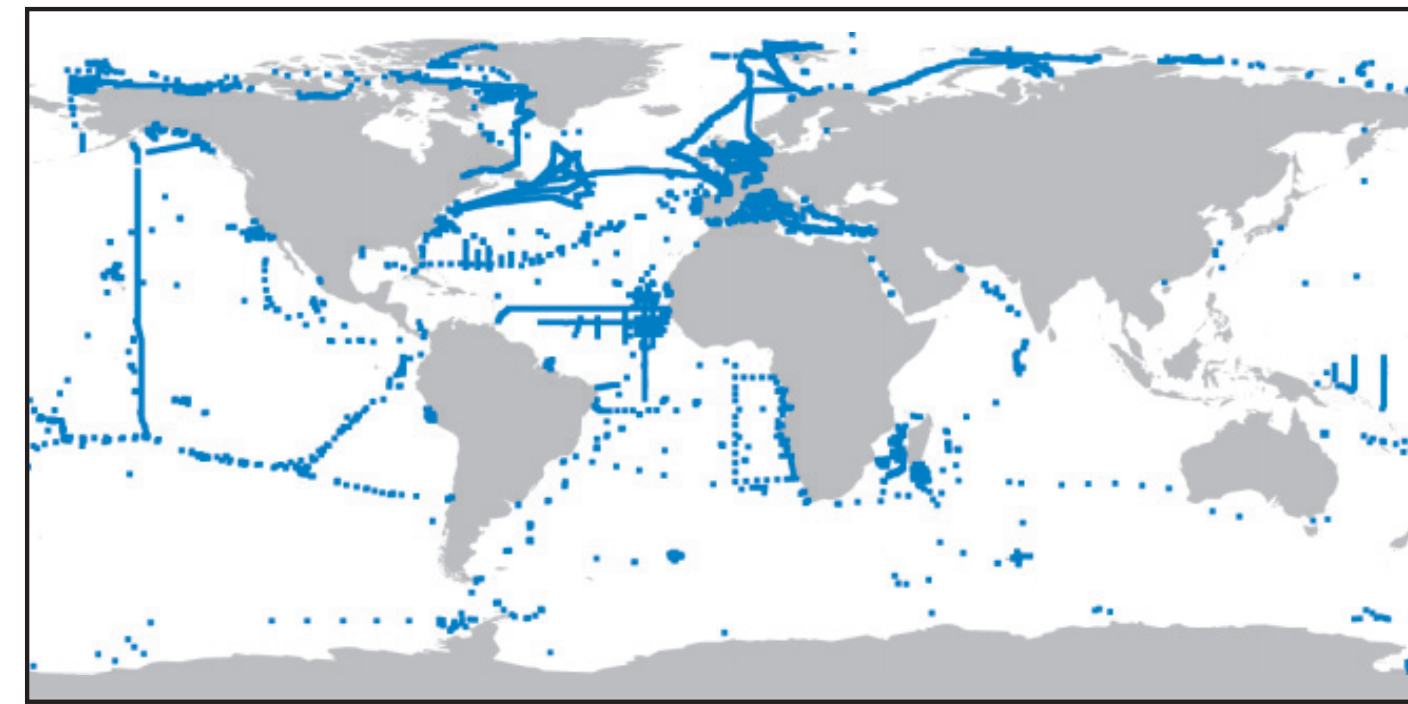
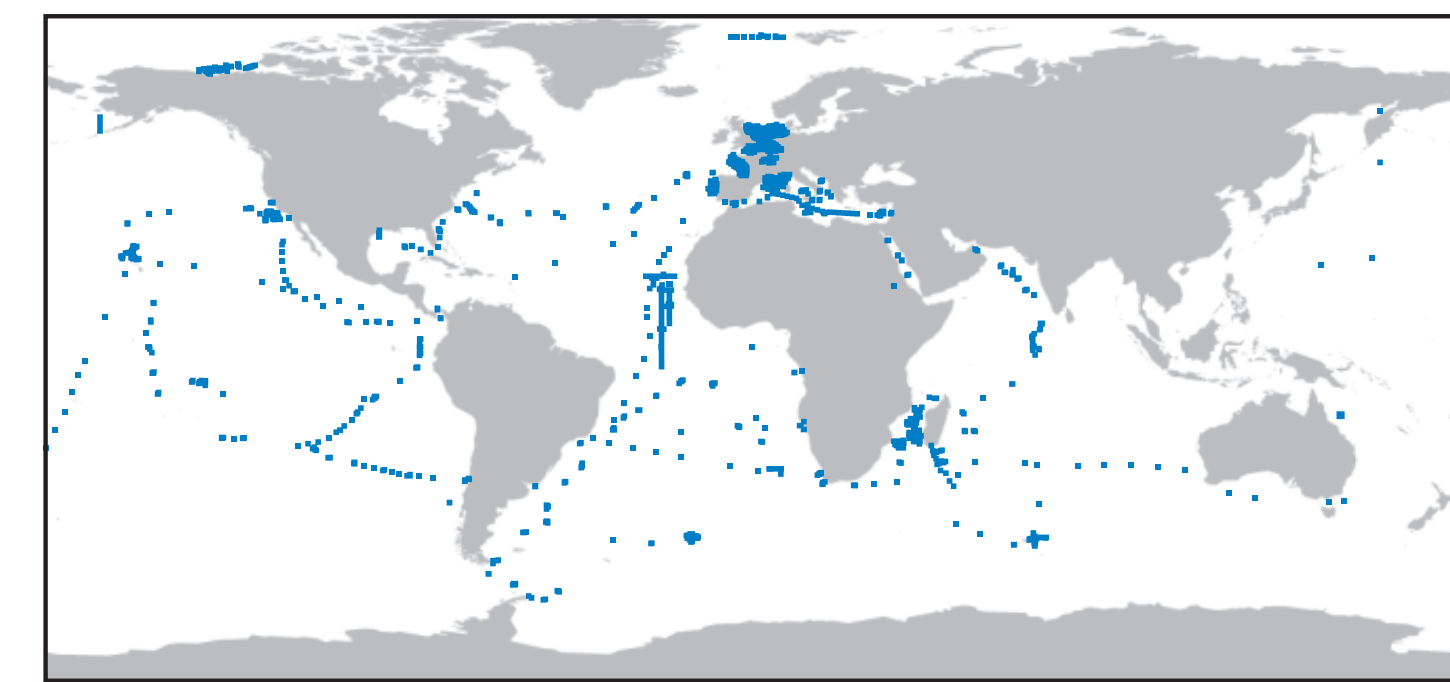
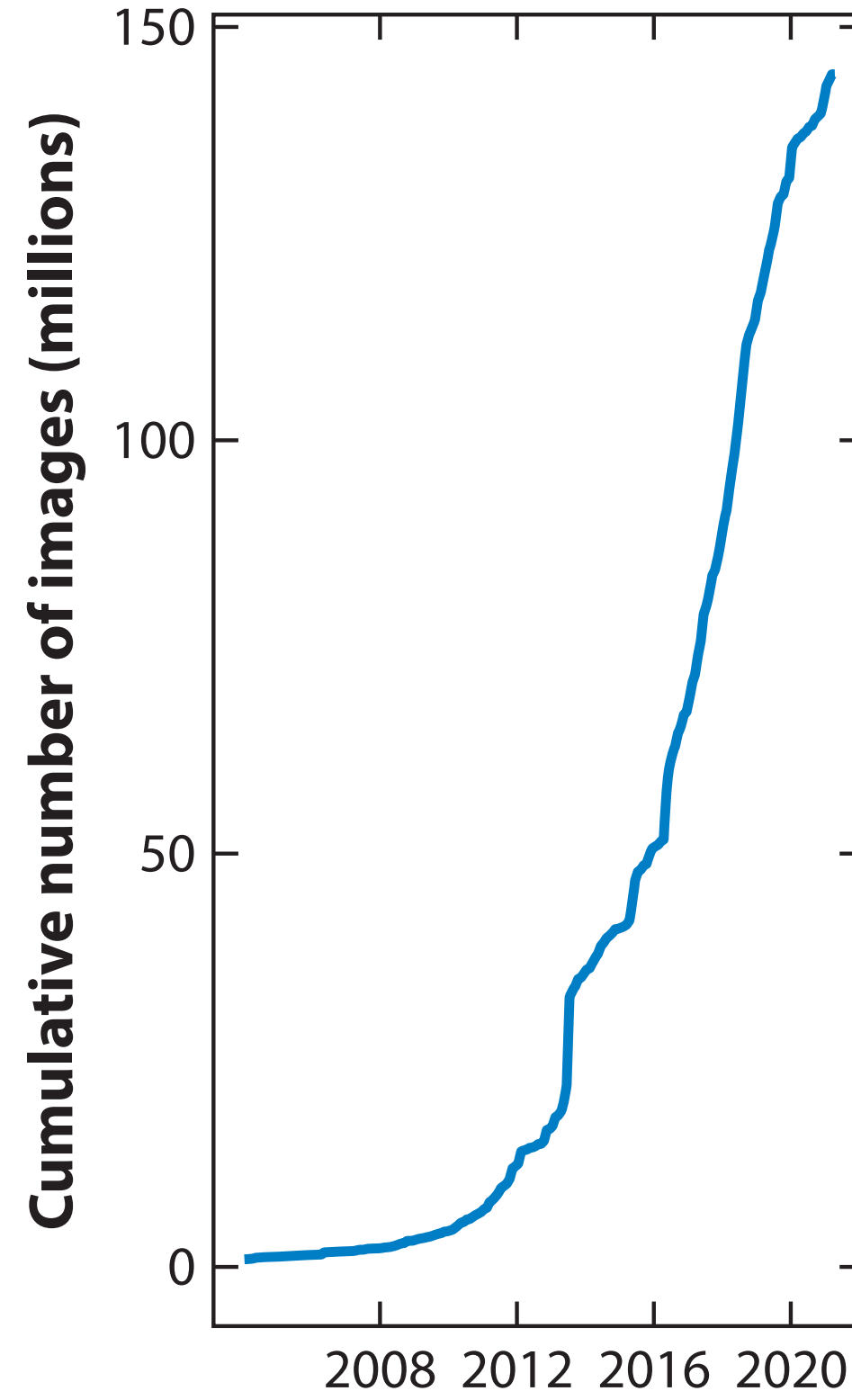
101M classified

+5M per month

1,800 users from **550 institutions** worldwide

At any point of the day,
90 sessions from
50 users

Sorting of **5,000 to 10,000 images** per day
per operator



How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3

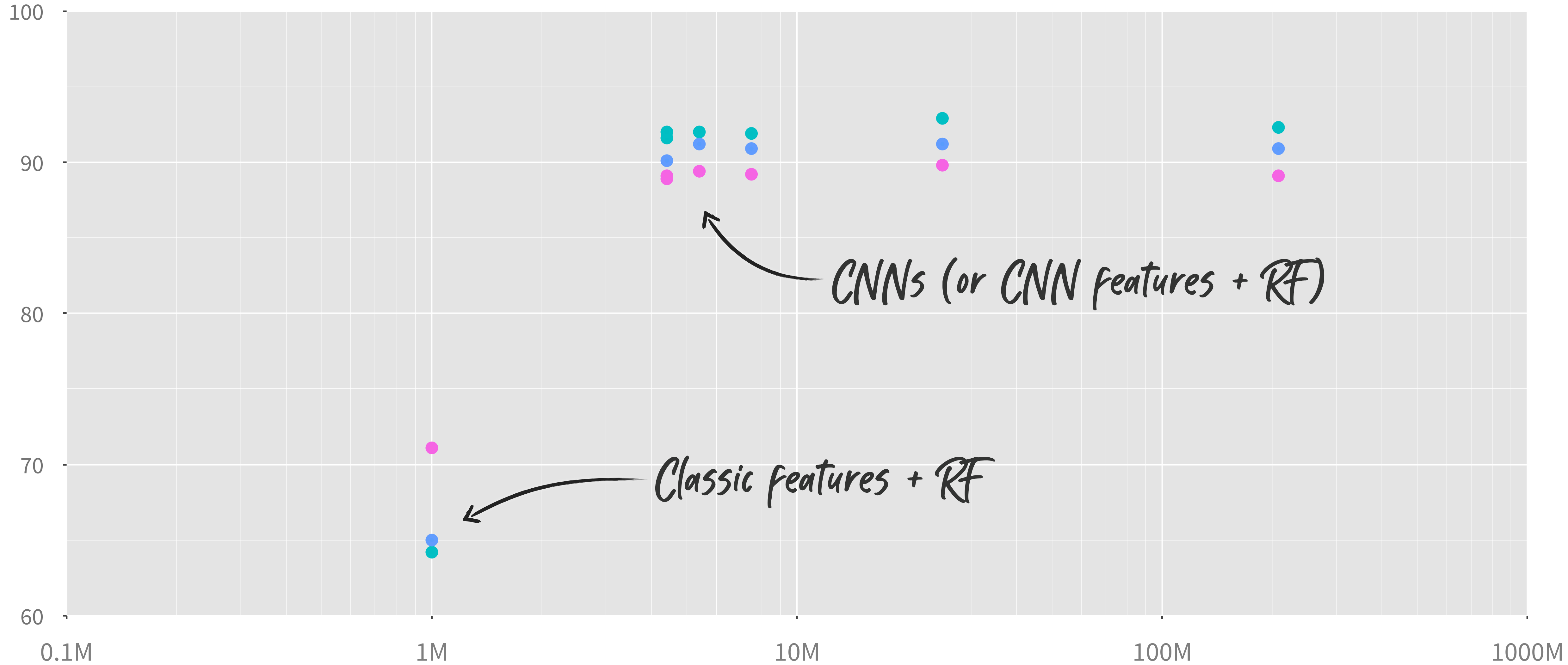
How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3
MobileNet v4 + 50	4.4M	88.9	90.1	901.6

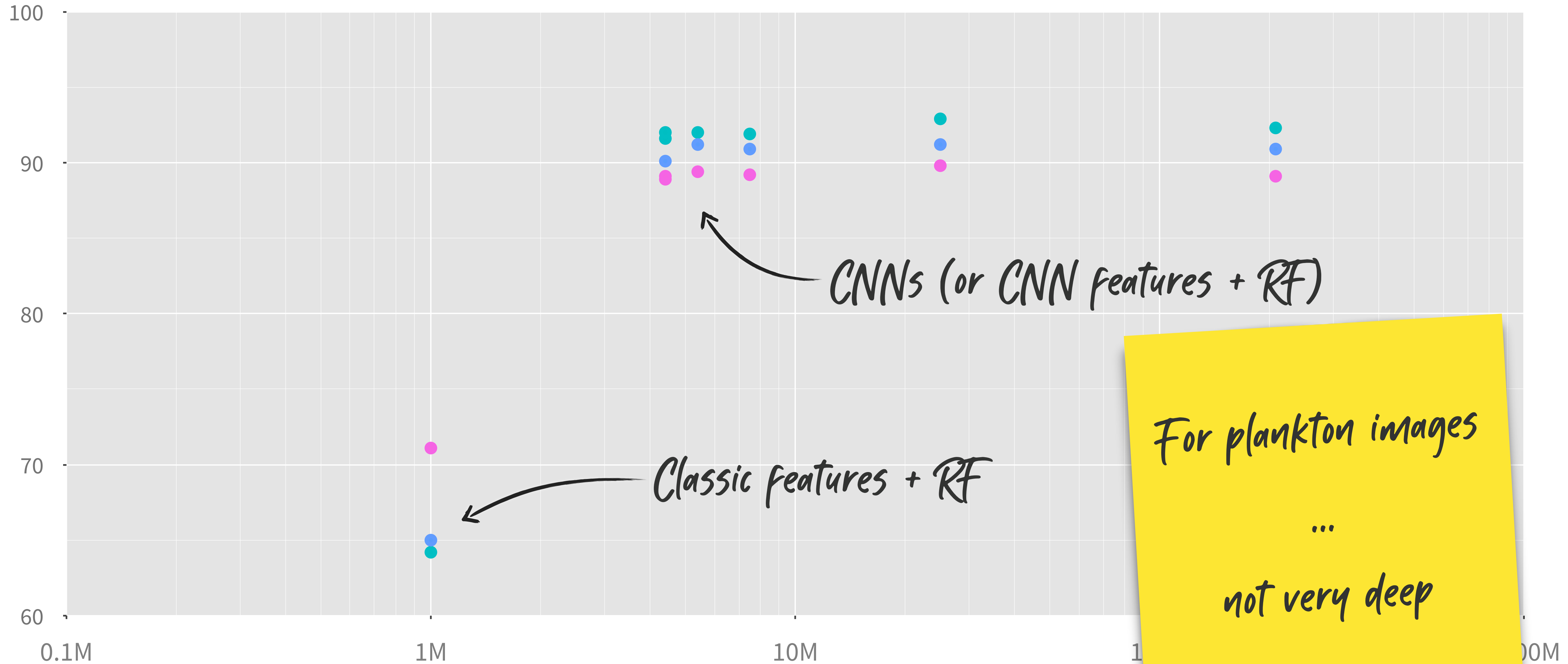
How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
Classic features + Random Forest	1M	71.1	65.0	64.2
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3
MobileNet v4 + 50	4.4M	88.9	90.1	901.6
MobileNet v4 + 1792 + PCA 50 + RF	~4.4M	89.1	90.1	92.0

How deep is enough?



How deep is enough?

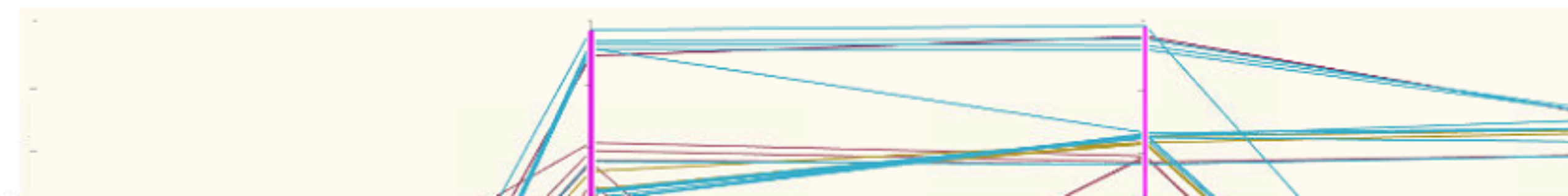
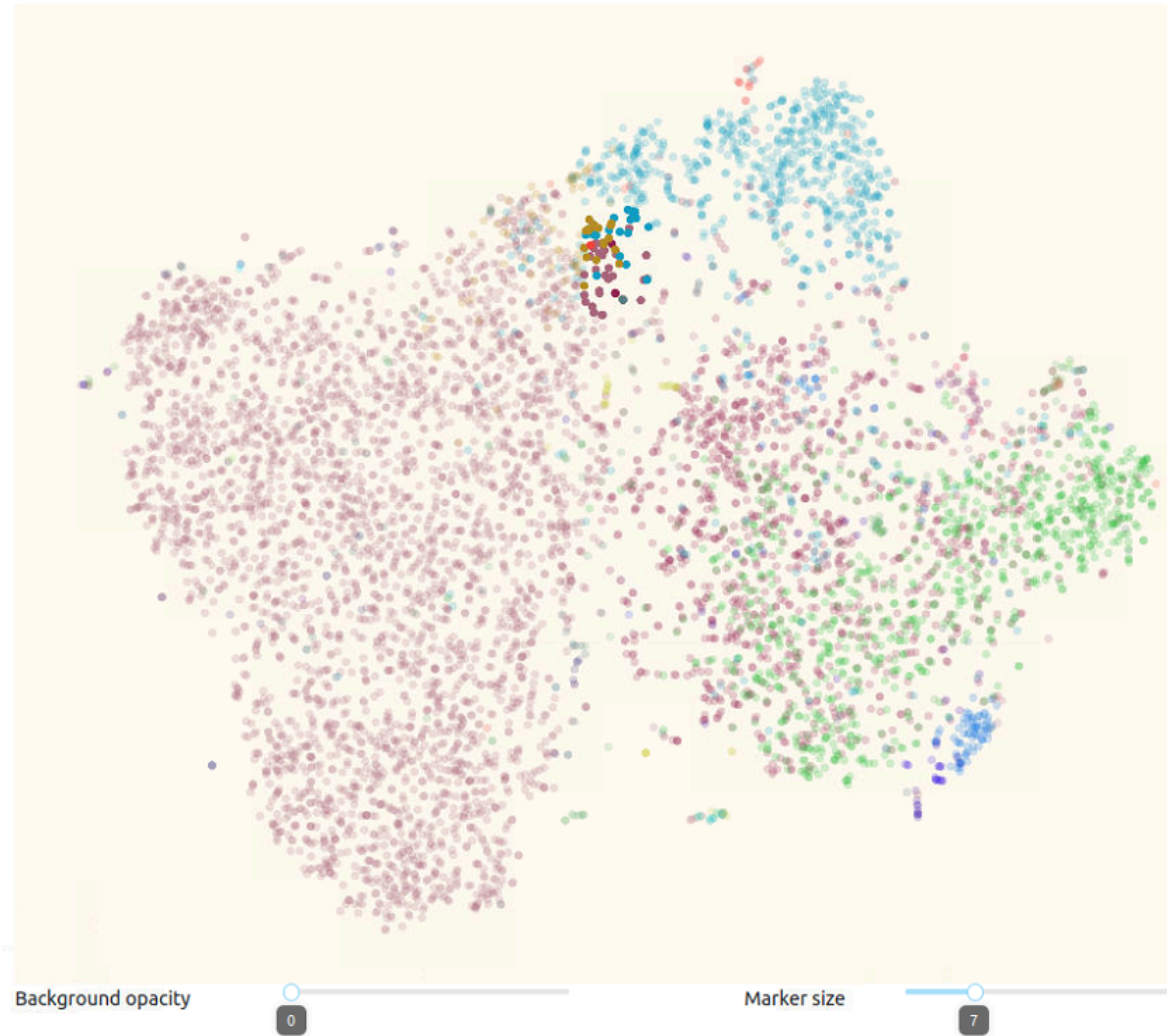


Future plans

Search by **similarity**/dissimilarity

Batch classification

Integration with a dedicated **taxonomic guide**



Future plans

Search by **similarity**/dissimilarity

Batch classification

Integration with a dedicated **taxonomic guide**

Save CSV after labeling Discard images after labeling

enter label here Label

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
coscinodiscus	cosc_sombra	coscinodiscus	coscinodiscus	cosc_sombra	cosc_sombra	sombra	coscinodiscus
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
coscinodiscus	cosc_sombra	coscinodiscus	sombra	cosc_sombra	coscinodiscus	coscinodiscus	coscinodiscus
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
cosc_sombra	cosc_sombra	cosc_sombra	sombra	cosc_sombra	cosc_sombra	cosc_sombra	sombra
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
cosc_sombra	sombra	sombra	sombra	sombra	sombra	detritos	sombra
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
sombra	sombra	sombra	sombra	sombra	multiplos	detritos	sombra
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
sombra	sombra	detritos	detritos	detritos	langanho		

Marked images first Similarity

Invert marks

Future plans

Search by **similarity**/dissimilarity

Batch classification

Integration with a dedicated **taxonomic guide**

EcoTaxa 2.6 Zooscan_Reference (1444921, 0, 0, 0 / 1444921)

Project Filtered Filter: Status=Validated X

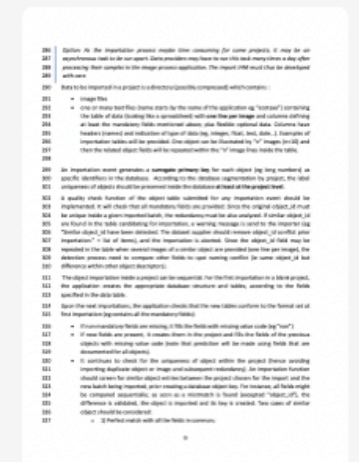
Update view & apply filter Select all Image Nam Display Status Validated 100 % 50

Taxonomy filter Other filters

Actinopterygii	1925
egg < Actinopterygii	4587
scale	48
Annelida	2167
Alciopidae	93
Poebius	91
Tomopteridae	550
larvae < Annelida	328
part < Annelida	989
Appendicularia	0
Fritillariidae	12132
Oikopleuridae	33108
tail < Appendicularia	8284
trunk < Appendicularia	1283
Branchiostoma lanceolatum	18
Bryozoa	0
cyphonaute	8888
Chaetognatha	52391
head < Chaetognatha	1266
tail < Chaetognatha	3694
Cnidaria < Metazoa	0
Actiniaria	140
Hydrozoa	3856
Liriope < Geryoniidae	0 218
Liriope tetraphylla	104
Obelia	974
Rhopalonematidae	0
Aglaura	3033

Grid of images showing various marine organisms, including larvae and appendicularians, with 1 mm scale bars. The organisms are labeled with their taxonomic classification and sample IDs.

- Cumacea a1m_10112015_300_micra_d2_1_1376
- zoea < Brachyura a1m_10112015_300_micra_d2_1_1516
- Cumacea a1m_10112015_300_micra_d2_1_1554
- zoea < Brachyura a1m_10112015_300_micra_d2_1_2174
- Cumacea a1m_10112015_300_micra_d2_1_2566
- Cumacea a1m_10112015_300_micra_d2_1_643
- Mysida a1m_10112015_300_micra_d2_1_3747
- Cumacea a1m_10112015_300_micra_d2_1_783



247 **2.5 Sharing reference list of categories**

248 The updated list of categories must be portable in order to be able to update an application running

249 on a field computer without any Internet connexion.

250 **2.6 Quality flag**

251 The **quality** of the identification is noted by a **flag** (*annotation_status*) which is automatically set by

252 the application in most cases.

253 The quality flag can be read at the importation stage if exists in the spreadsheet provided with the

254 images.

255 The possible values are:

256 - No identification (“noid”)

257 - Identification from the automated prediction output (“predicted”)

258 - A manual doubtful identification needs to be confirmed by other annotators (“doubtful”)

259 - Identification is fully validated (“validated”)

260

261 • Flag changes from “predicted” to “validated” when the vignette is manually dragged into a

262 category or renamed or when the “validate all” button is pressed.

263 • Flag changes to “doubtful” when selected in the object page.

264 • Flag changes from “noid” to “predicted” when an automatic prediction is made

265 • Flag changes from “noid” to “validated” when the vignette is manually dragged into a

266 category (ie when it is manually and intentionally renamed).

Lessons learned

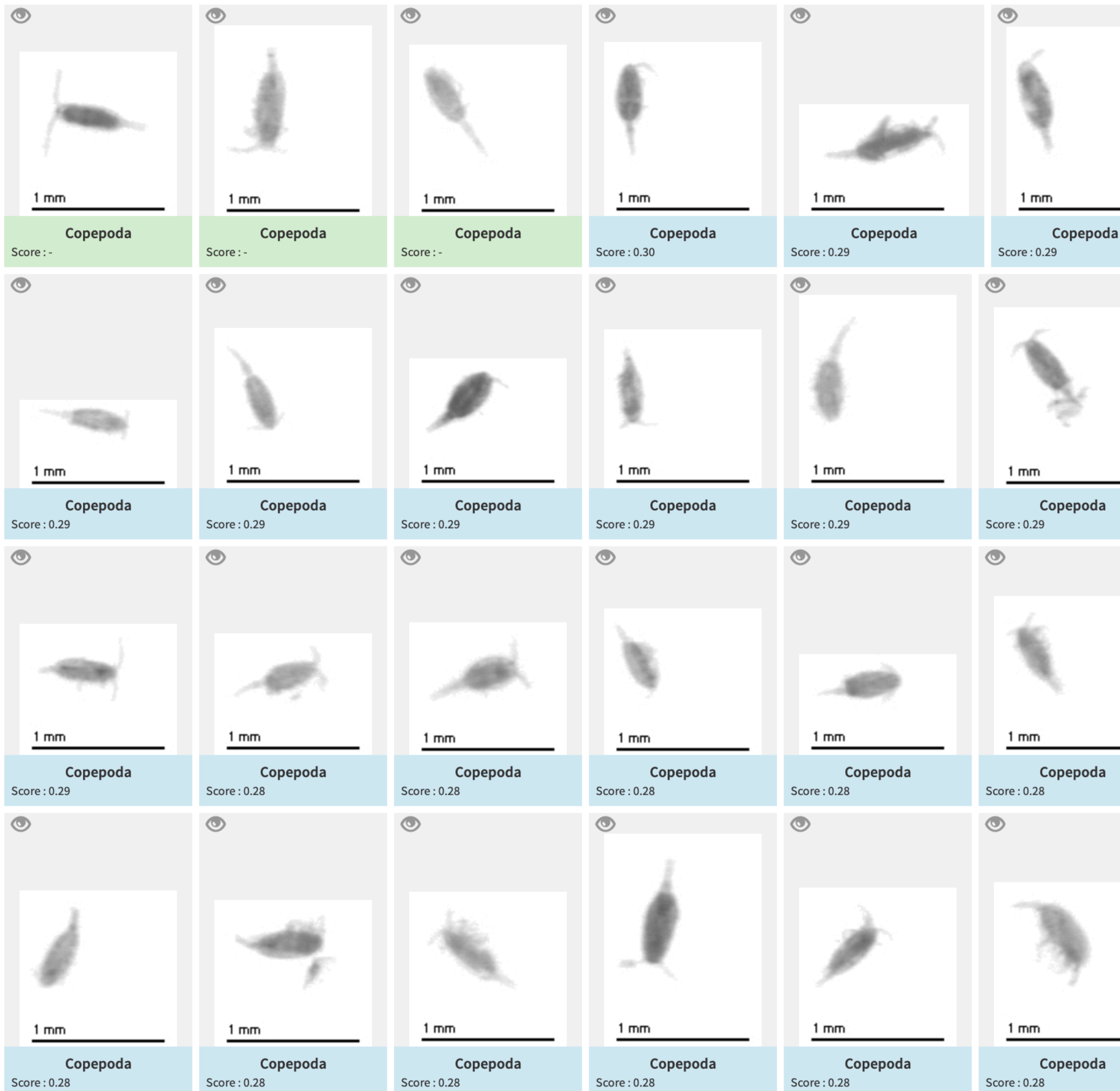
Subcontracting has the advantage of forcing to **specify** features

267 **2.7 Users roles and rights**

268 Four roles of users and associated rights/privileges are considered:

	visitor	annotator	project manager	application manager (unique)
View images	X	X	X	X
Annotate images (manual)		X	X	X
Register annotators			X	X
Annotate images (automatic)			X	X
Import new objects in project			X	X
Set project status			X	X
Create subset from project			X	X
Manage the taxonomy list			X	X
Create classes			X	X
Create project				X
Register project managers				X
		X		

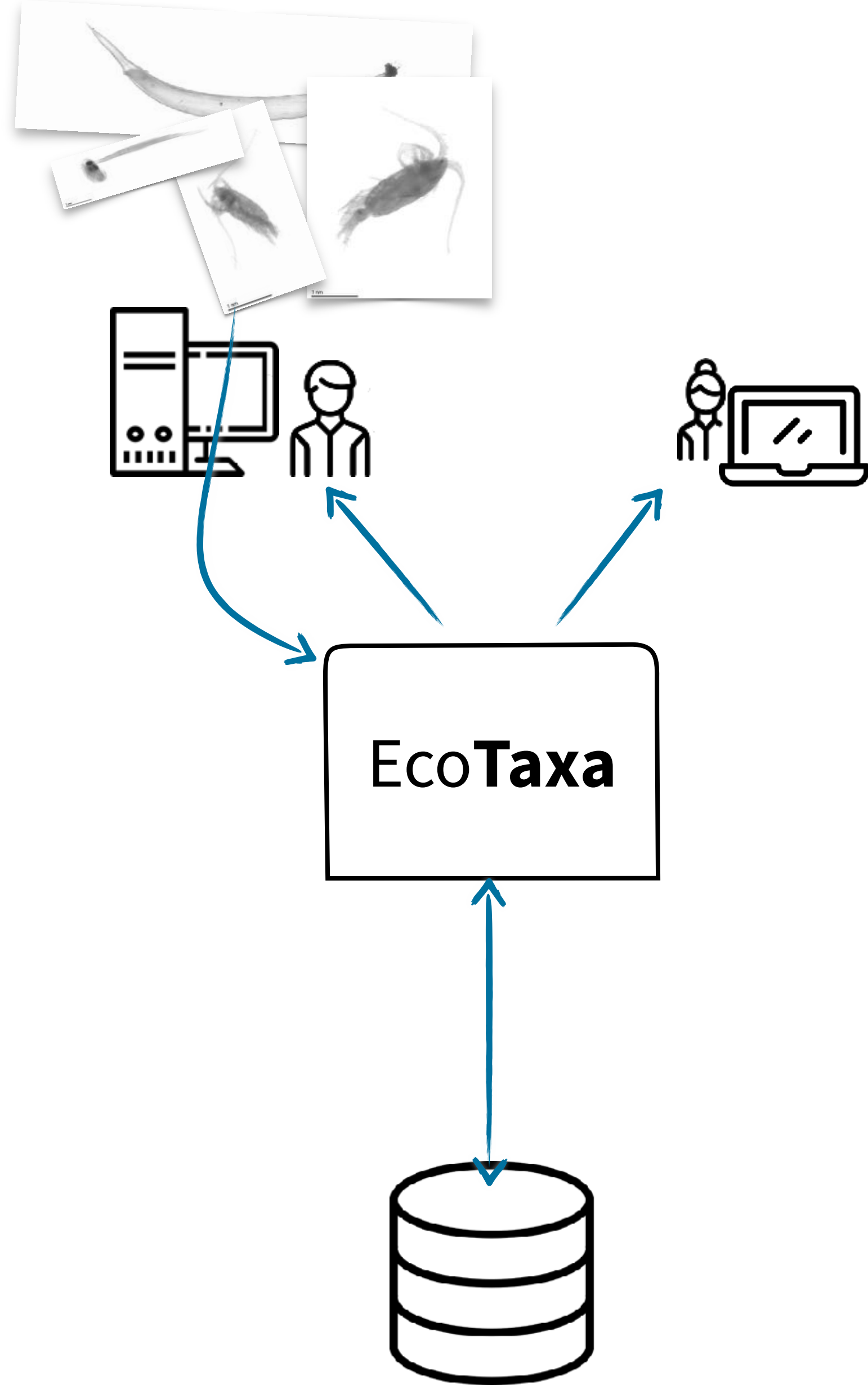
(with notification)



Lessons learned

Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**



Lessons learned

Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**

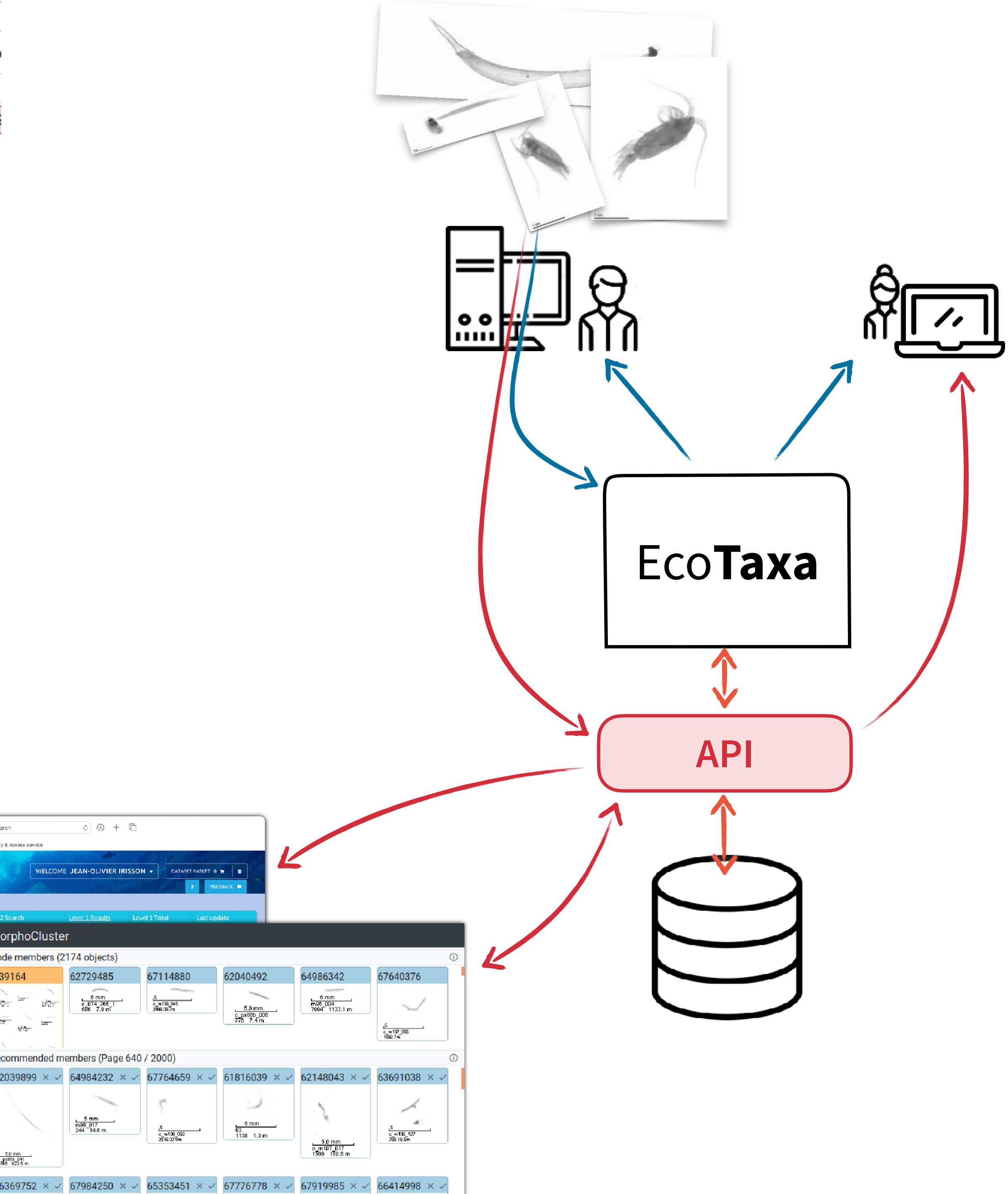
An **Application Programming Interface** is the center of the modern web

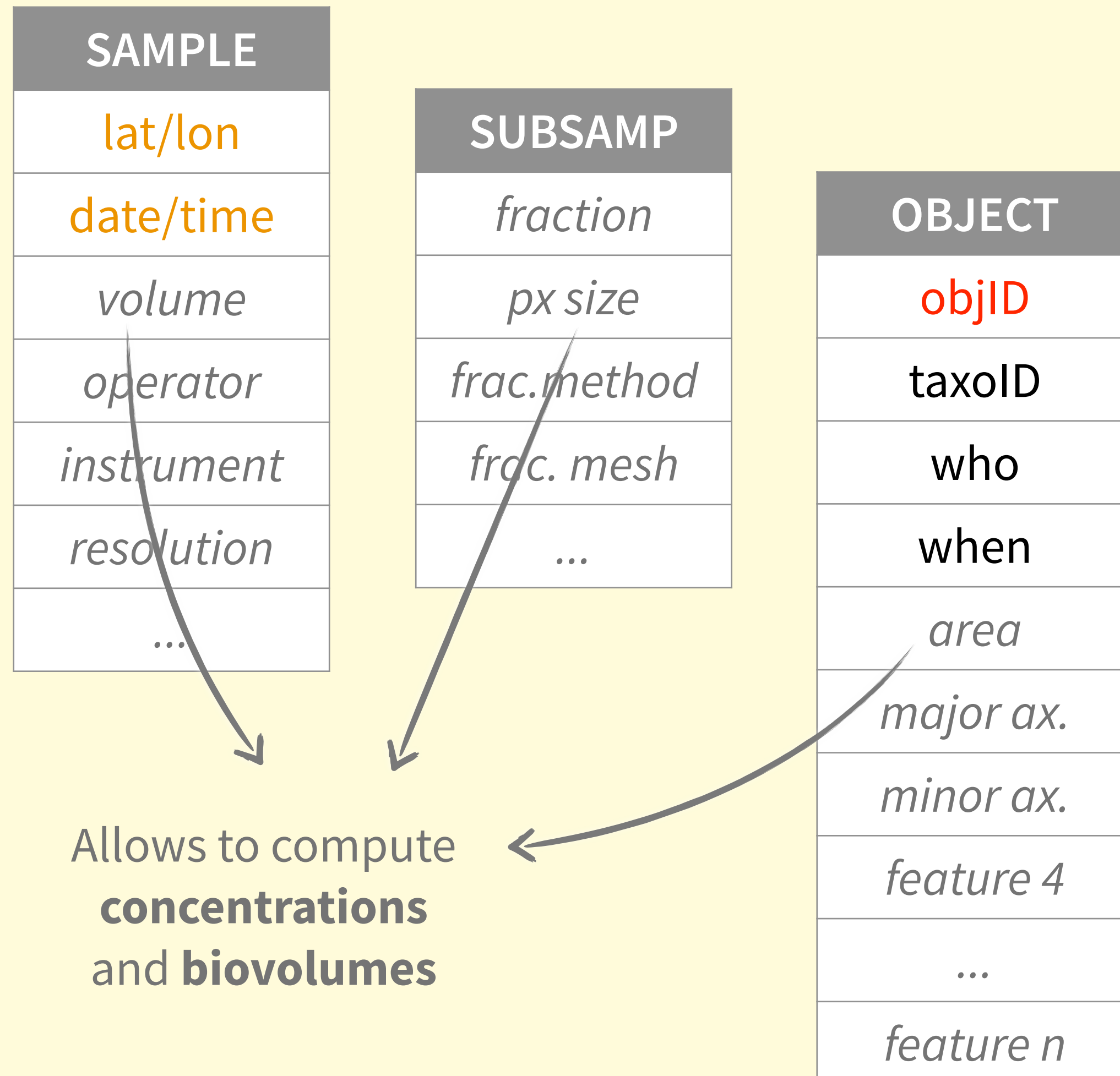
Lessons learned

Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**

An **Application Programming Interface** is the center of the modern web





Lessons learned

Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**

An **Application Programming Interface** is the center of the modern web

Flexibility is good, too much flexibility is **difficult**



Lessons learned

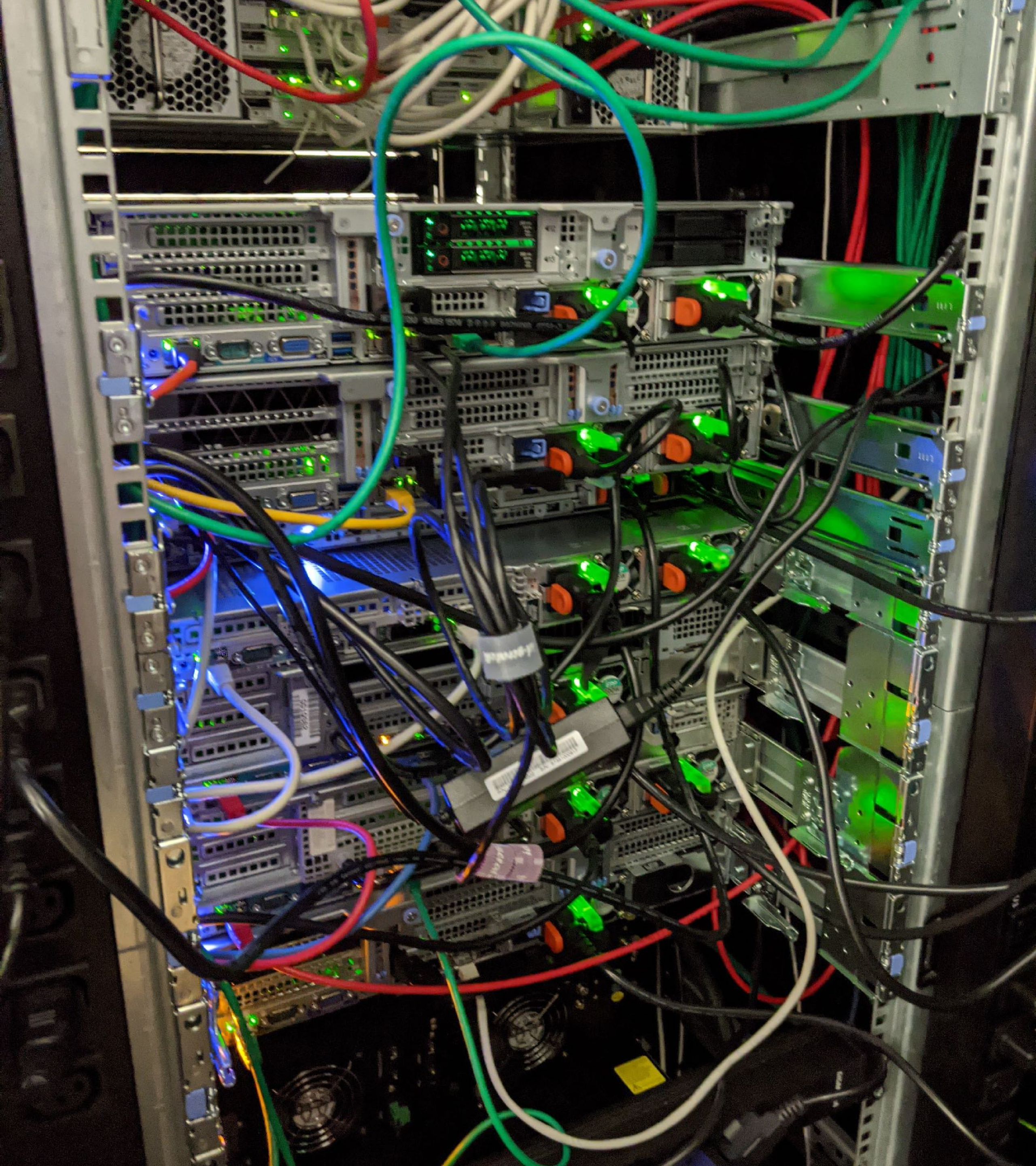
Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**

An **Application Programming Interface** is the center of the modern web

Flexibility is good, too much flexibility is **difficult**

Proper (web) development is **expensive**



Lessons learned

Subcontracting has the advantage of forcing to **specify** features

Machine learning as an **assistance** rather than an end in itself; and **KISS**

An **Application Programming Interface** is the center of the modern web

Flexibility is good, too much flexibility is **difficult**

Proper (web) development is **expensive**

Institutional **hosting** is a problem

Merci

irisson@normalesup.org



SCIENCES
SORBONNE
UNIVERSITÉ

