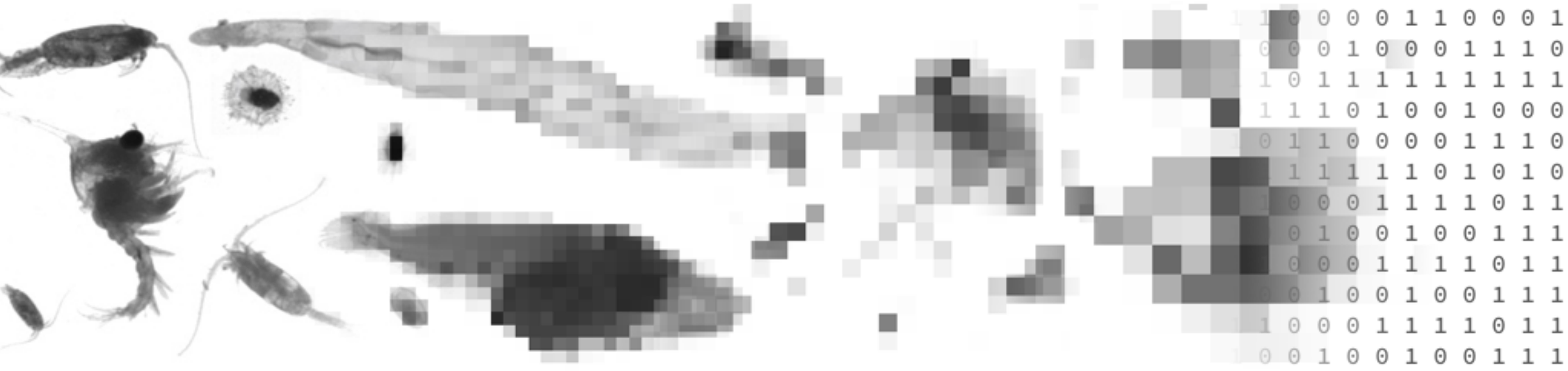
SRI Conference, 2022-06-21
Jean-Olivier Irisson (and https://site.wwwpic.net/team)

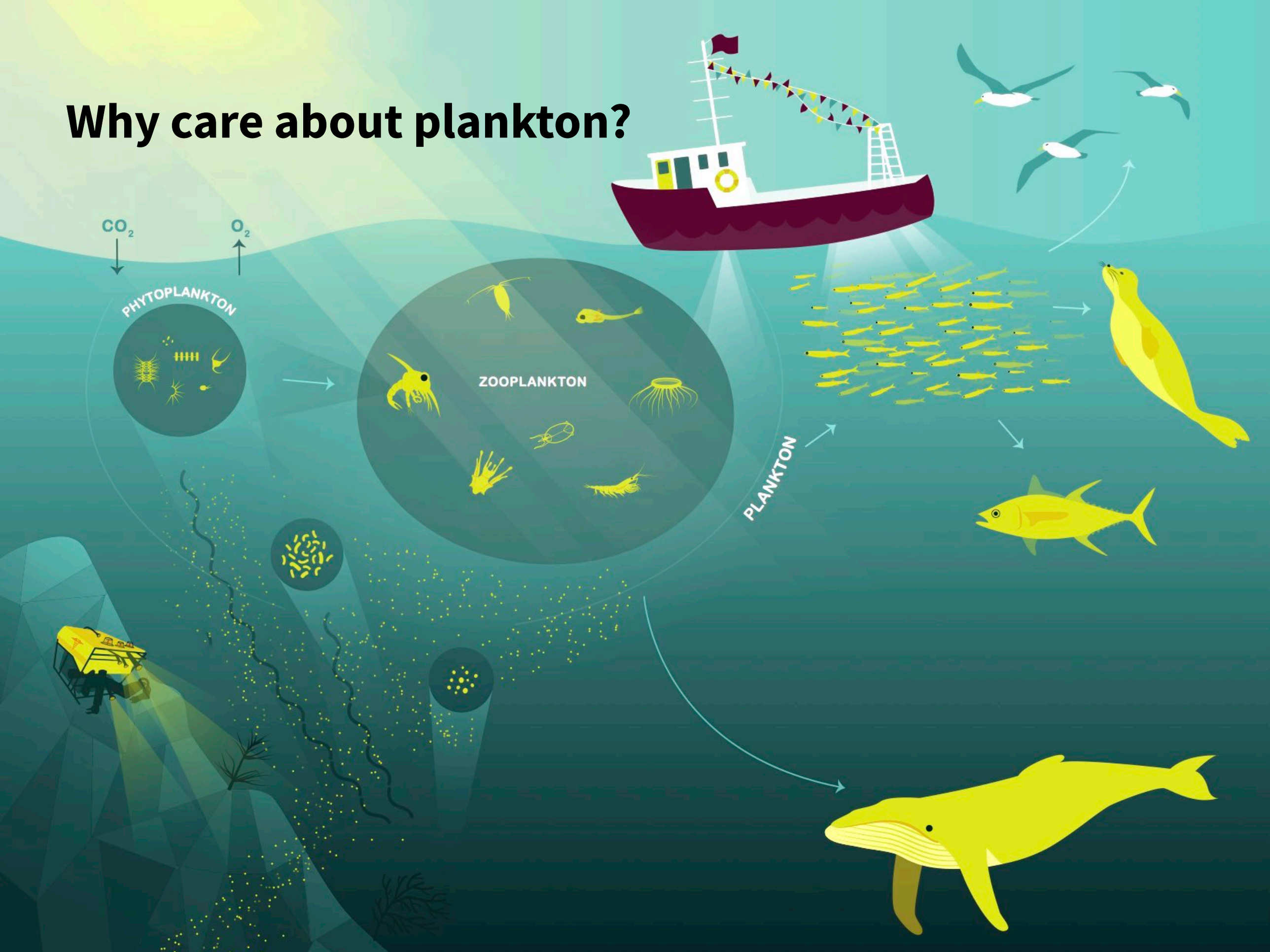
Observing plankton in the open ocean with quantitative imaging

From images to data

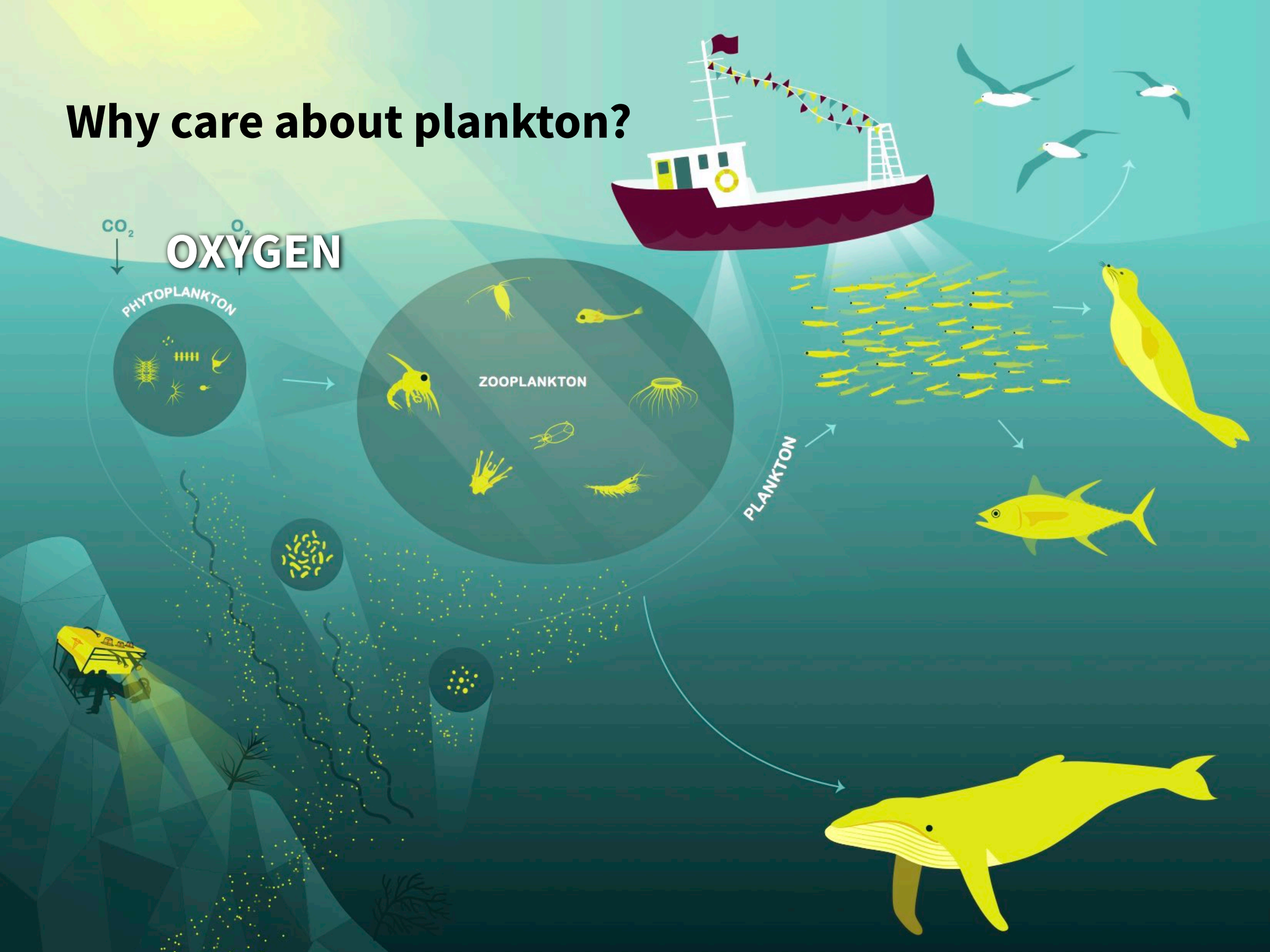
WWWPIC



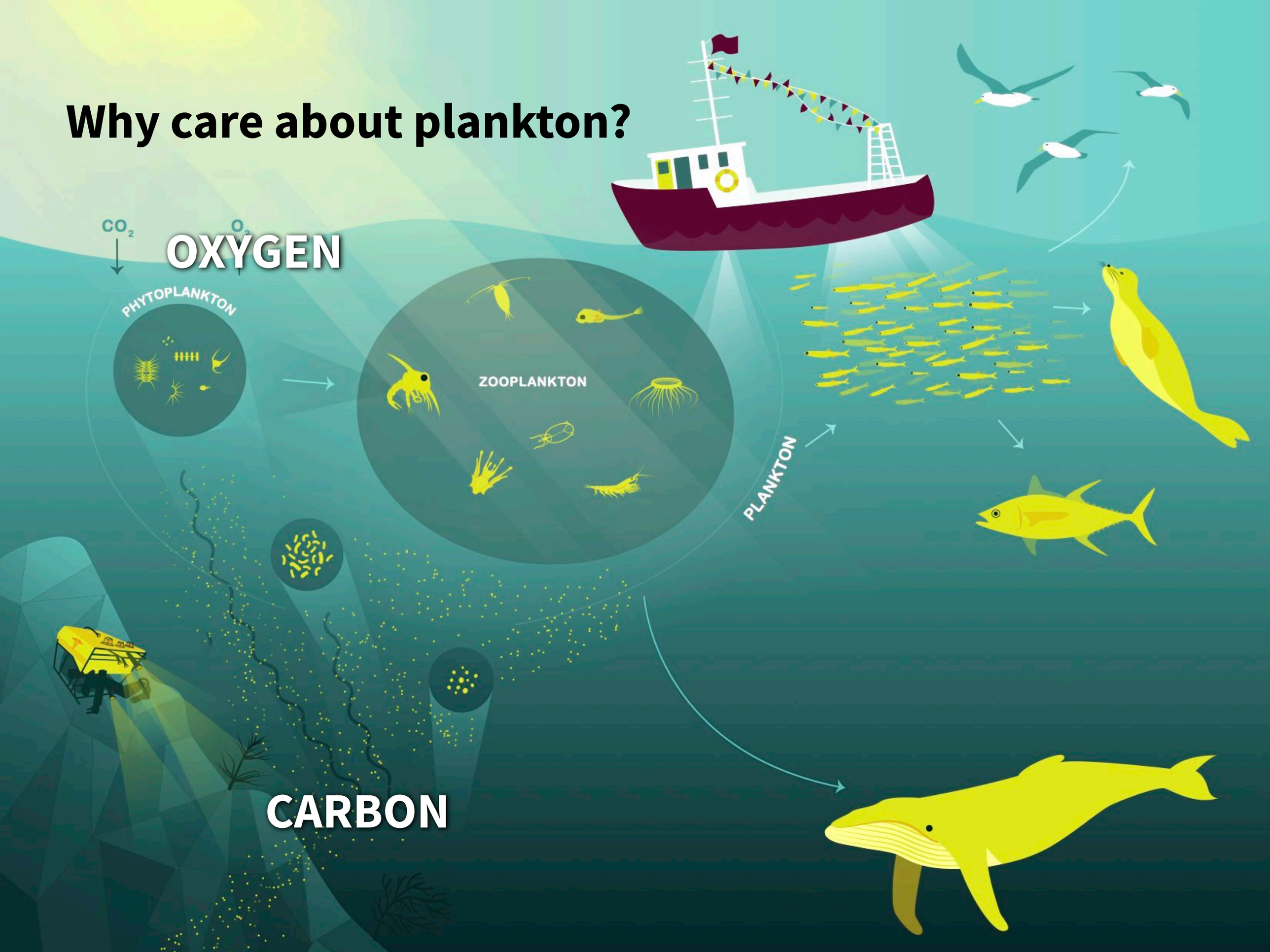
Why care about plankton?



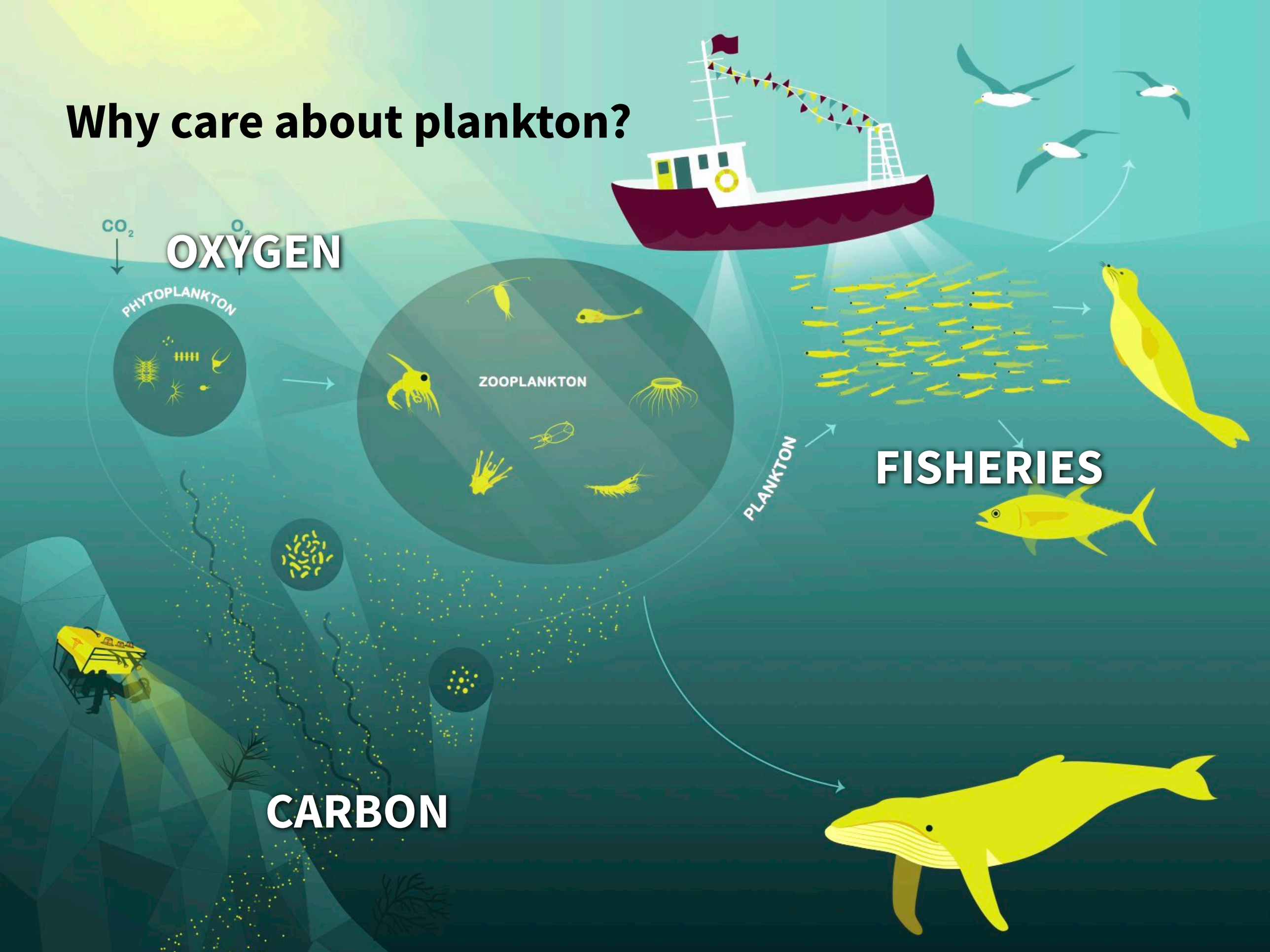
Why care about plankton?



Why care about plankton?



Why care about plankton?



Why care about plankton?

CO_2 O_2

OXYGEN

PHYTOPLANKTON

ZOOPLANKTON

PLANKTON

FISHERIES

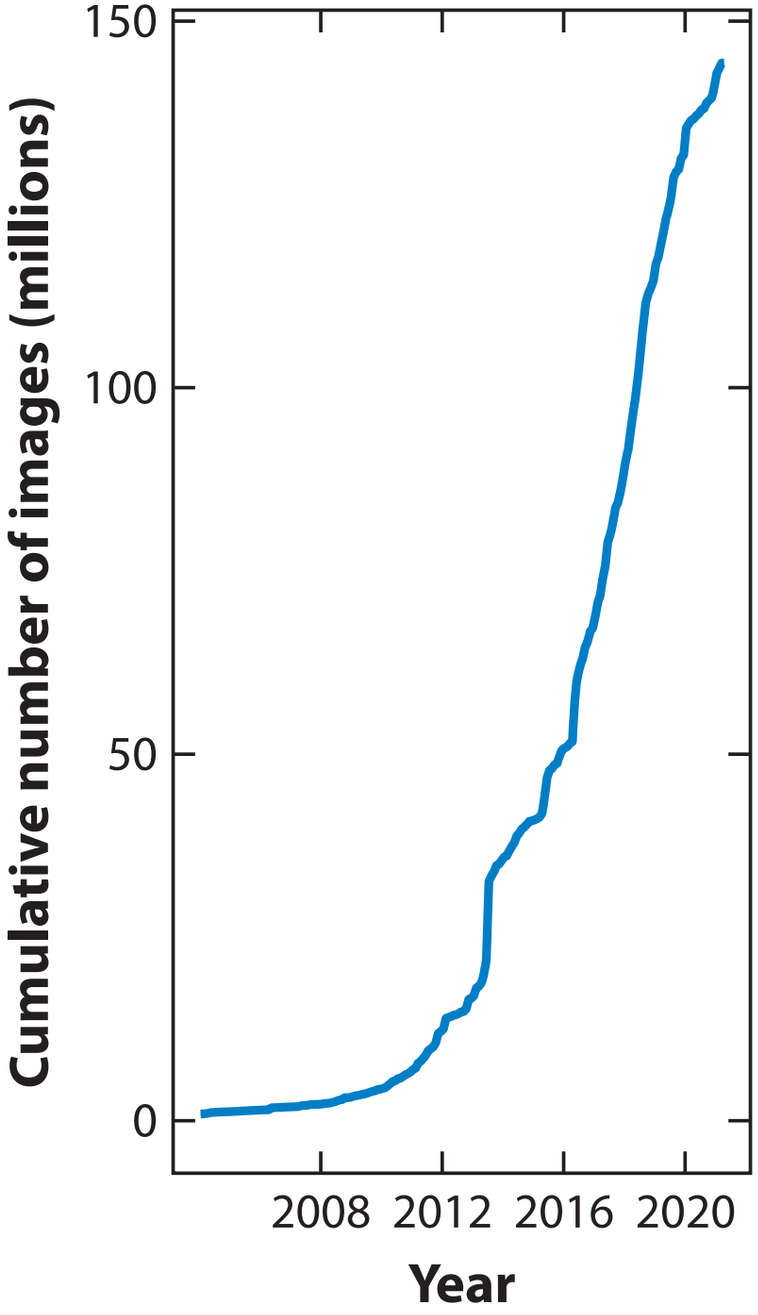
CARBON

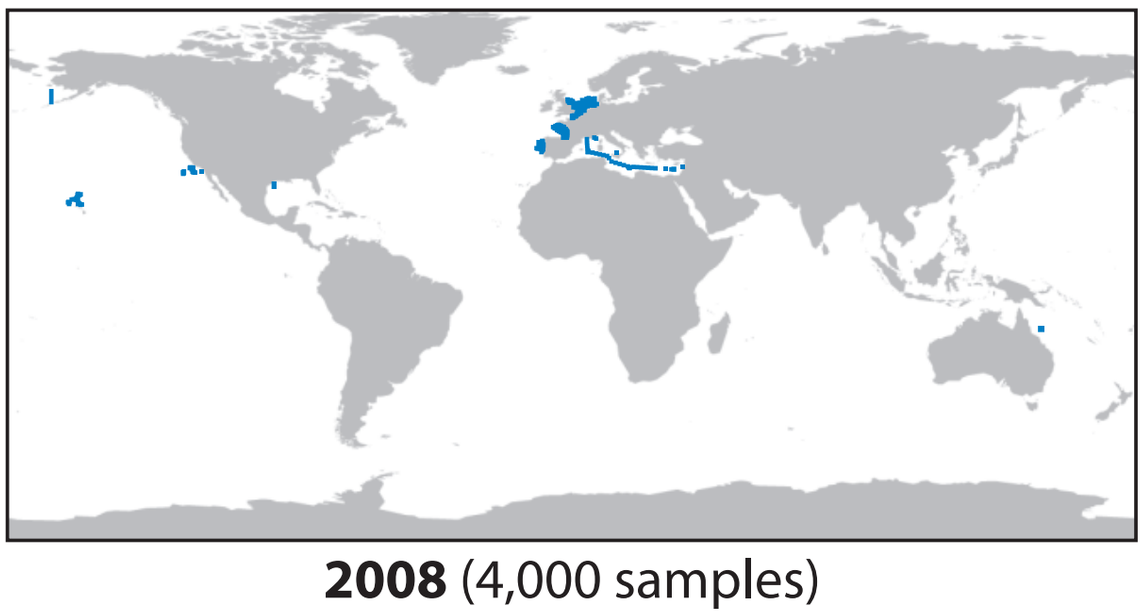
INDICATOR

Quantitative plankton imaging

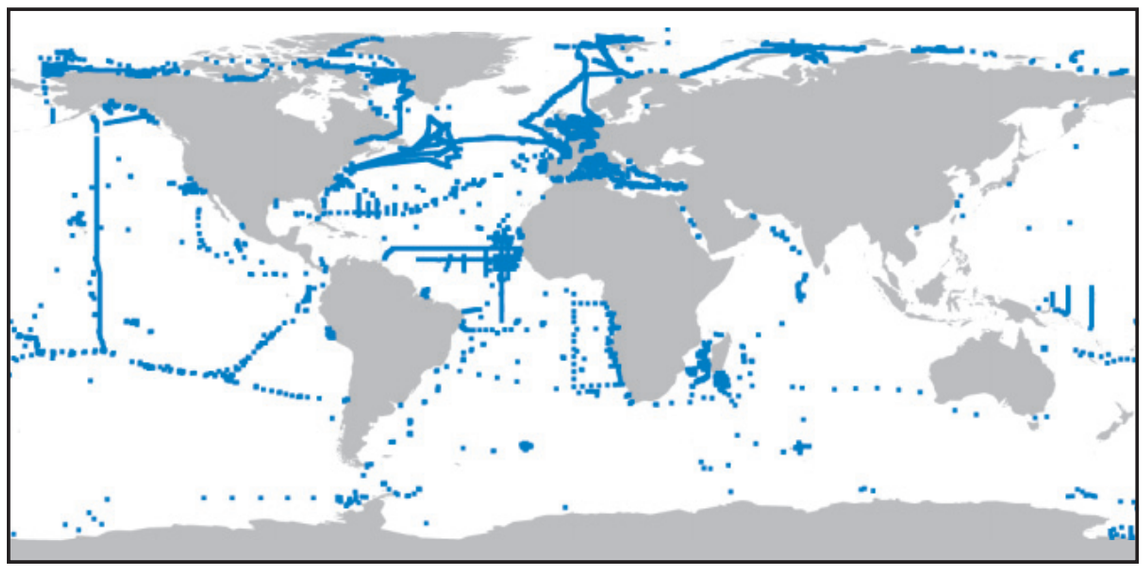
ZooScan = 1 Bpx/y, 1.5M objects/y
UVP = 8.6Bpx/y, ~10M objects/y
ISIIS=25Tpx/y, 100M objects/y

Steep growth in data acquisition

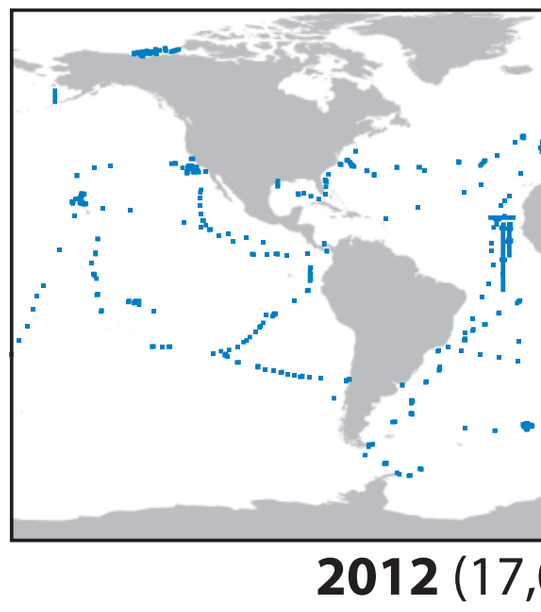




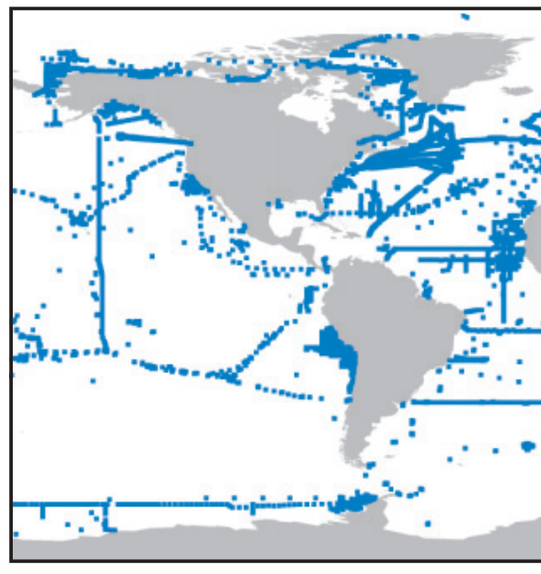
2008 (4,000 samples)



2016 (56,000 samples)

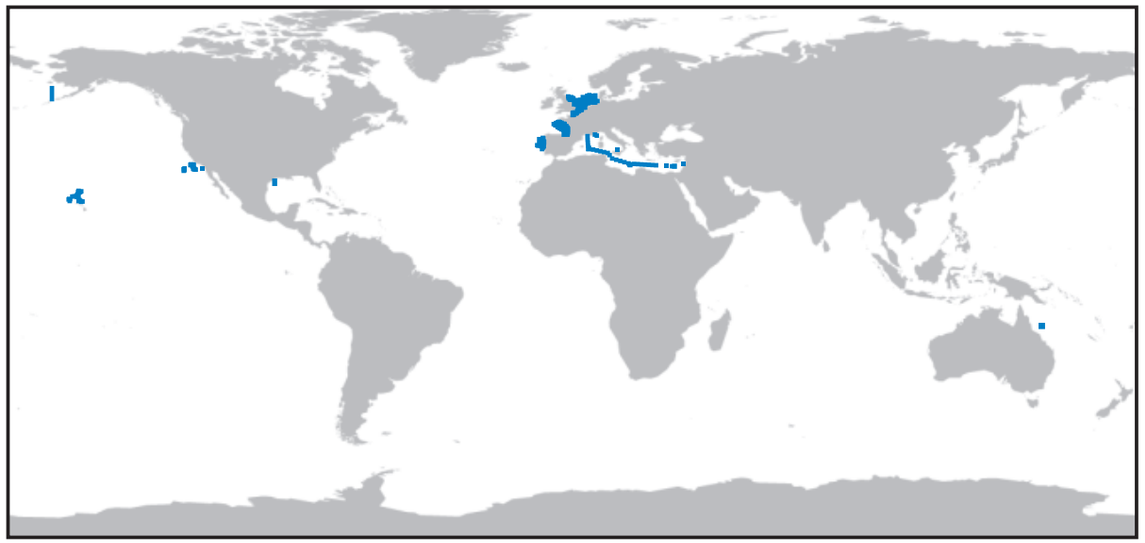


2012 (17,

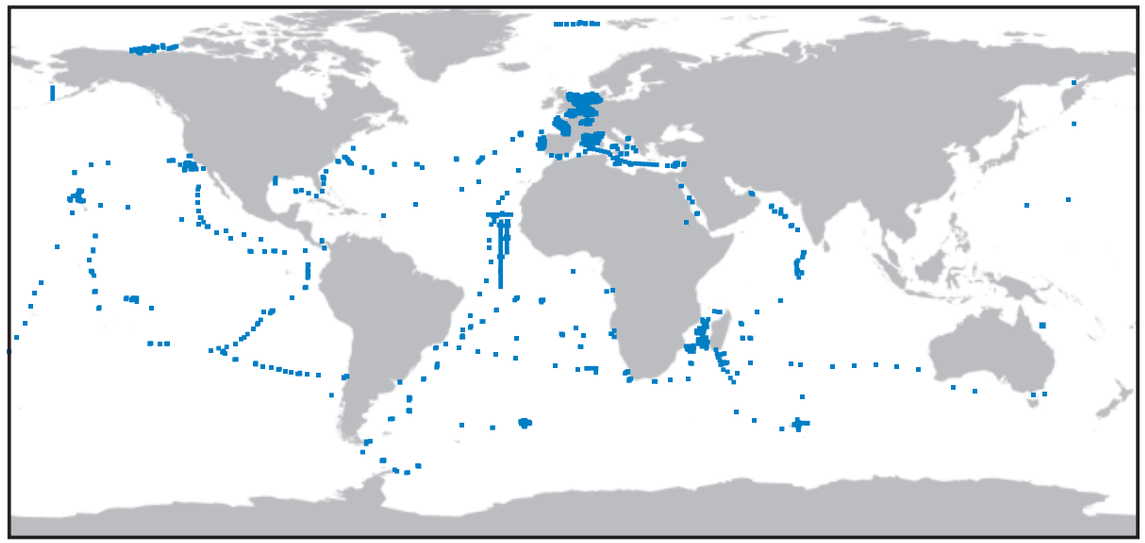


2020 (91,

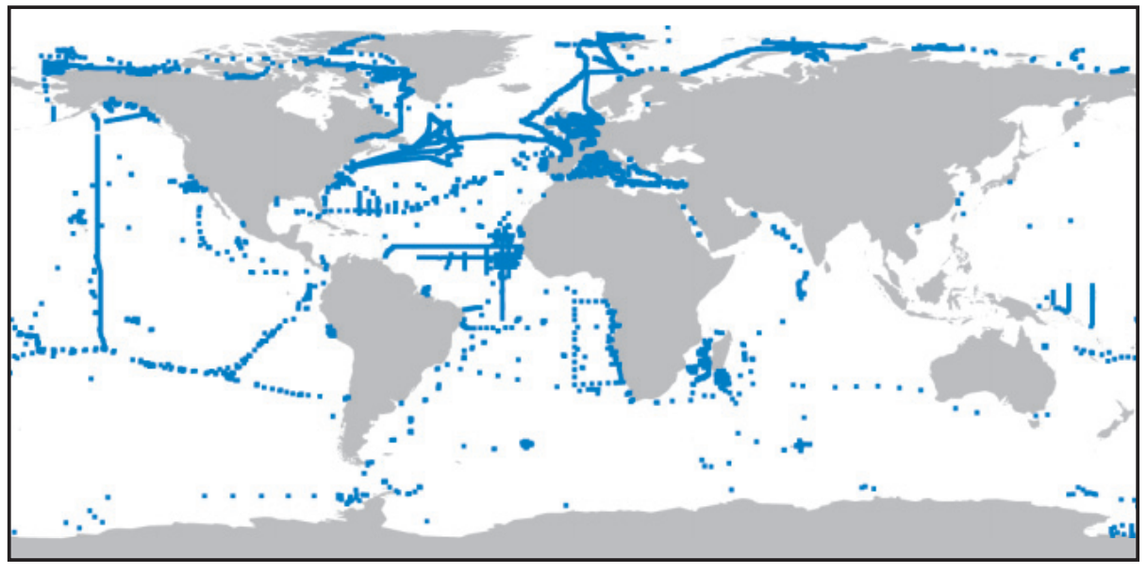
Steep growth in data acquisition



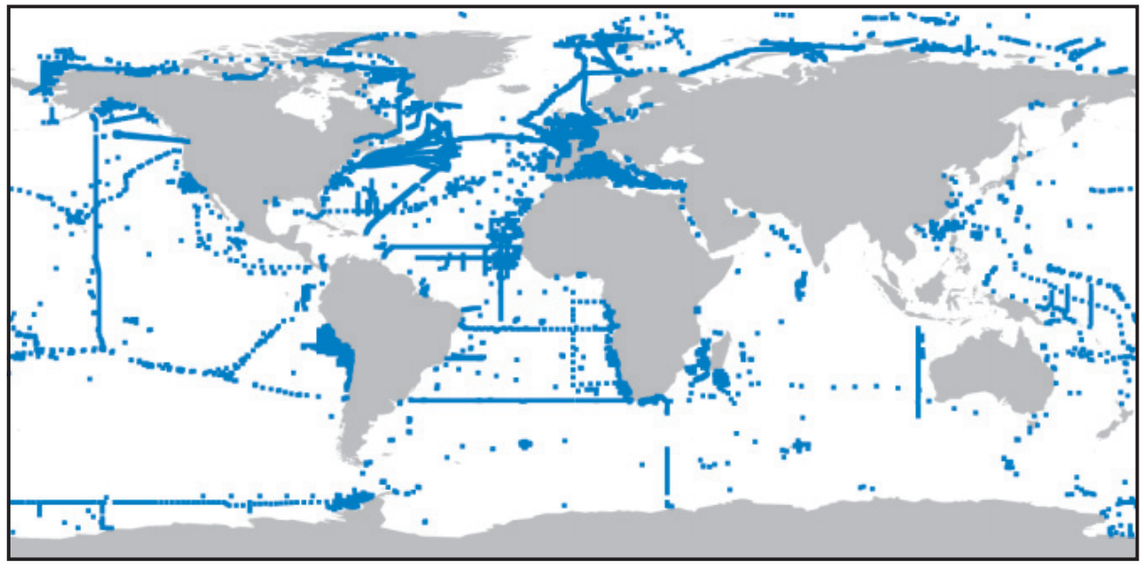
2008 (4,000 samples)



2012 (17,000 samples)

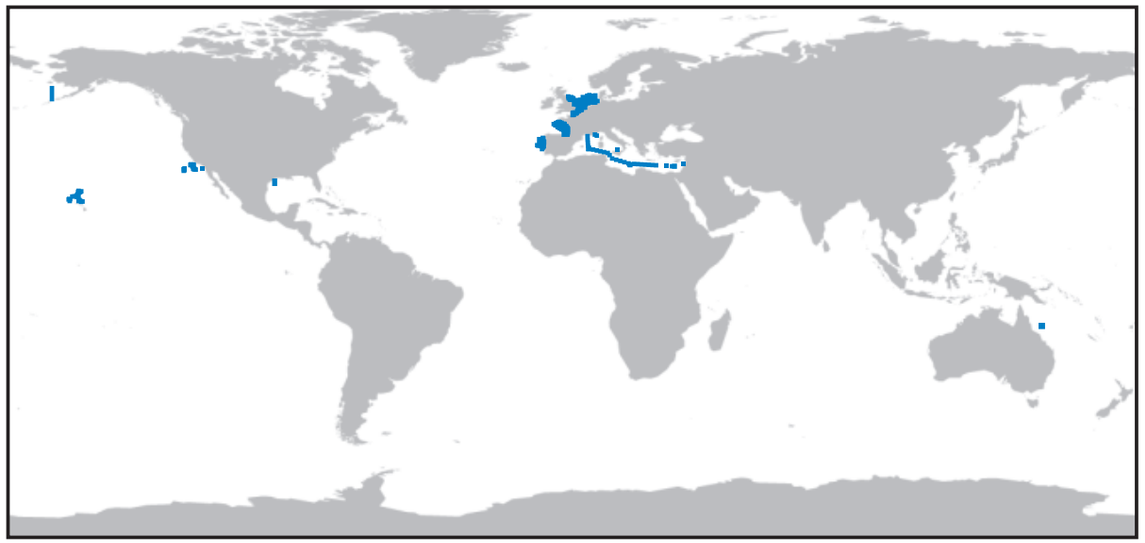


2016 (56,000 samples)

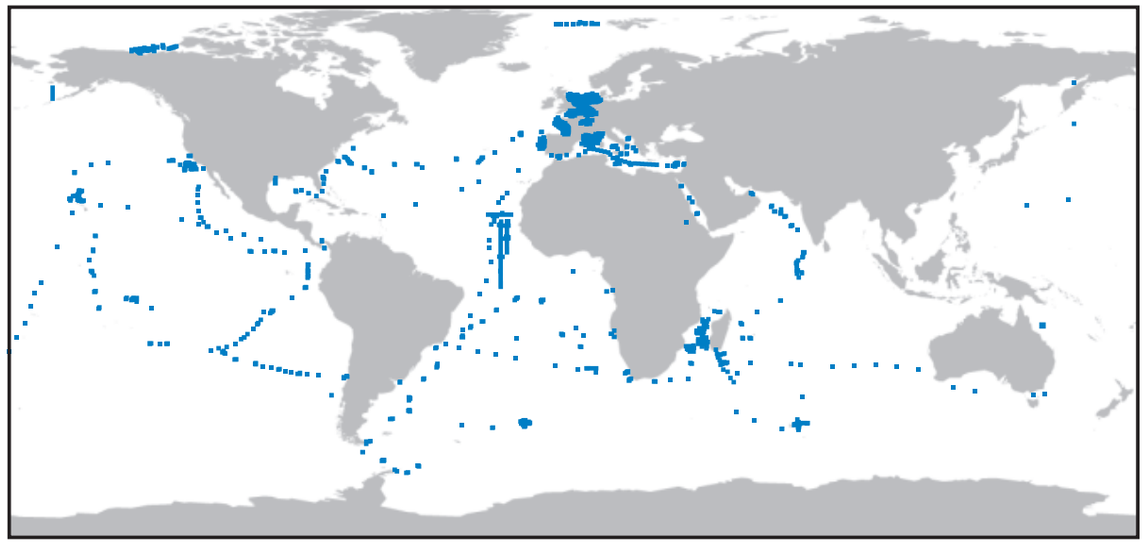


2020 (91,000 samples)

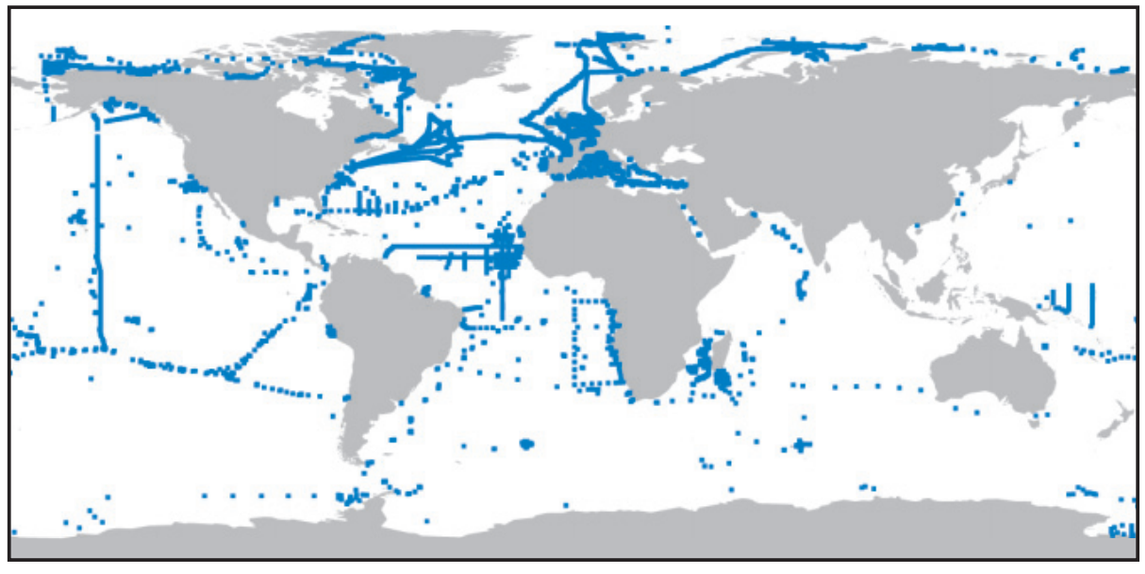
Steep growth in data acquisition



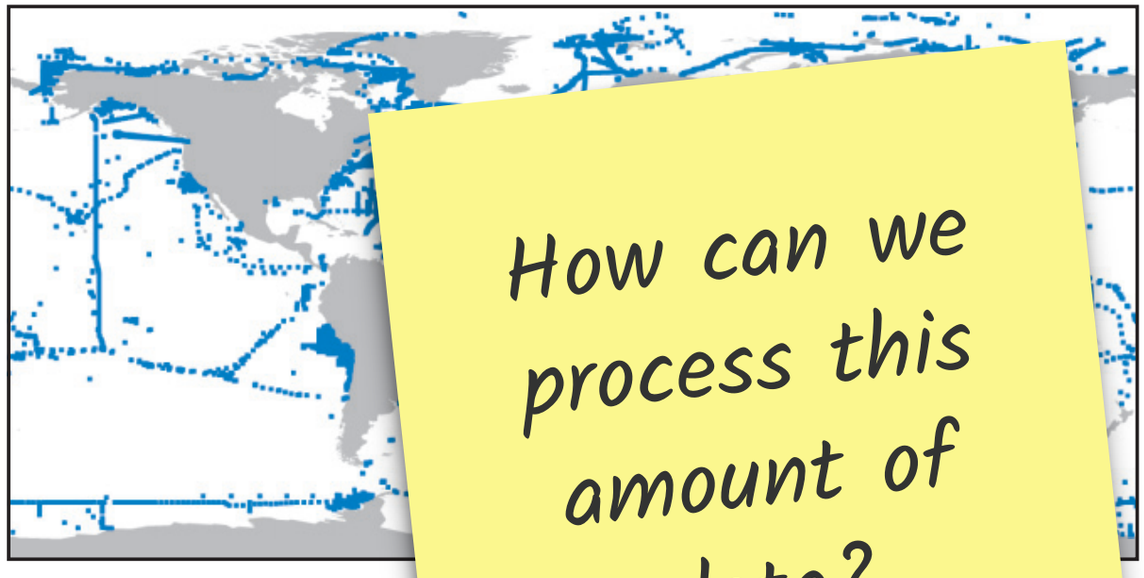
2008 (4,000 samples)



2012 (17,000 samples)



2016 (56,000 samples)

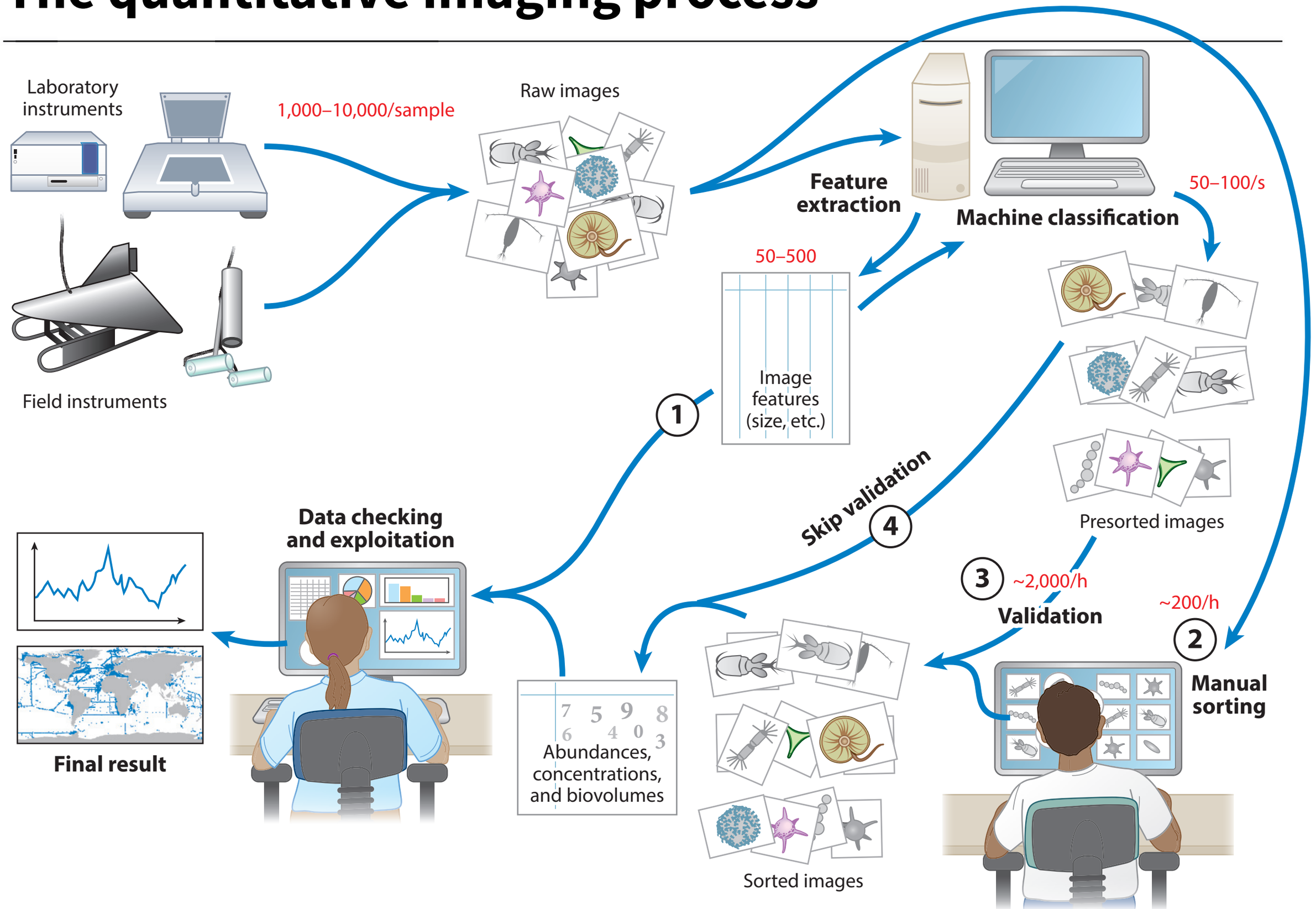


202

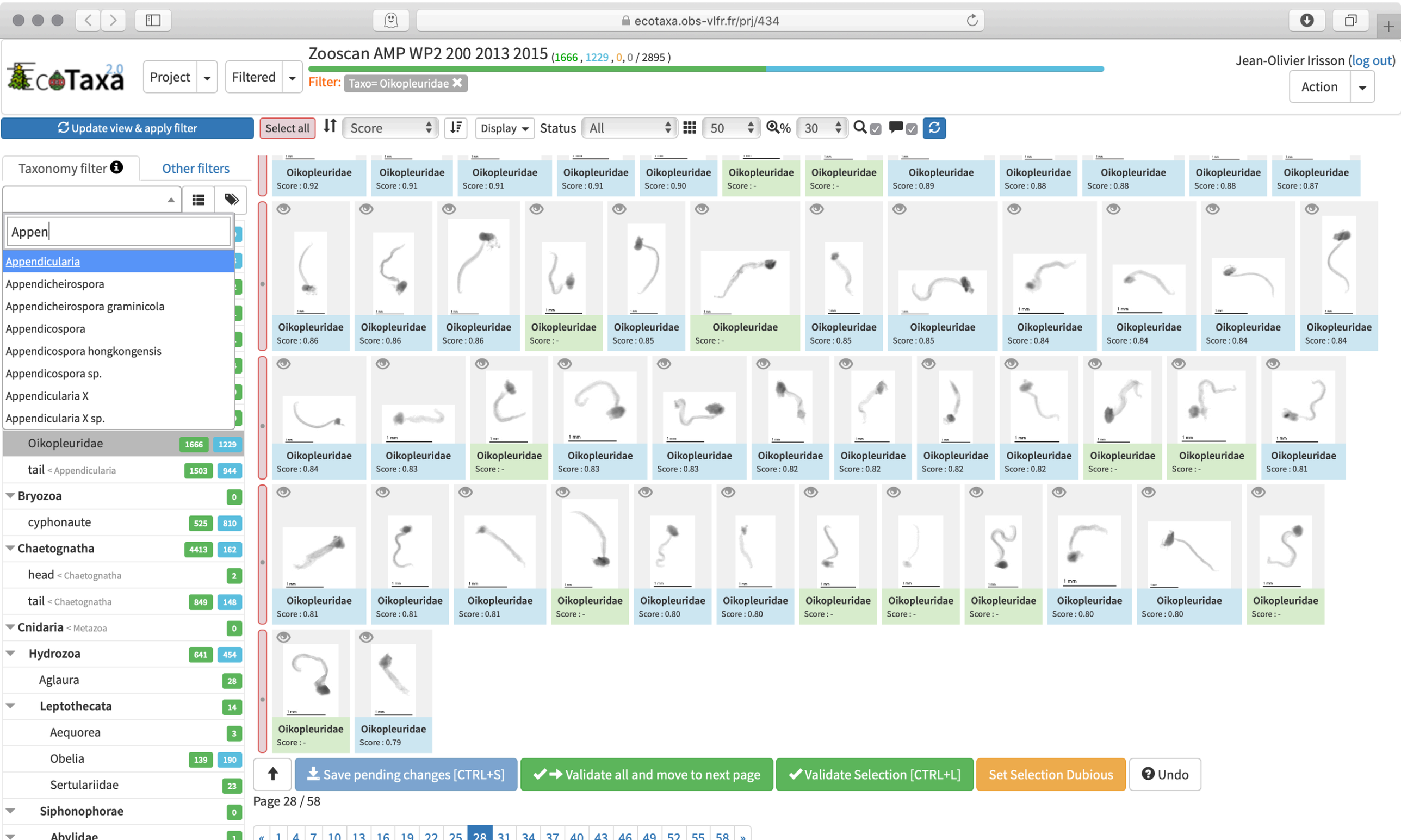
020

How can we process this amount of data?

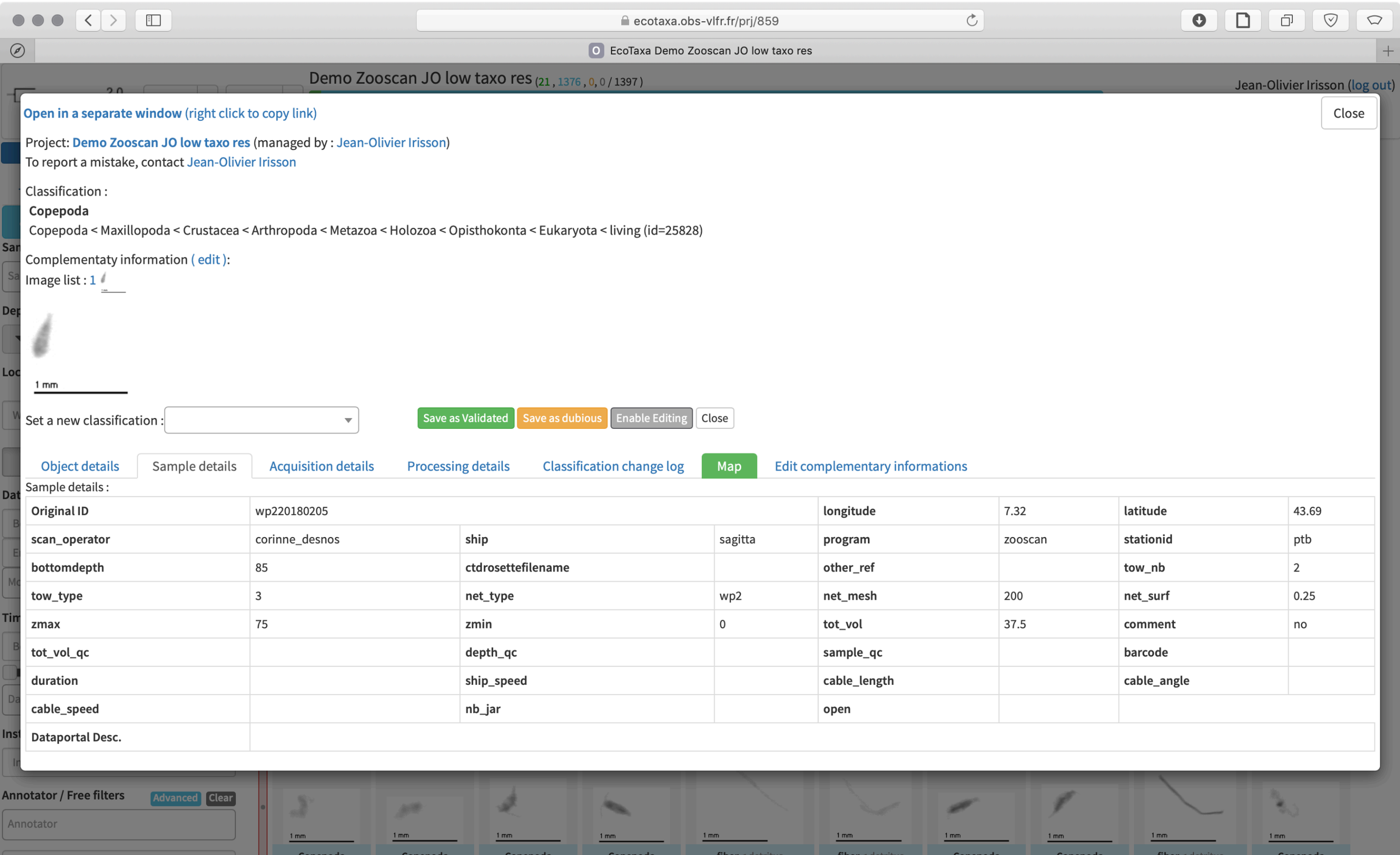
The quantitative imaging process



Web-based infrastructure for the classification of ecological images



Web-based infrastructure for the classification of ecological images



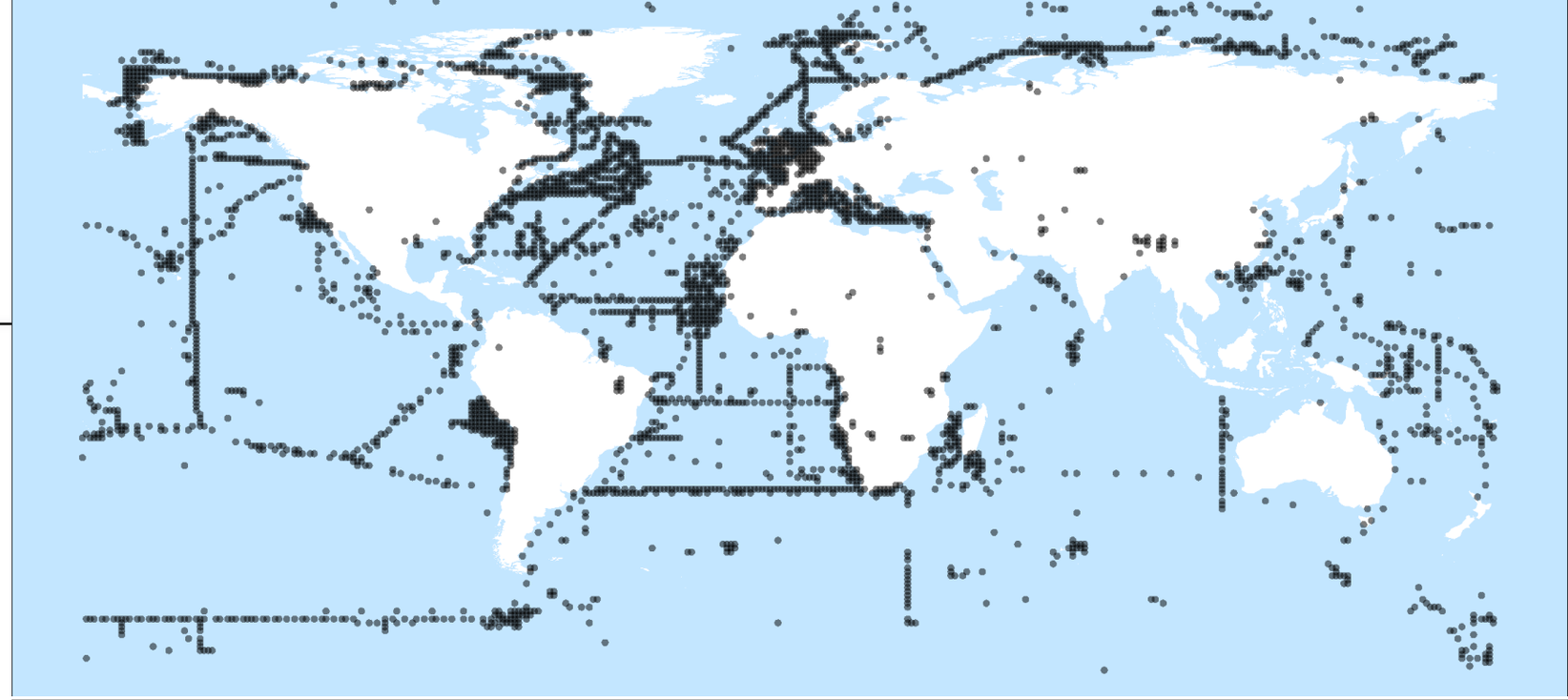
Eco**Taxa today**

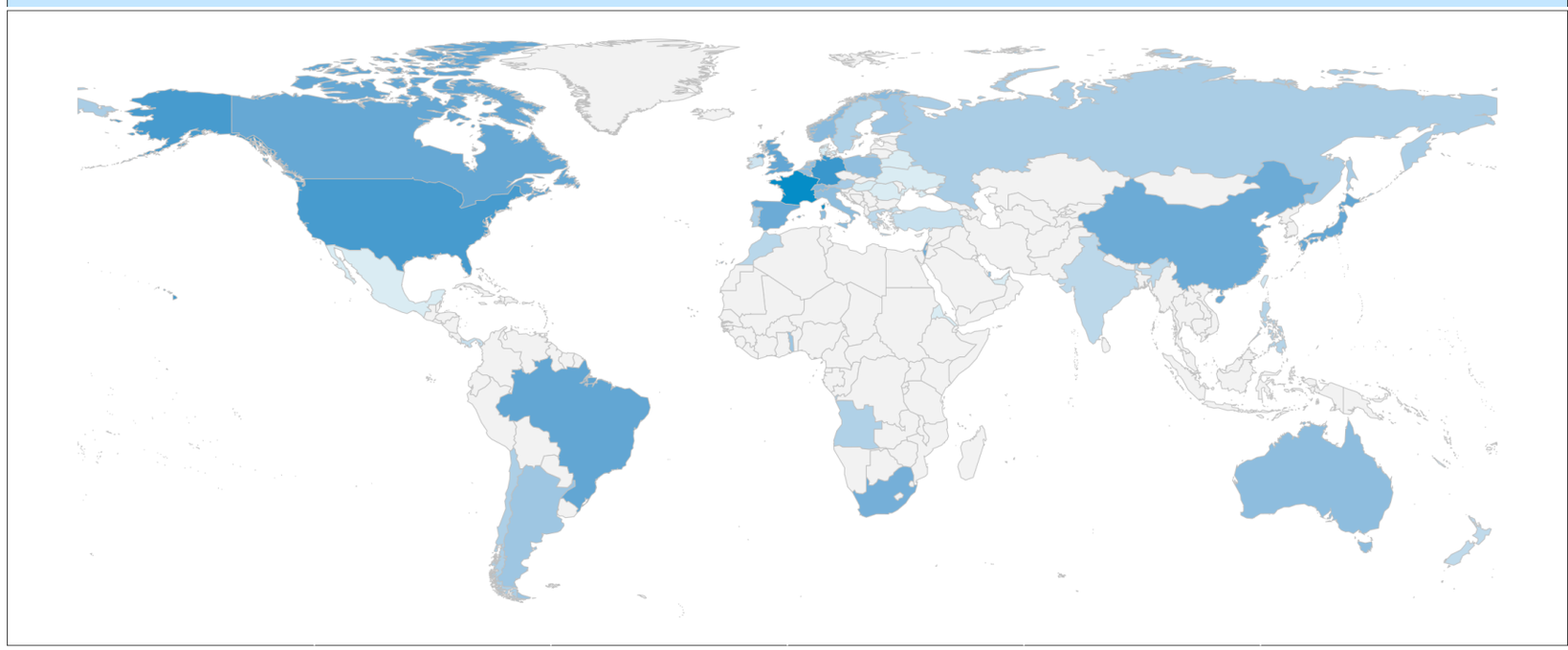
213M objects, 88M human-validated (41%)

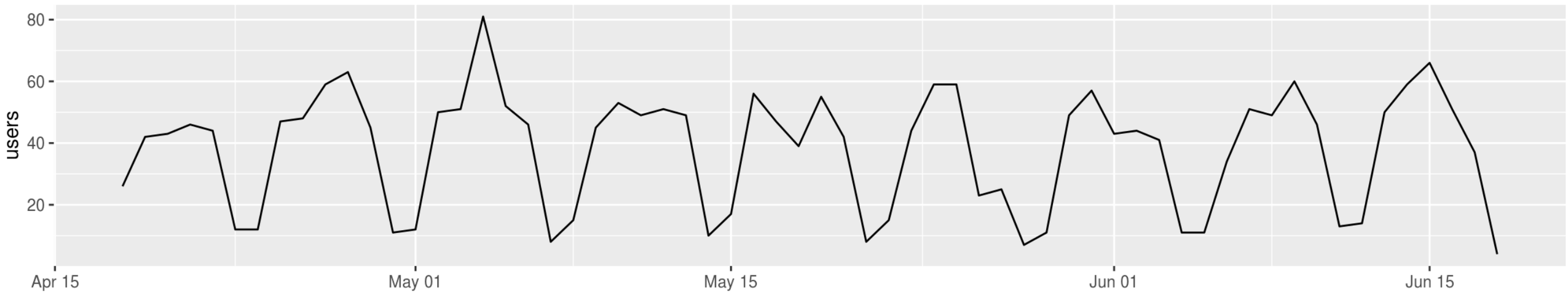
>100k locations in the world's oceans

1600 users, **500 organisations**, from various countries

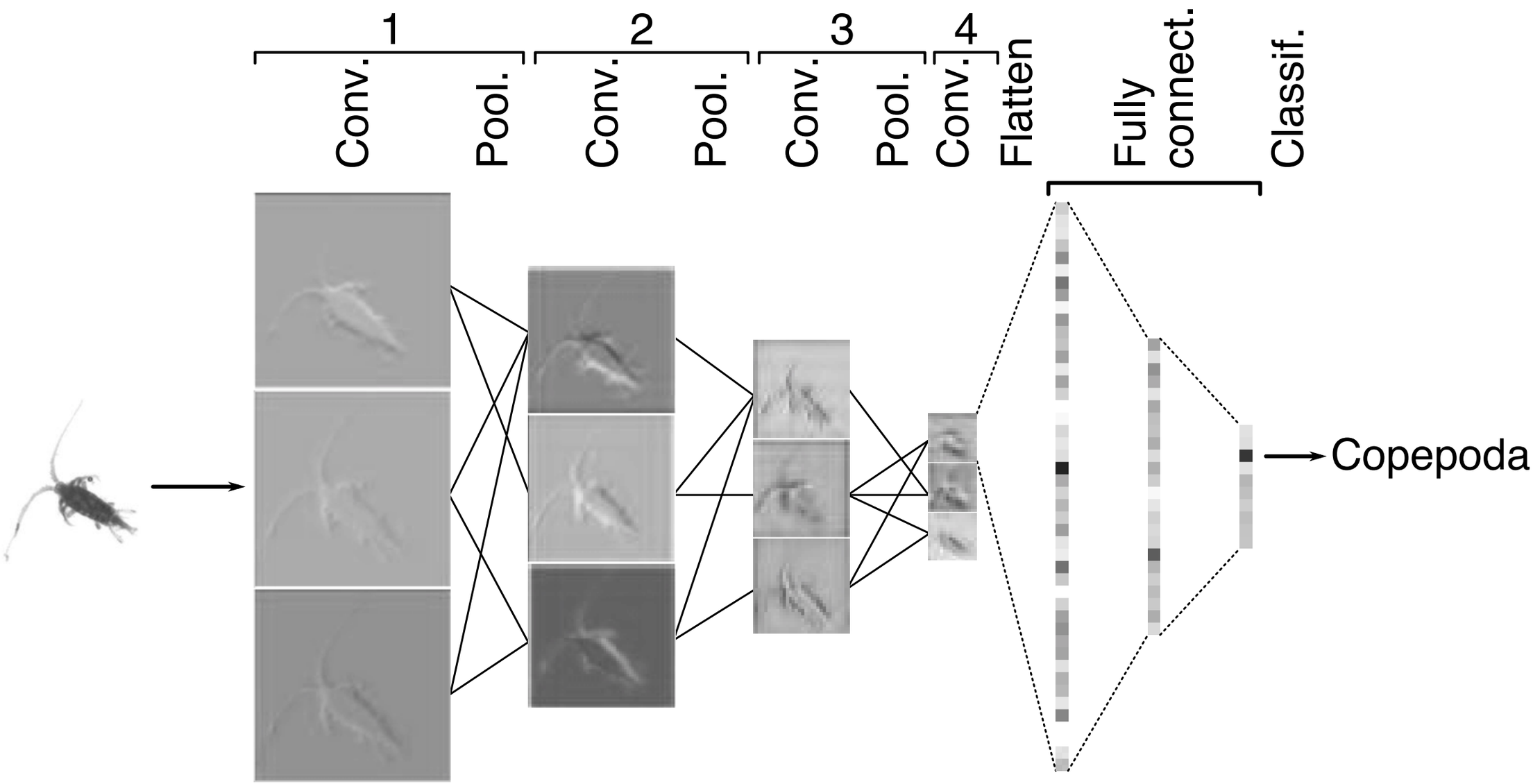
50 concurrent users at all times



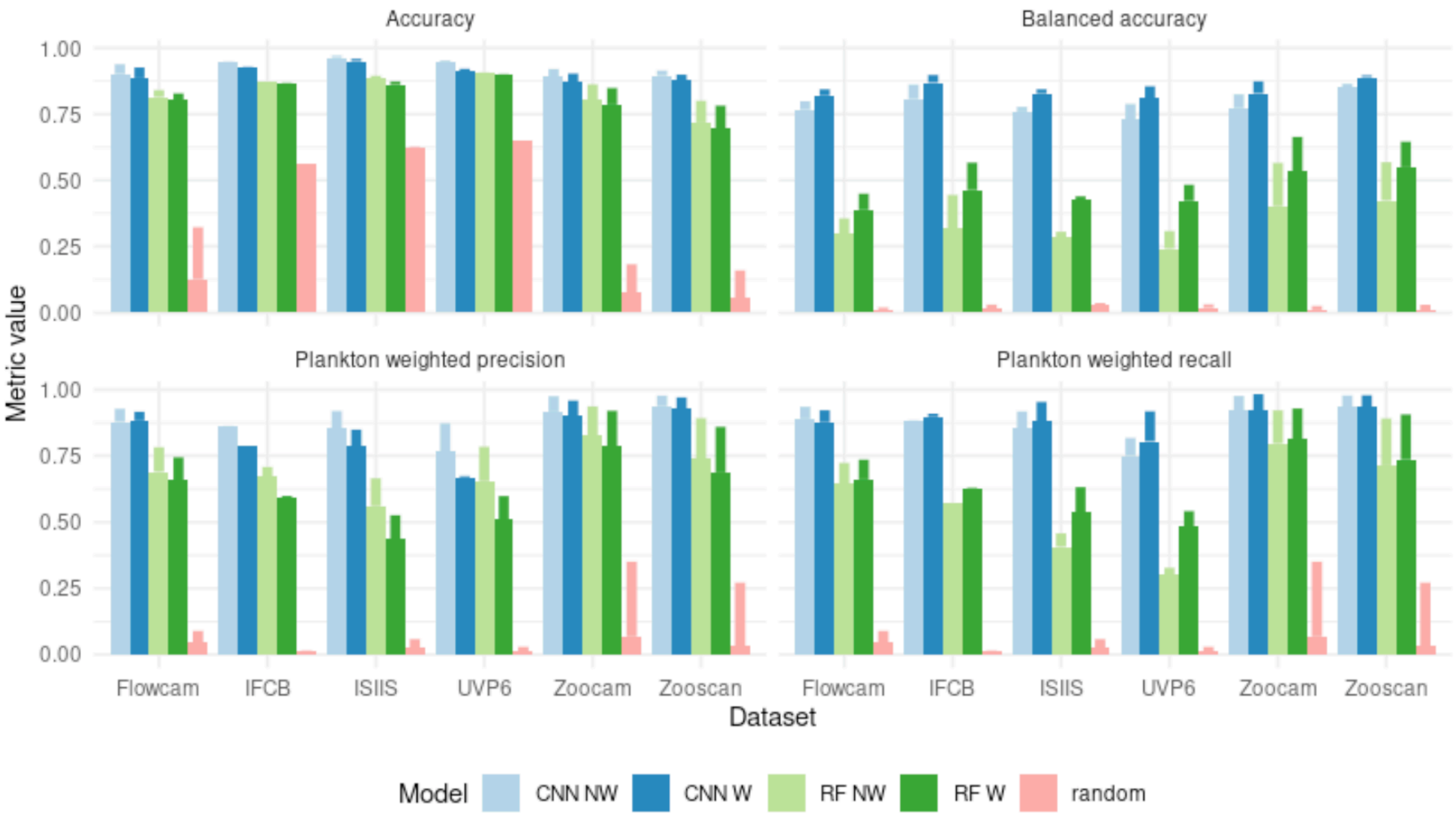




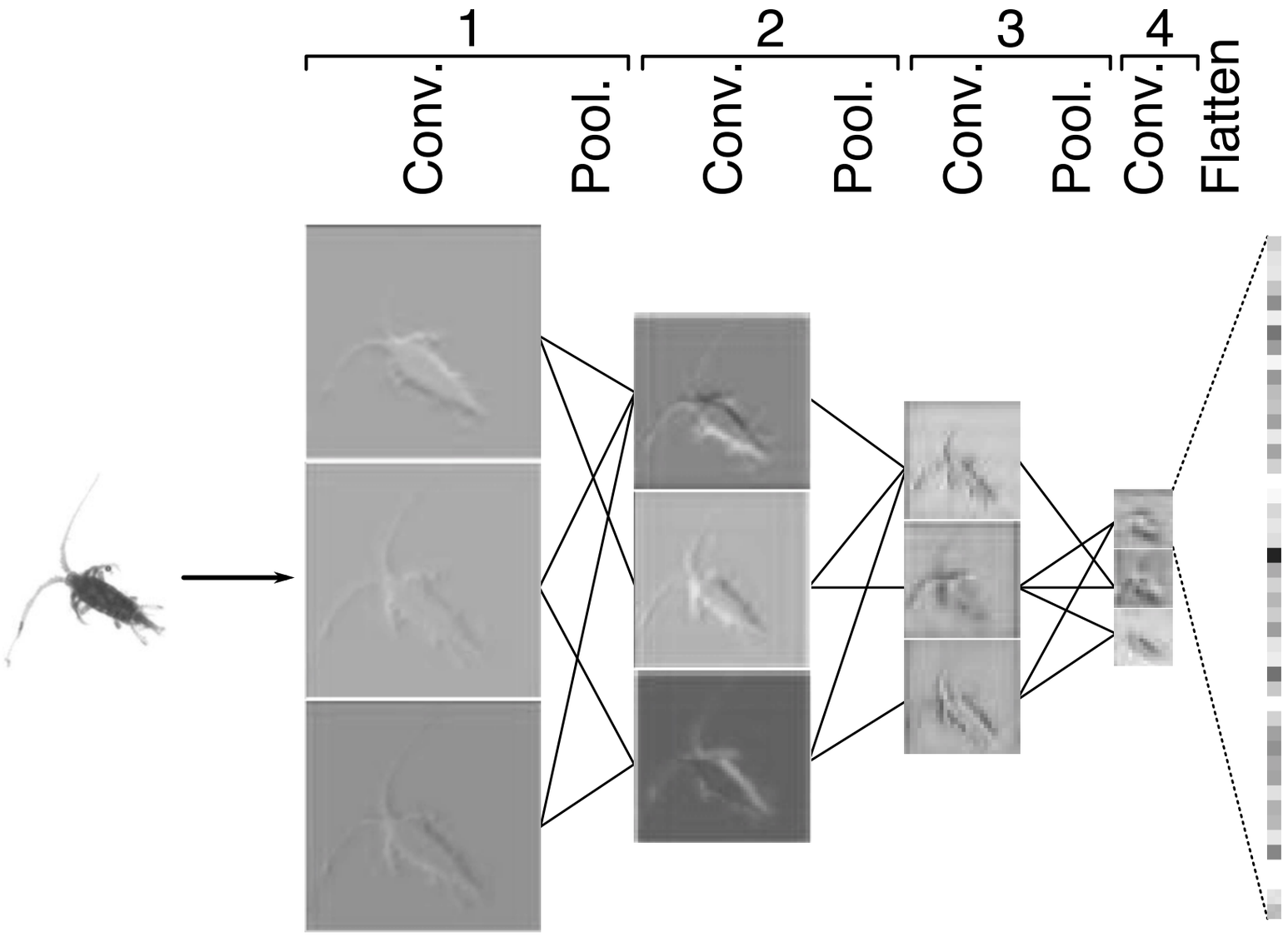
Improved Machine Learning backbone



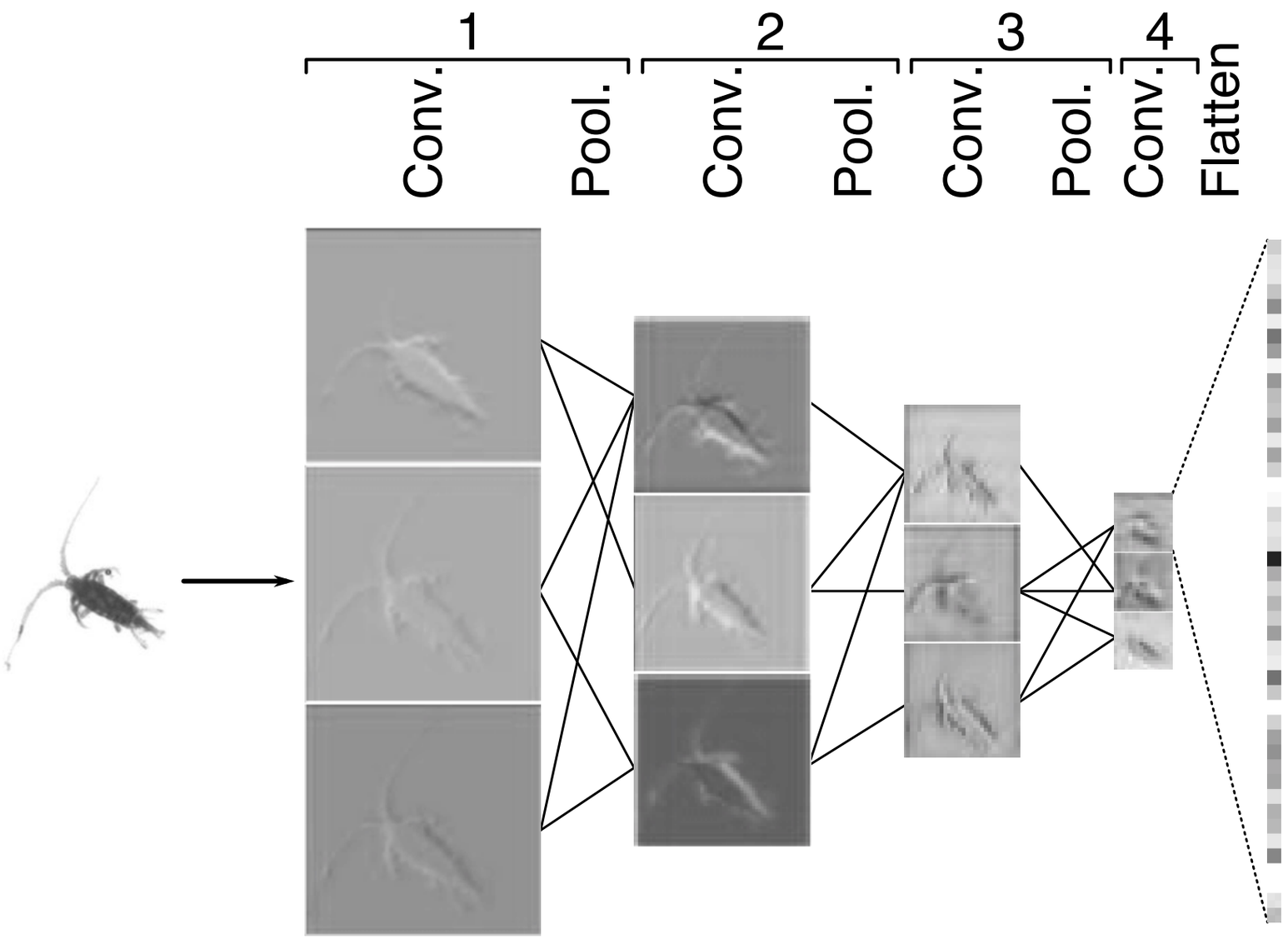
Improved Machine Learning backbone



Improved Machine Learning backbone



Improved Machine Learning backbone



Combine Convolutional Neural Networks with a fast Random Forest classifier

Application Programming Interface (API)

Programmatically

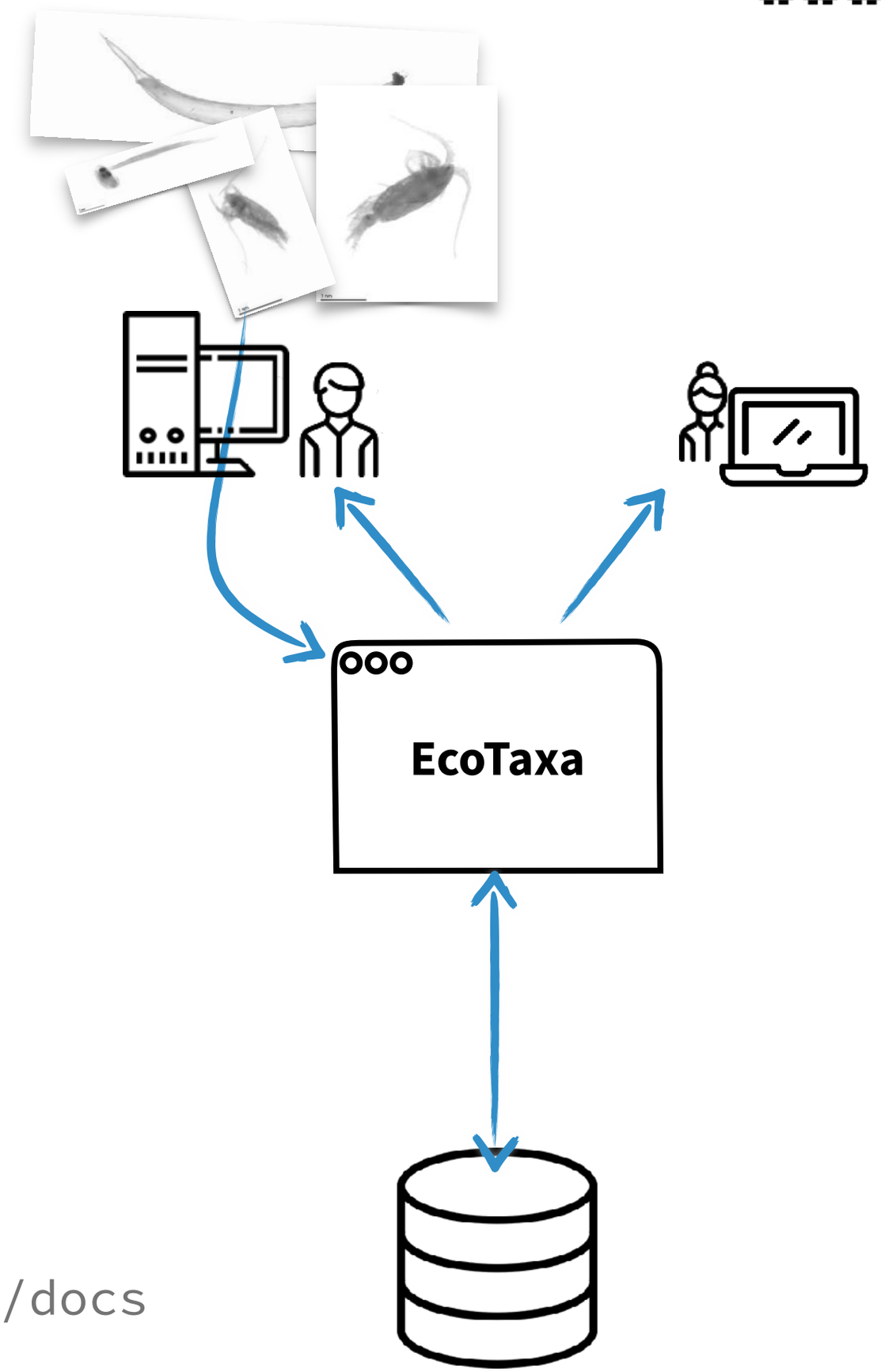
search projects, samples, users, taxa

import data

query subsets of data

classify images automatically

export datasets, in particular to external databases



`https://ecotaxa.obs-vlfr.fr/api/docs`

Application Programming Interface (API)

Programmatically

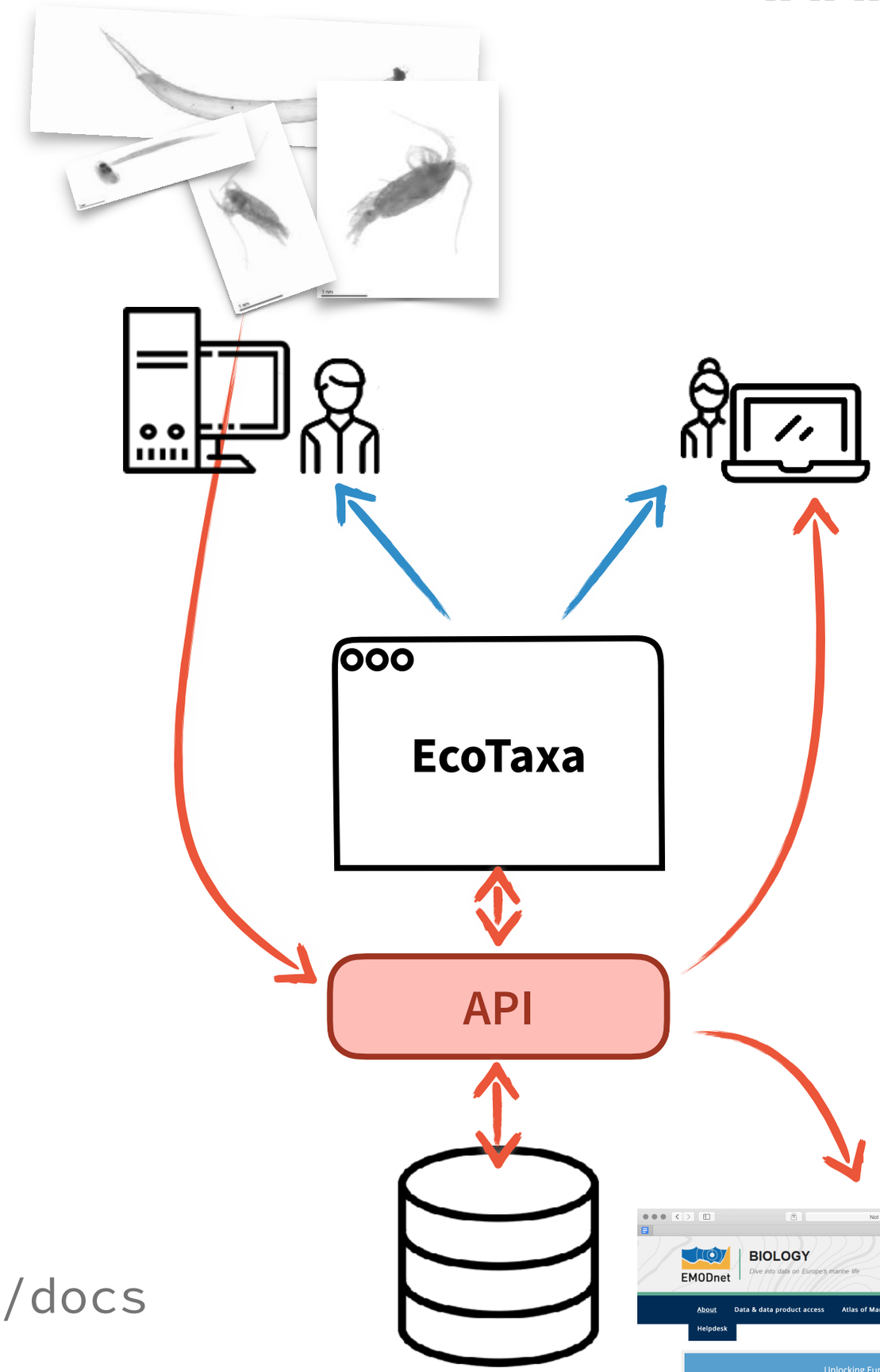
search projects, samples, users, taxa

import data

query subsets of data

classify images automatically

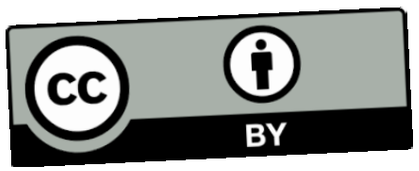
export datasets, in particular to external databases



`https://ecotaxa.obs-vlfr.fr/api/docs`

Export open data as DarwinCore Archive

Add **license** to the data sets, favouring open licenses



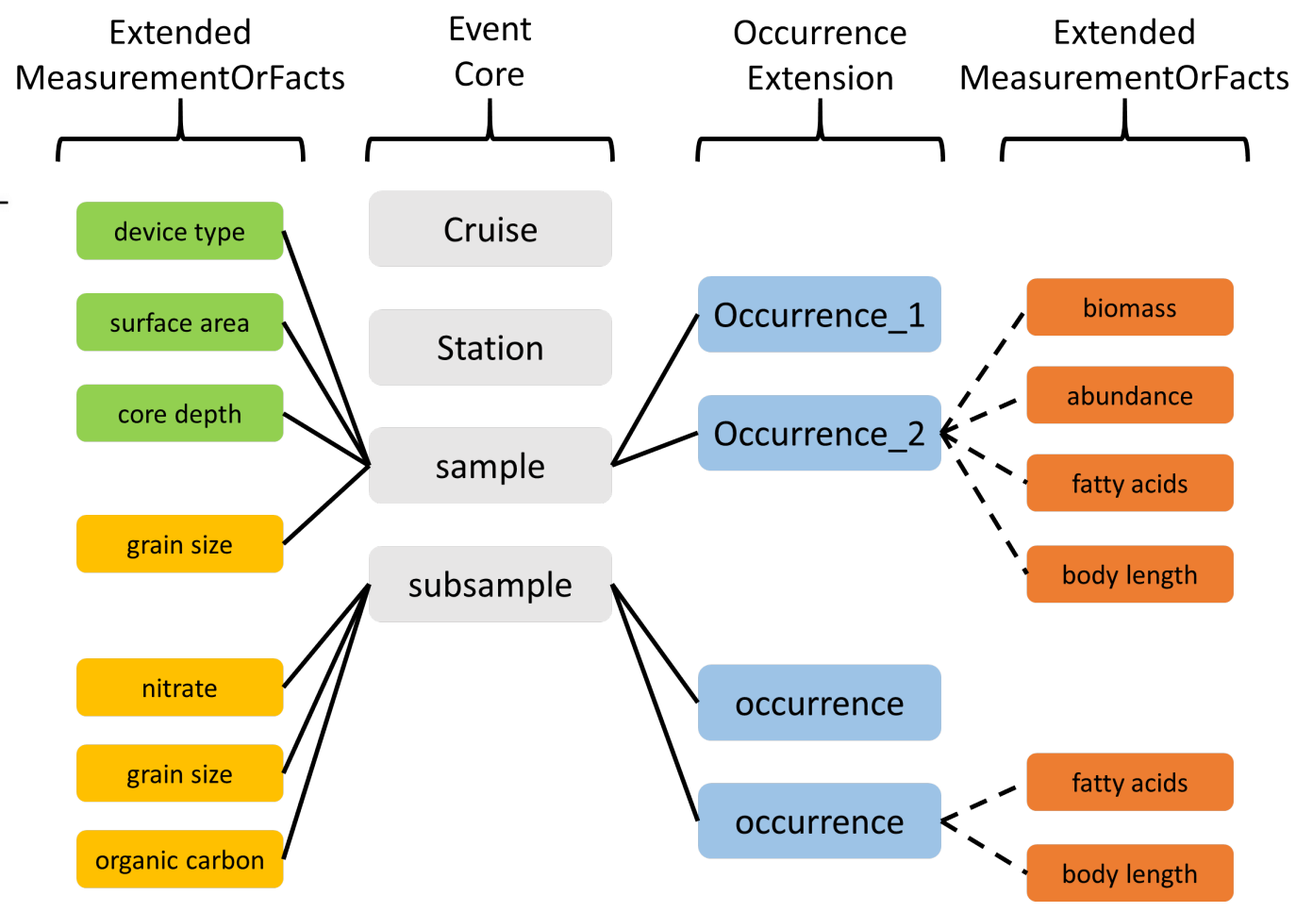
DarwinCore Archive is

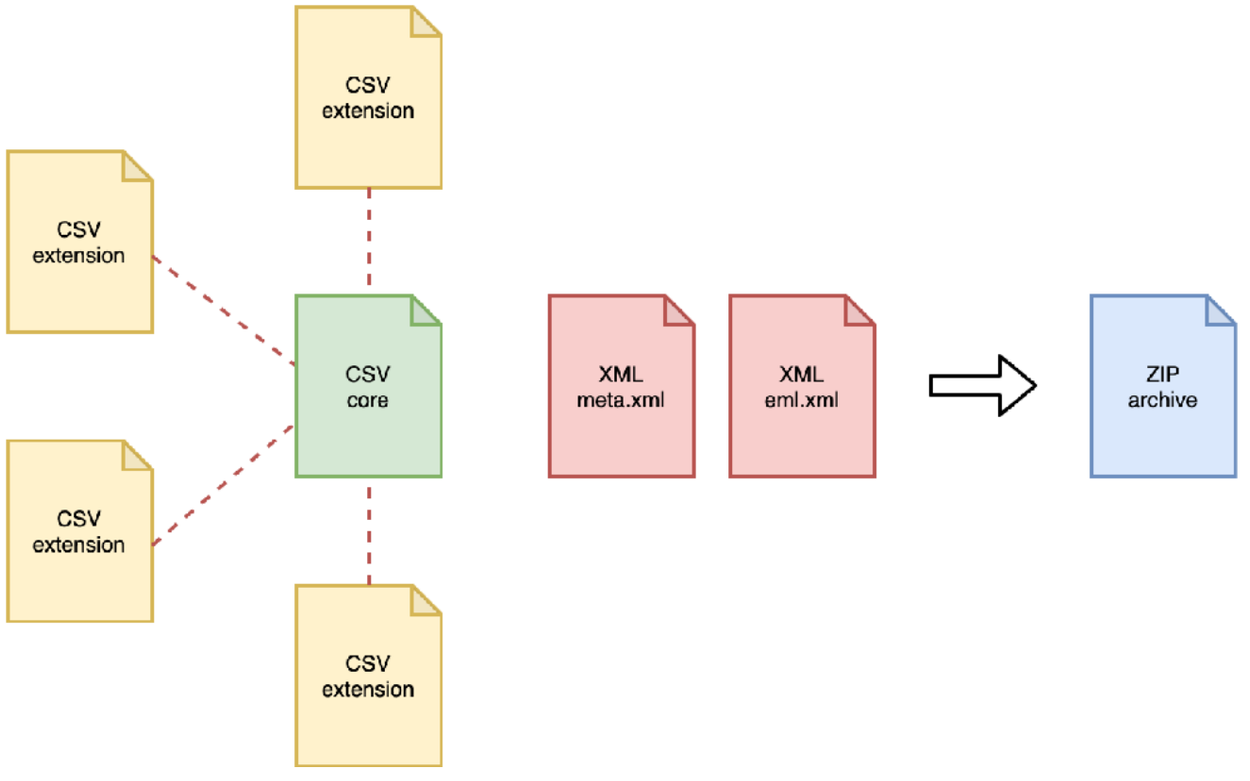
The **standard** format for biodiversity information

Organised **hierarchically**, like EcoTaxa

Very **descriptive** but **complex** to produce

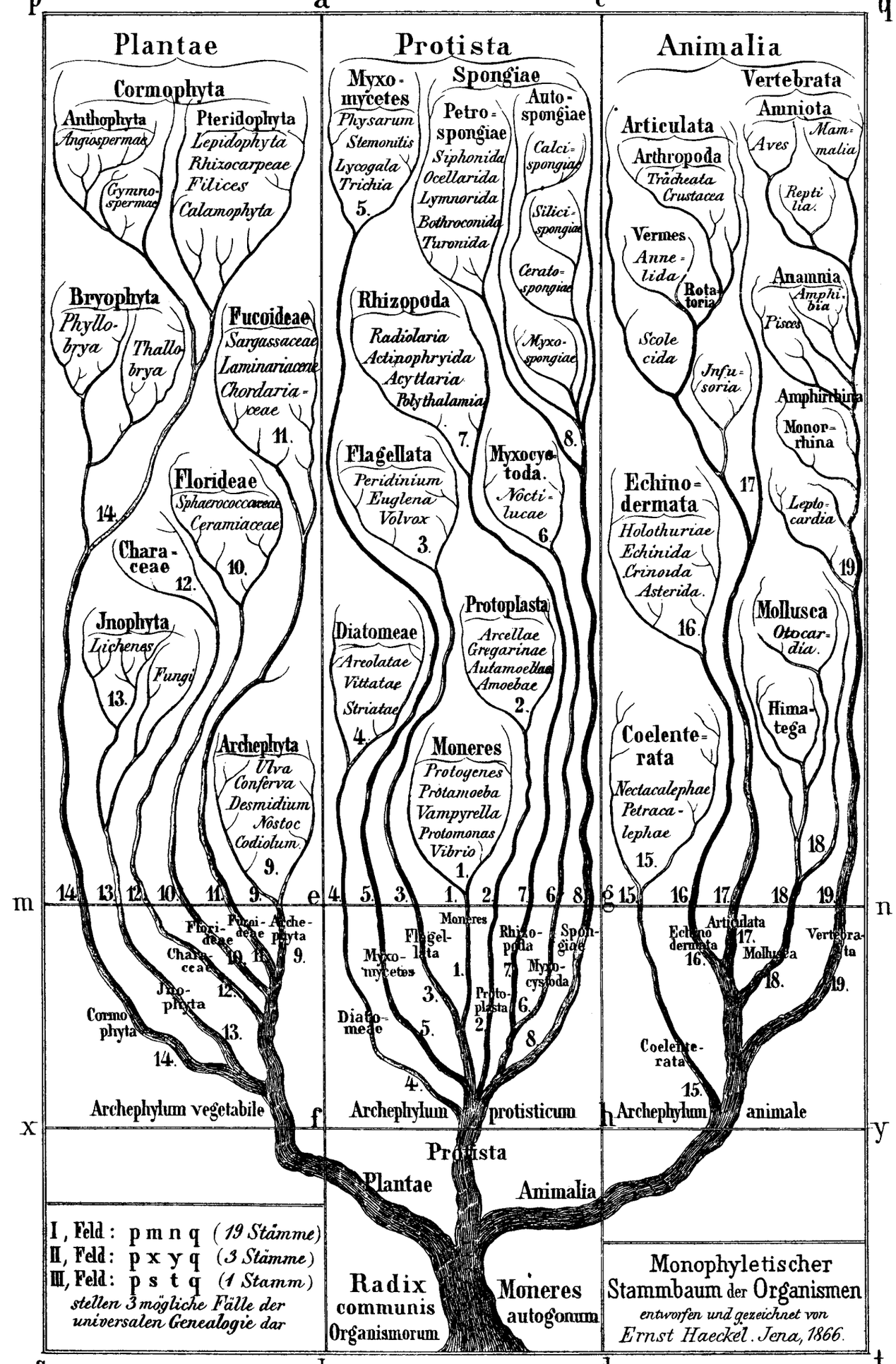
EcoTaxa helps users create DwCA, hence making their data more **FAIR**





Difficulties

Choosing a **universal taxonomy**



Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely

Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely

Deploying on **public infrastructures**

Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely

Deploying on **public infrastructures**

Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely

Deploying on **public infrastructures**

Finding (good) **programmers** !

Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely

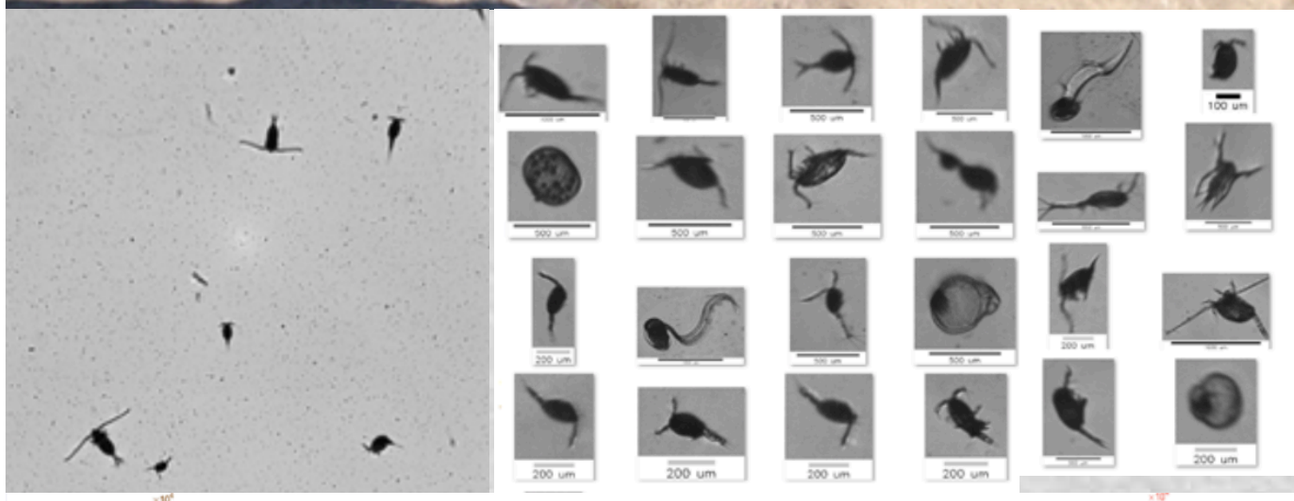
Deploying on **public infrastructures**

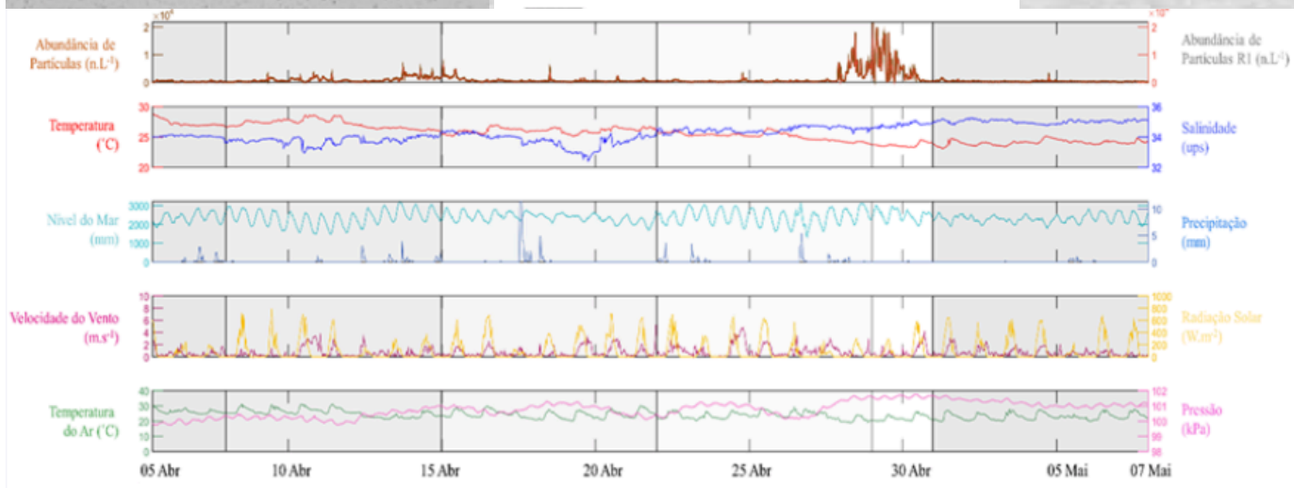
Finding (good) **programmers** !

Video conferencing does not solve a 12h **time difference**

Use 1: Environmental monitoring

High resolution **time series** at coastal sites, to detect **invasive** or **harmful** species





Use 1: Environmental monitoring

High resolution **time series** at coastal sites, to detect **invasive** or **harmful** species

Environmental **assessment** prior to exploitation at **deep sea mining** sites

#OzOceans2020 #SchoolsOutScience #VisioningCoralSea

ROV SuBastian Dive 373 (Part A) - Cairns Seamount, Australia - FK200802

527 views • Streamed live on 6 Aug 2020

Schmidt Ocean

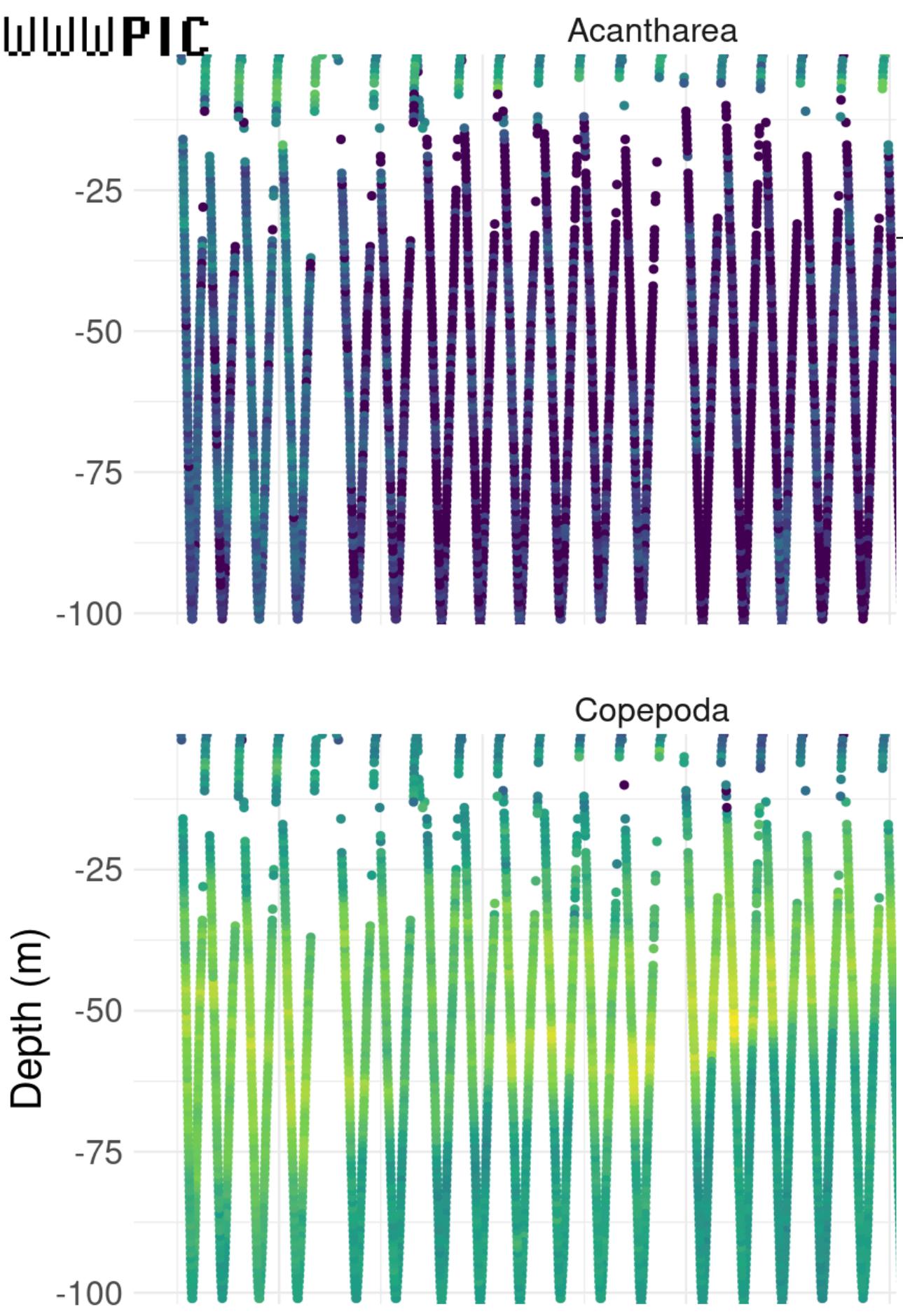
Join RV Falkor and ROV SuBastian for the fourth dive of the Seamounts, Canyons & Reefs of the Coral Sea expedition as we dive on Cairns Seamount, which is located in the Coral Sea Marine Park around 150 km east of Cairns. Here we'll be observing in high resolution, and strategically

SHOW MORE

Use 2: Biology studies at *sub*mesocale

Imager deployed on automated underwater **glider**

Capture one image **every 2 seconds for ~4 months** along a transect

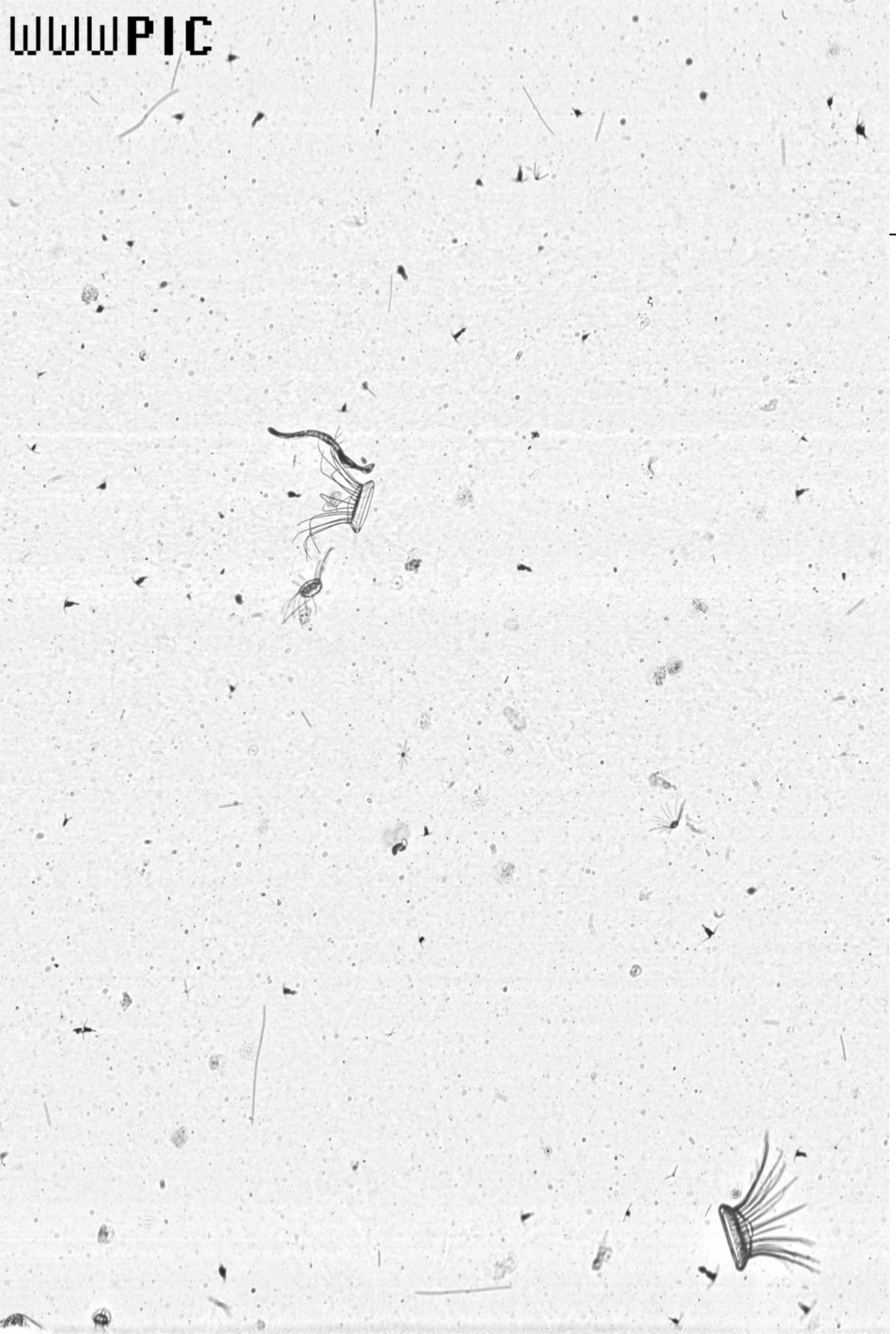


Use 2: Biology studies at *sub*mesocale

Imager deployed on automated underwater **glider**

Capture one image **every 2 seconds for ~4 months** along a transect

Describe the **biology** on a front at the **same resolution** as the physics and chemistry



Use 2: Biology studies at *sub*mesocale

Imager deployed on automated underwater **glider**

Capture one image **every 2 seconds for ~4 months** along a transect

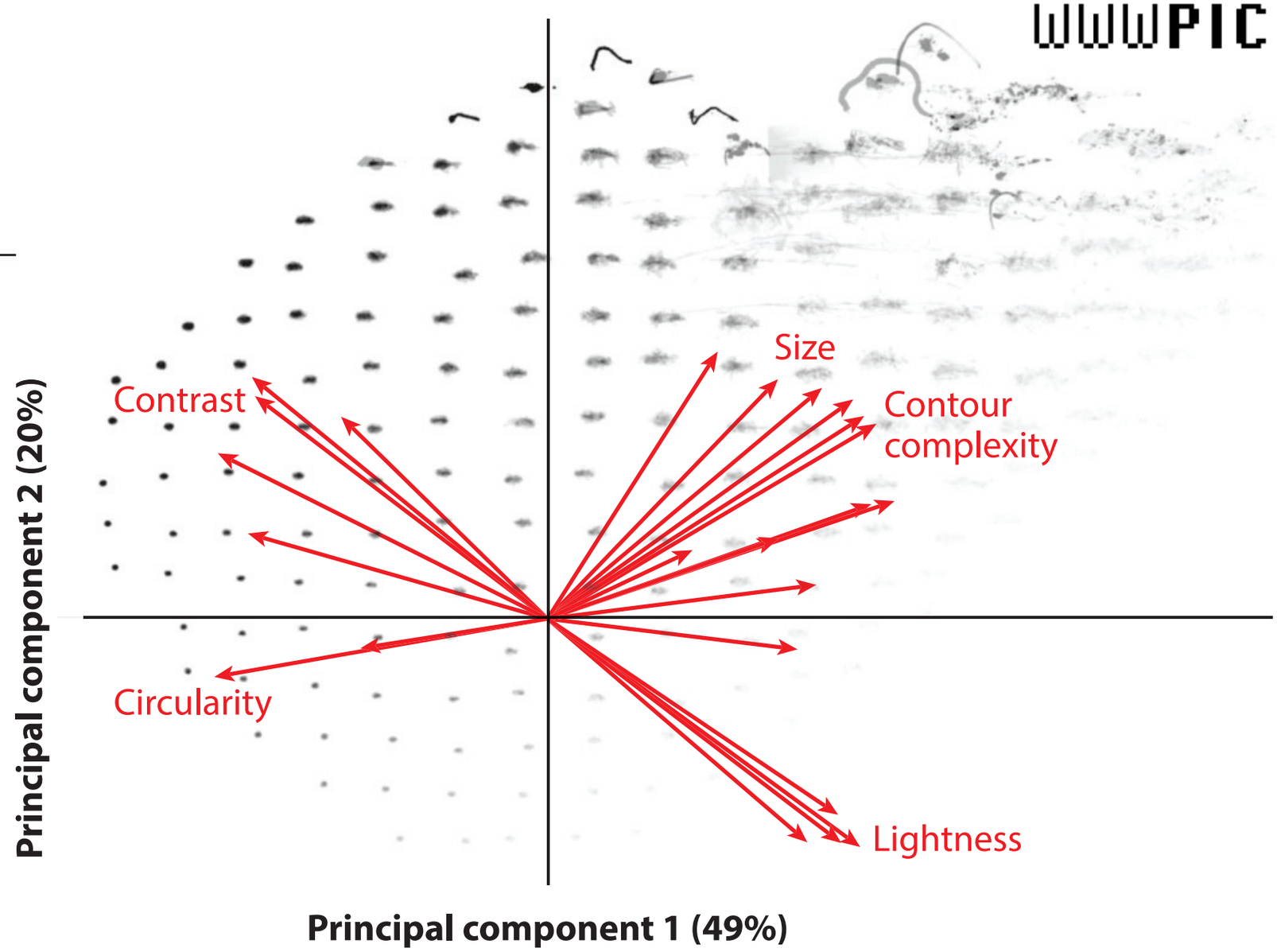
Describe the **biology** on a front at the **same resolution** as the physics and chemistry

Investigate **prey-predator** interactions at cm scale, in situ

Use 3: Carbon export

Worldwide database of ~40M **marine snow** images

Characterise their shape to better describe their **production** and **sinking**



Take home messages

Data management, curation and sharing is **just as important** as data collection

It is complex. Doing it well requires **time and effort**

Public **e-Infrastructures** are needed to serve the needs of environmental scientists

Machine learning is a great tool to **accelerate** (not replace) the work of human experts

https://site.wwwpic.net/