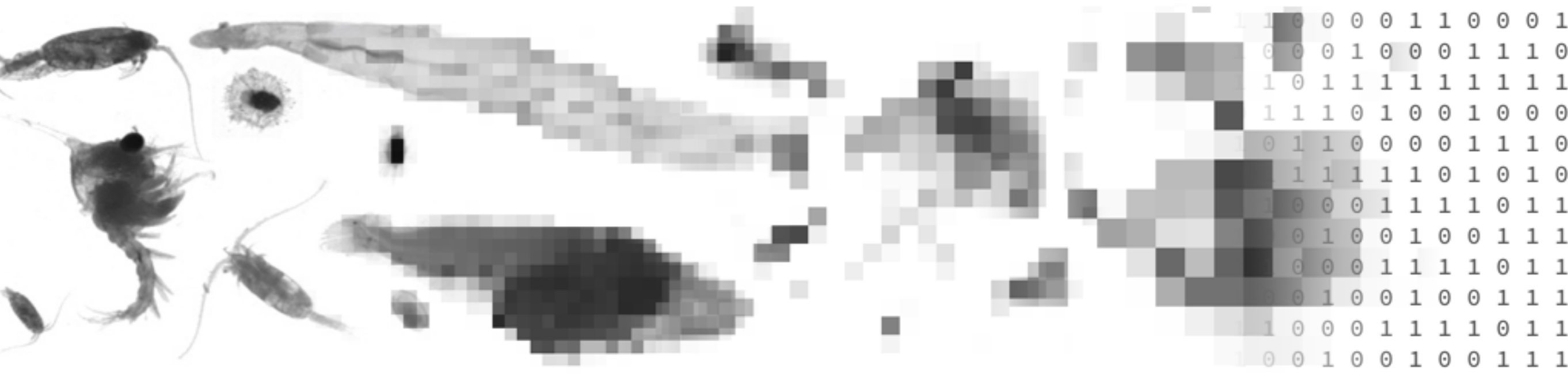


SRI Conference, 2022-06-21

Jean-Olivier Irisson (and <https://site.wwwpic.net/team>)

Observing plankton in the open ocean with quantitative imaging

From images to data



BELMONT
FORUM

SCIENCES
SORBONNE
UNIVERSITÉ

cnrs

Oregon State
University

USP

JAMSTEC

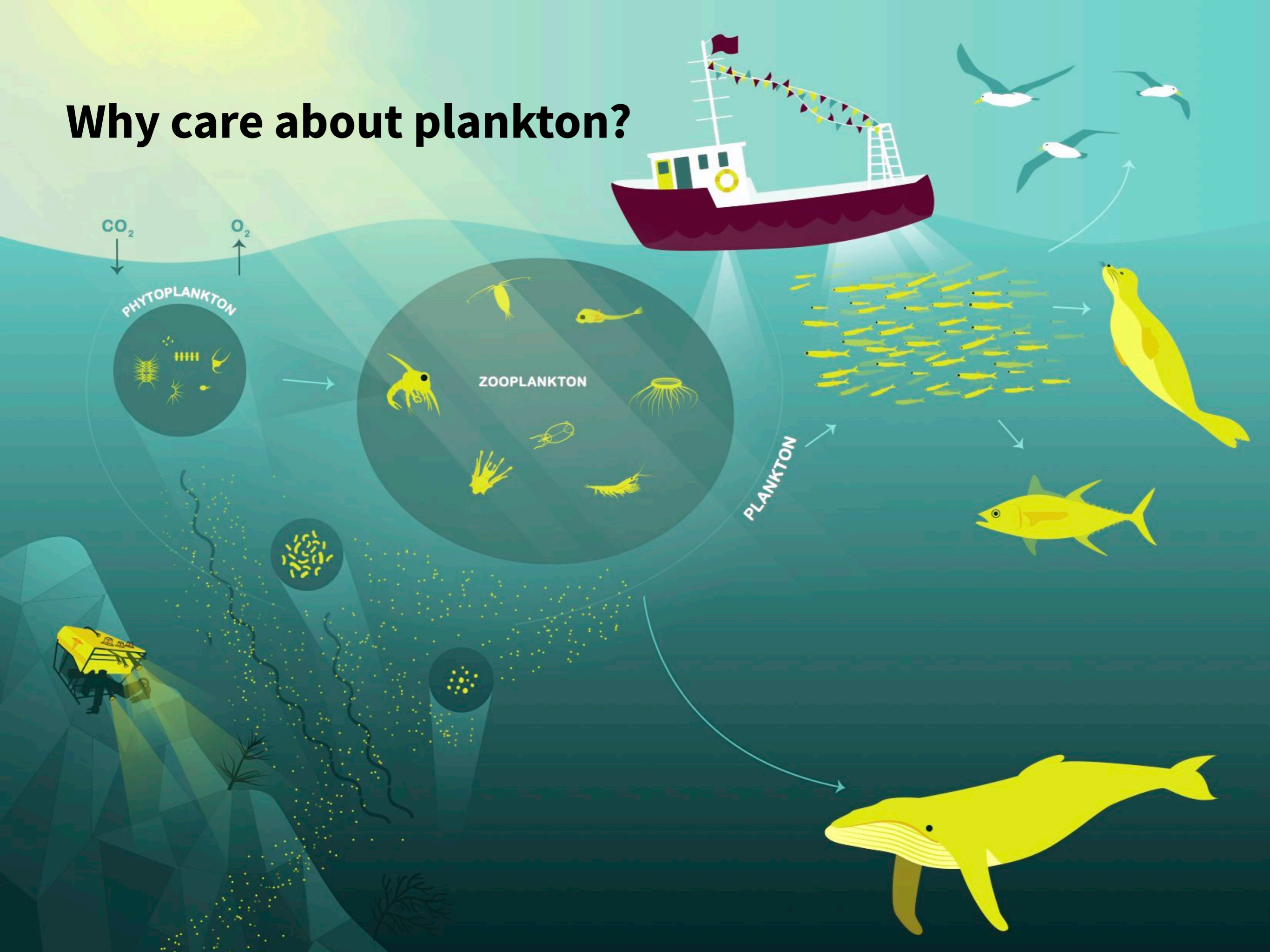
国立研究開発法人海洋研究開発機構

JAMSTEC

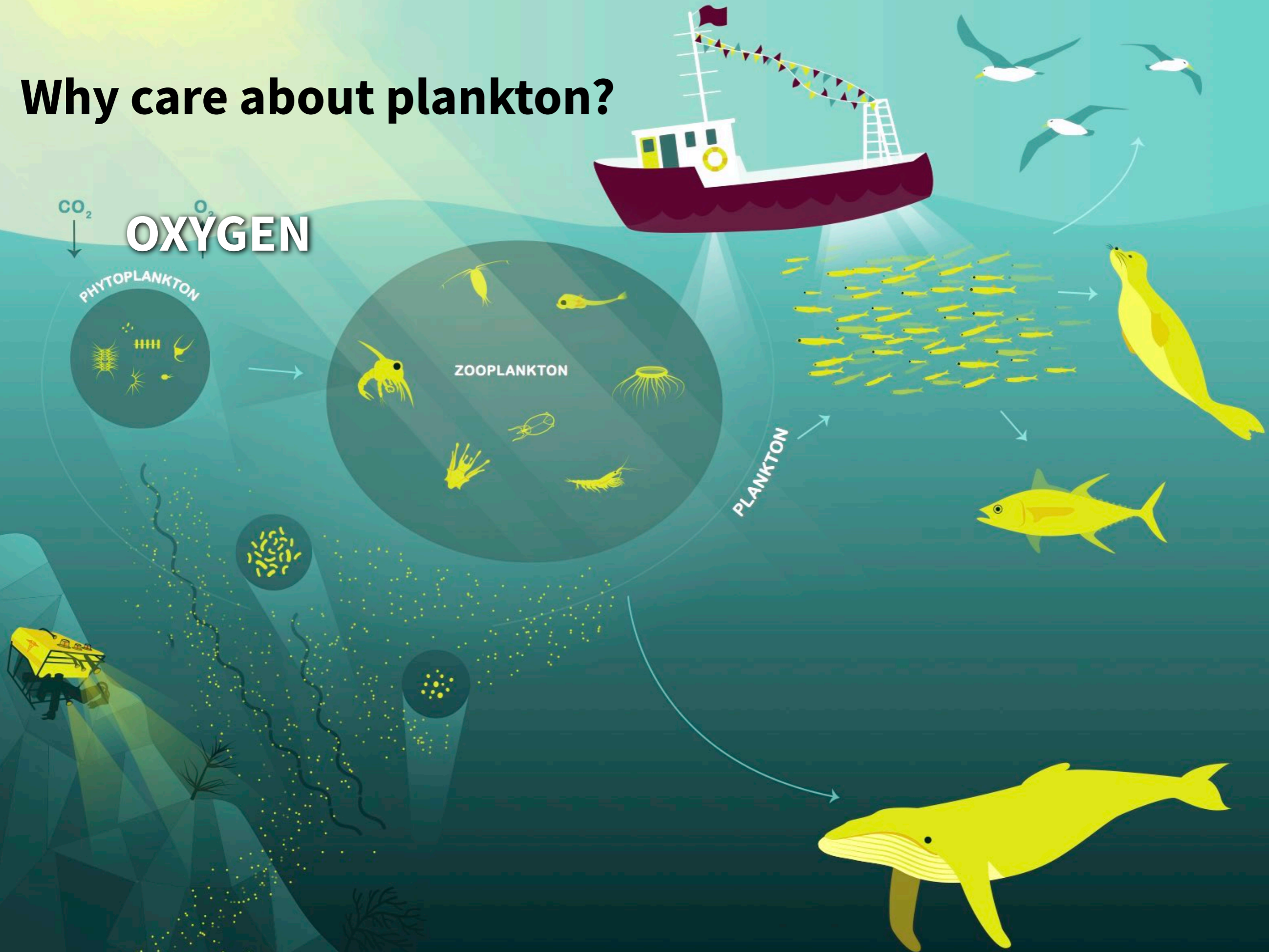
Japan Agency for Marine-Earth Science and Technology

WWWPIC

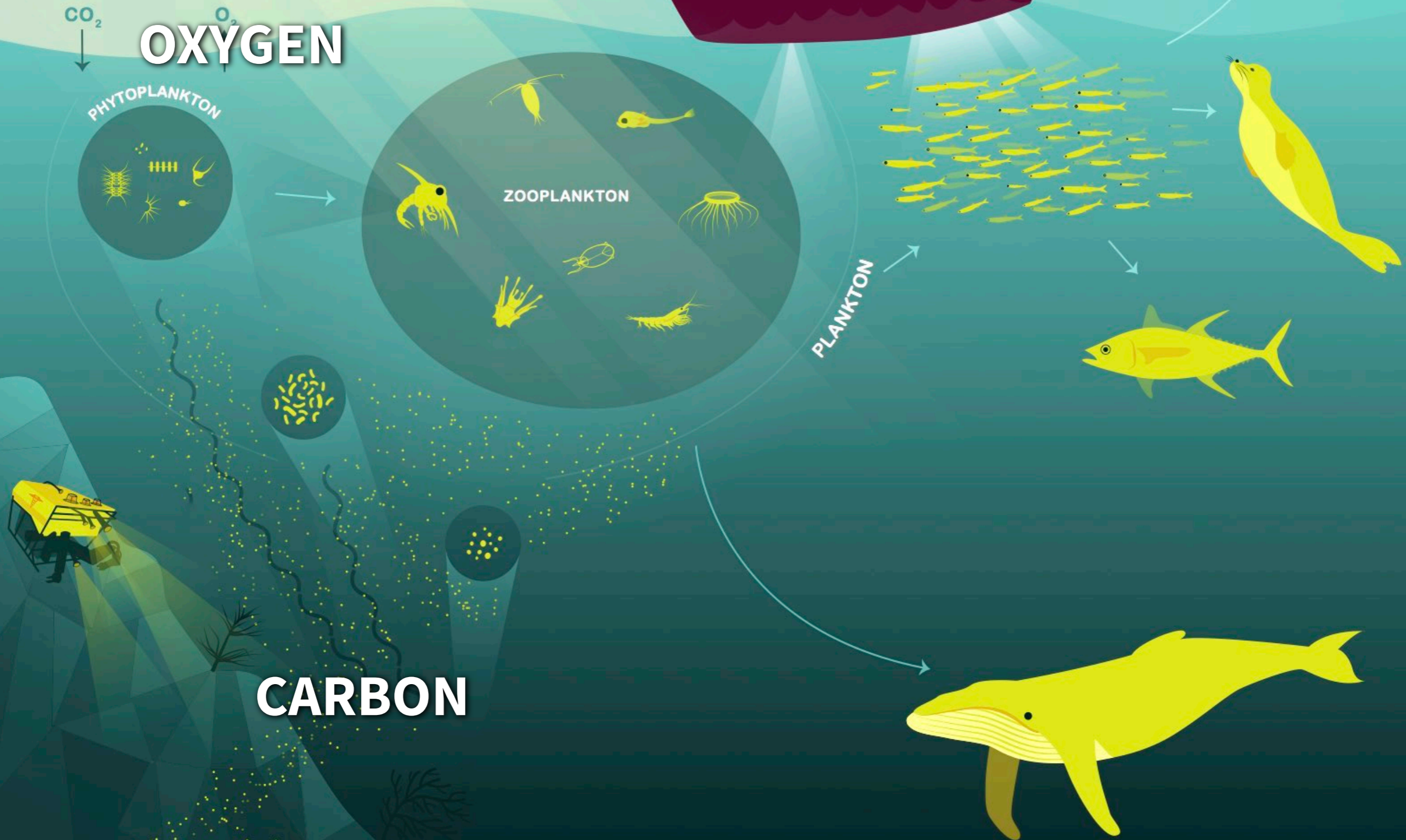
Why care about plankton?



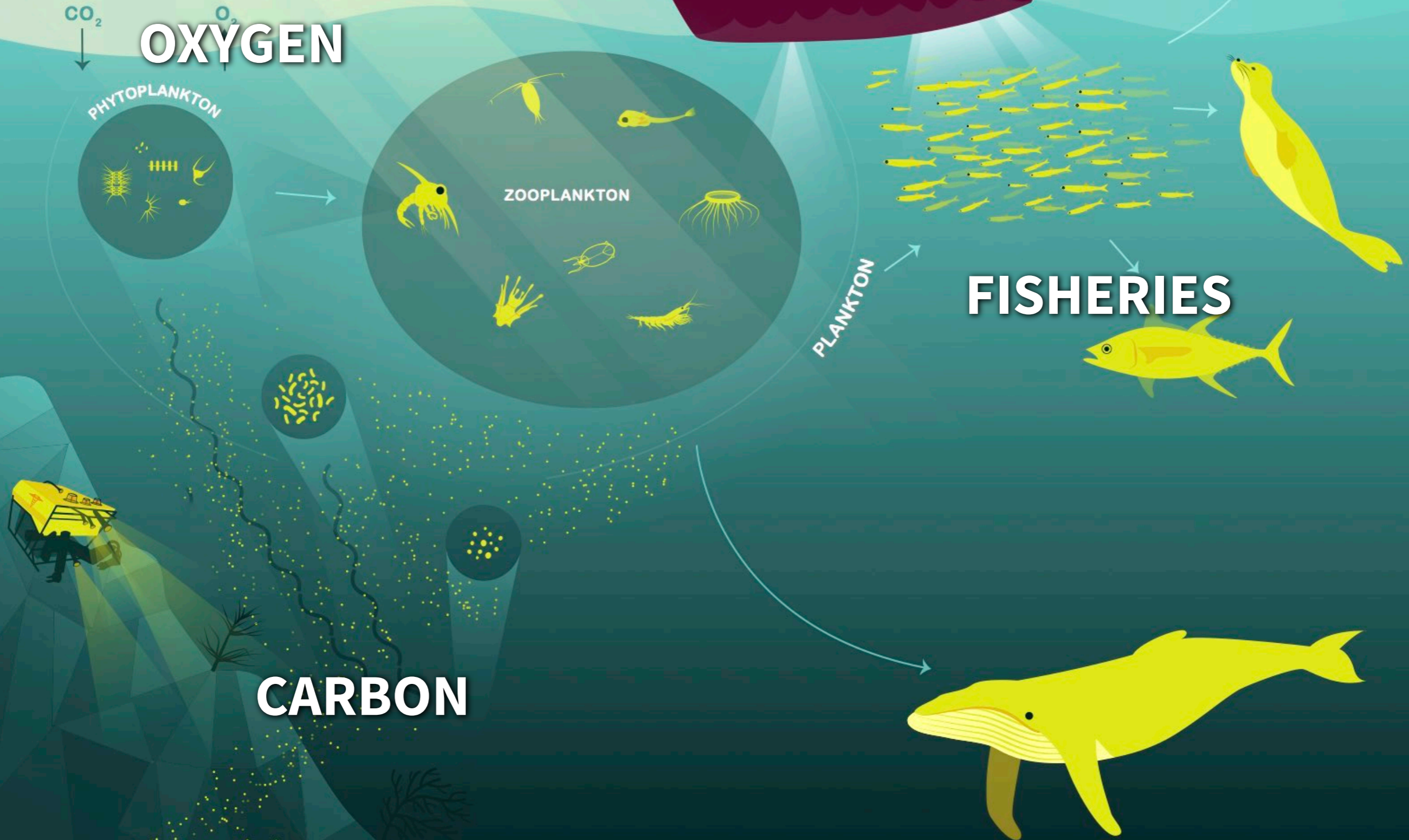
Why care about plankton?



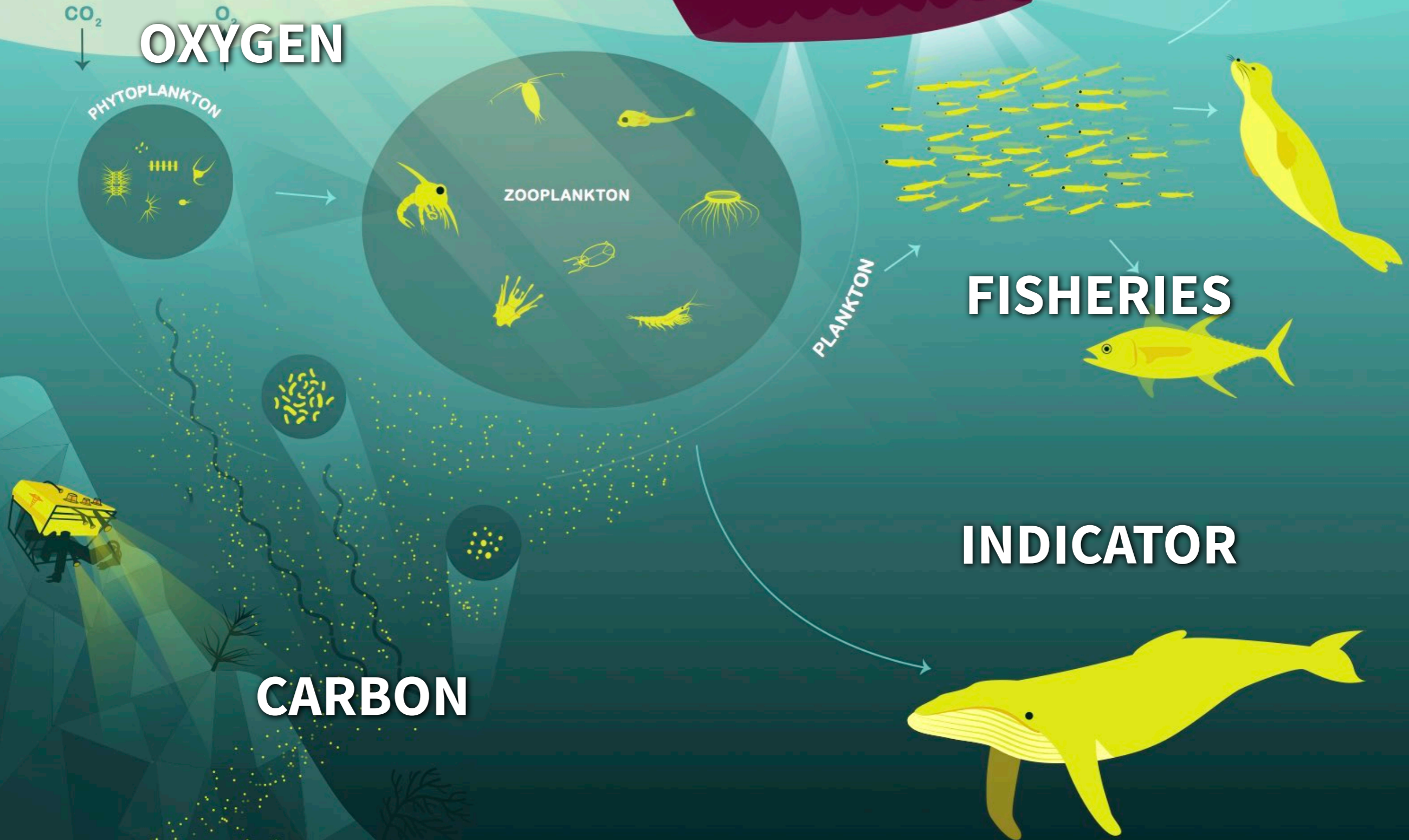
Why care about plankton?

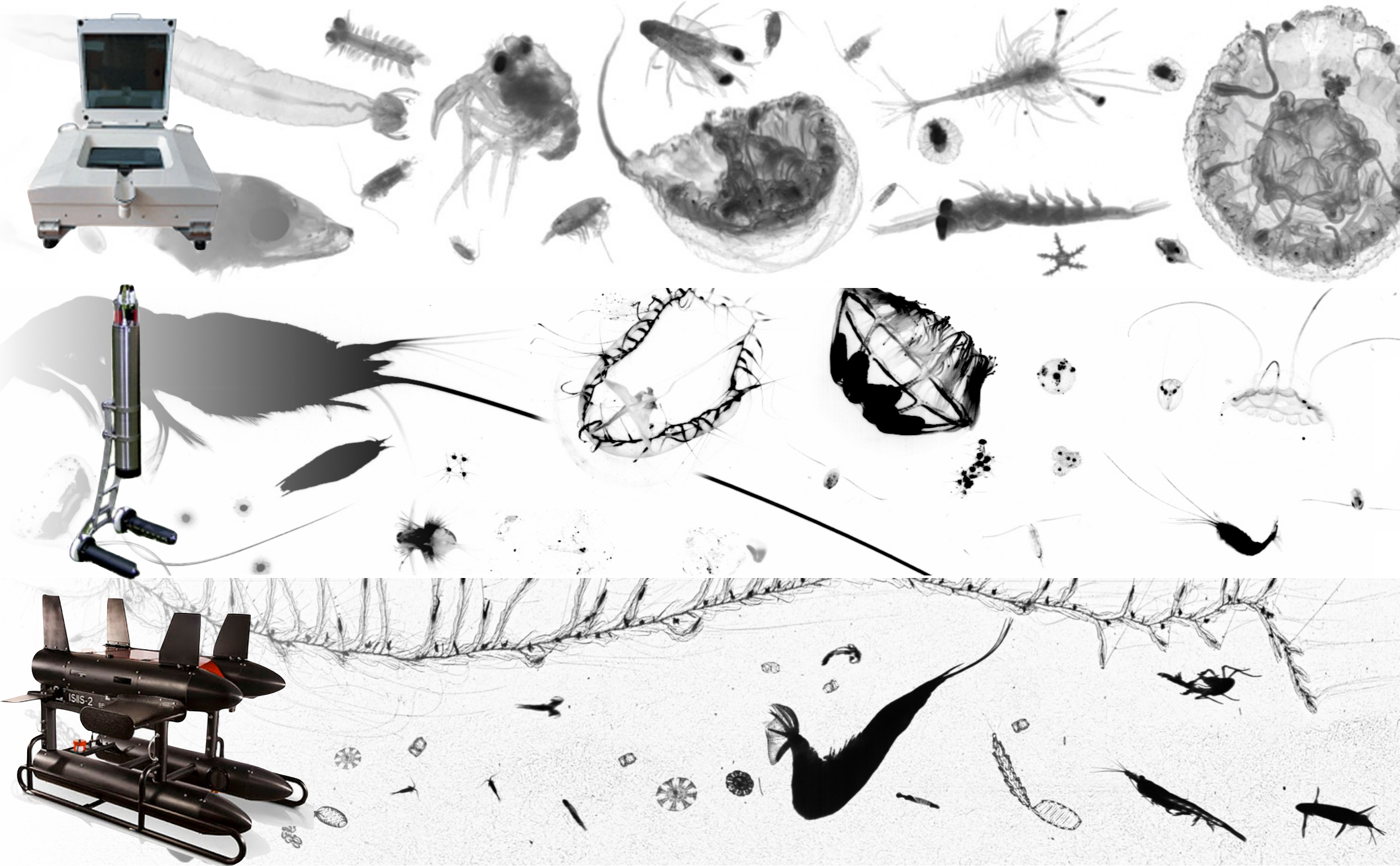


Why care about plankton?



Why care about plankton?

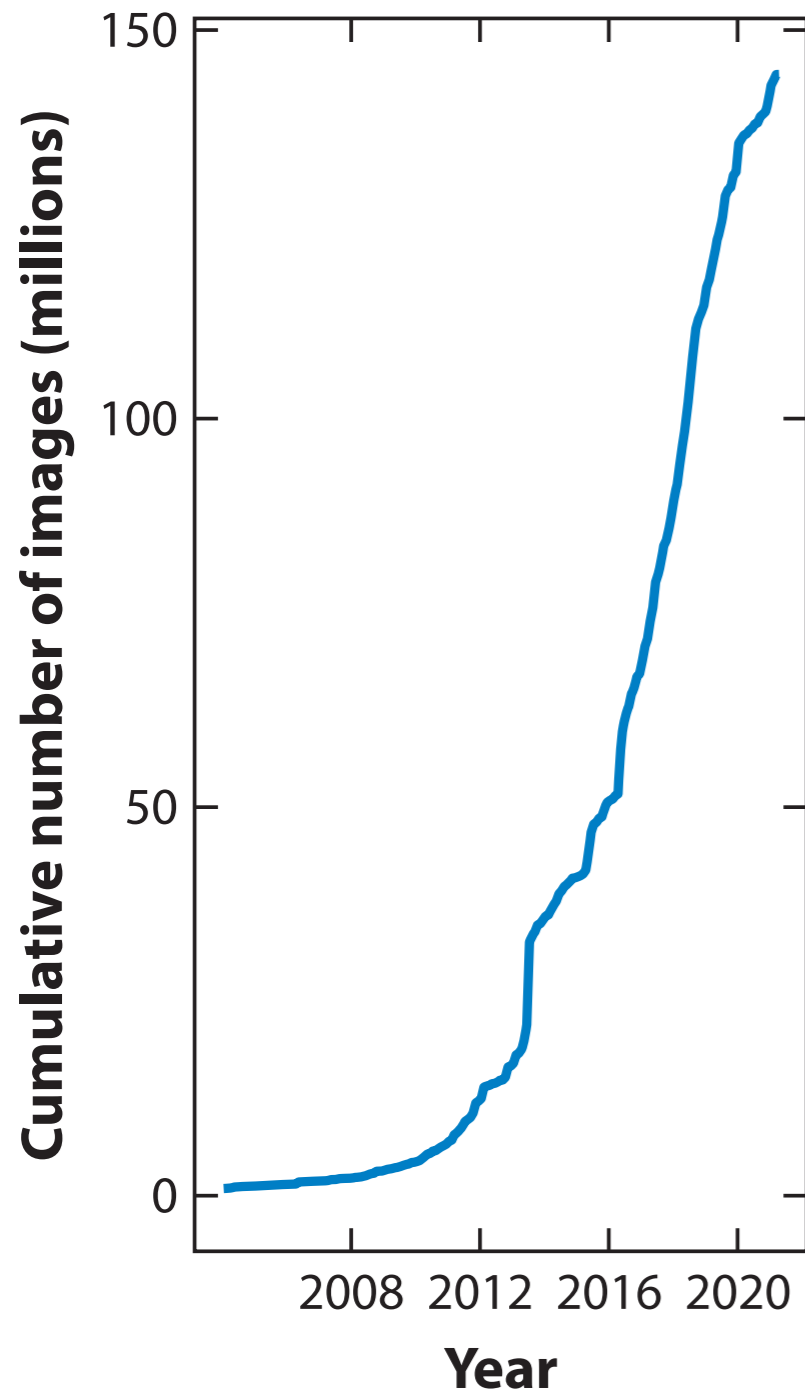




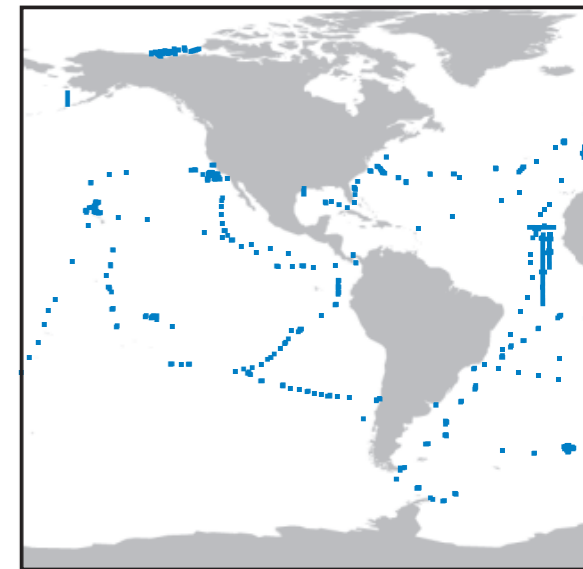
Quantitative plankton imaging

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

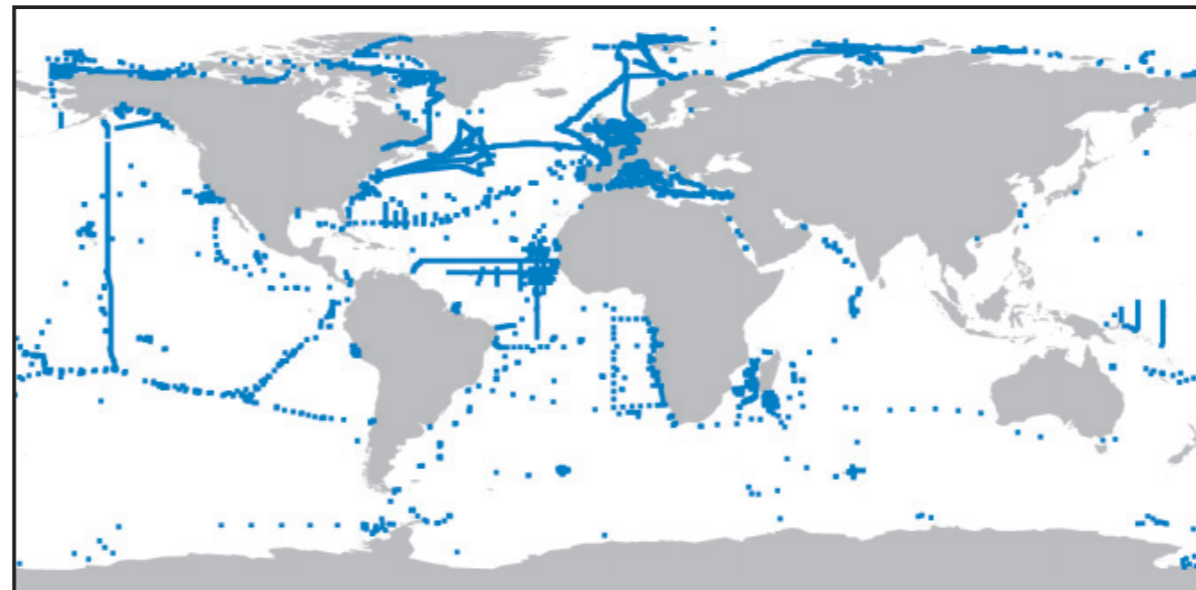
Steep growth in data acquisition



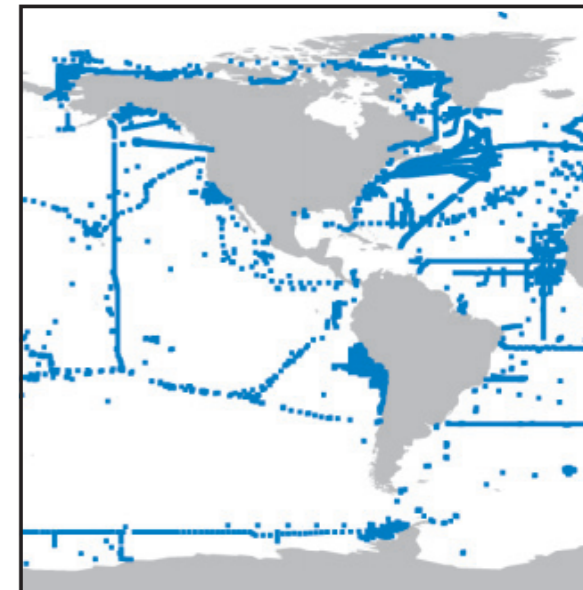
2008 (4,000 samples)



2012 (17,000 samples)



2016 (56,000 samples)

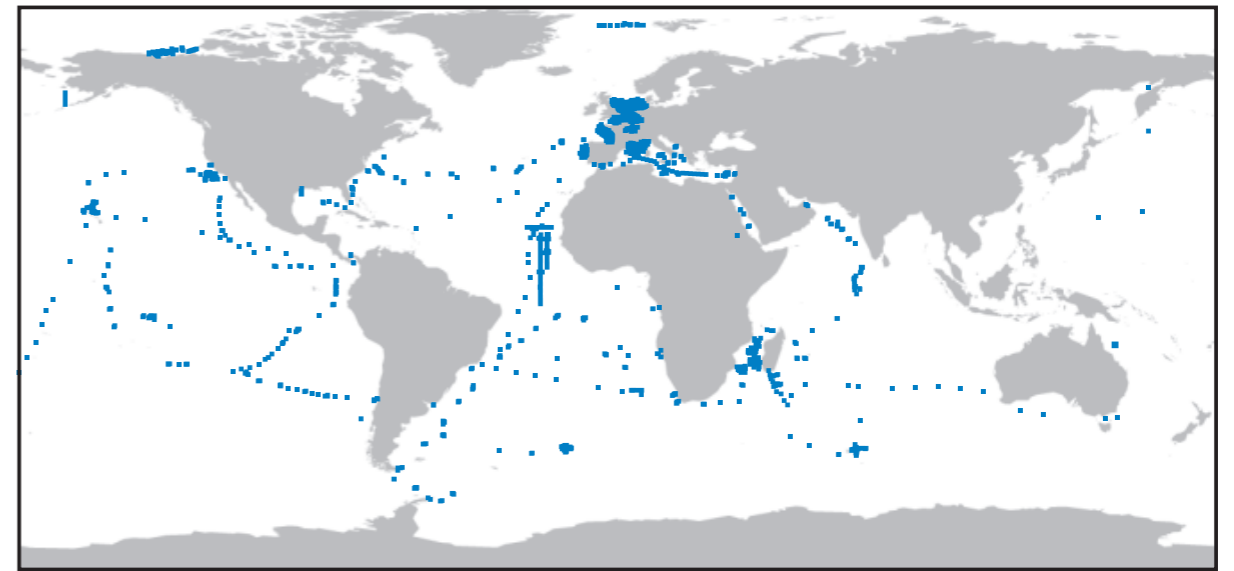


2020 (91,000 samples)

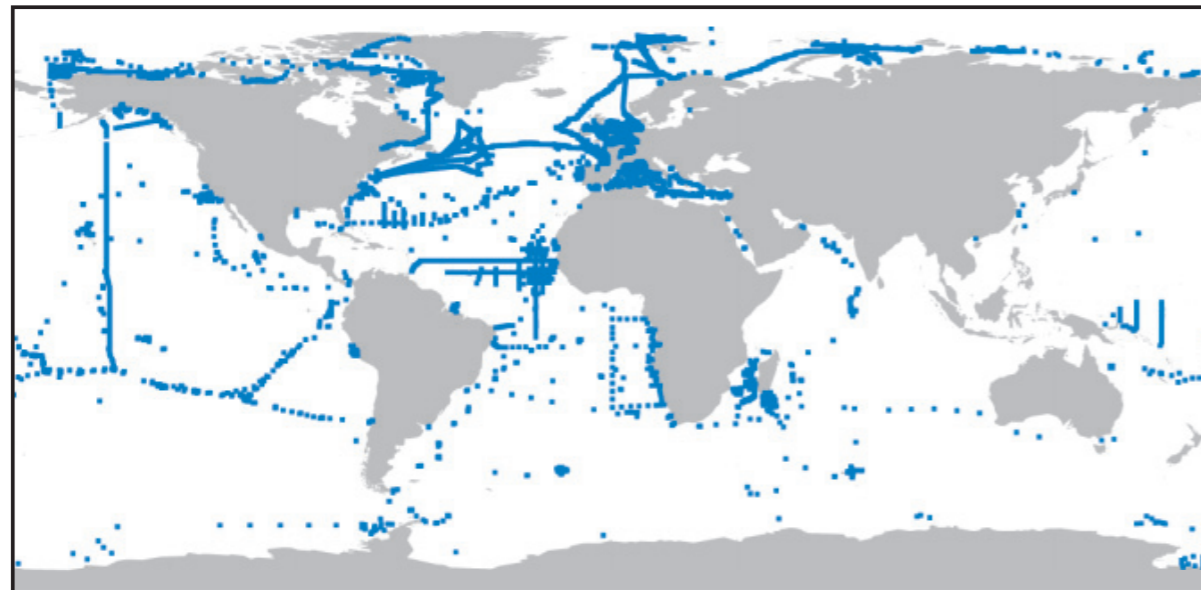
Steep growth in data acquisition



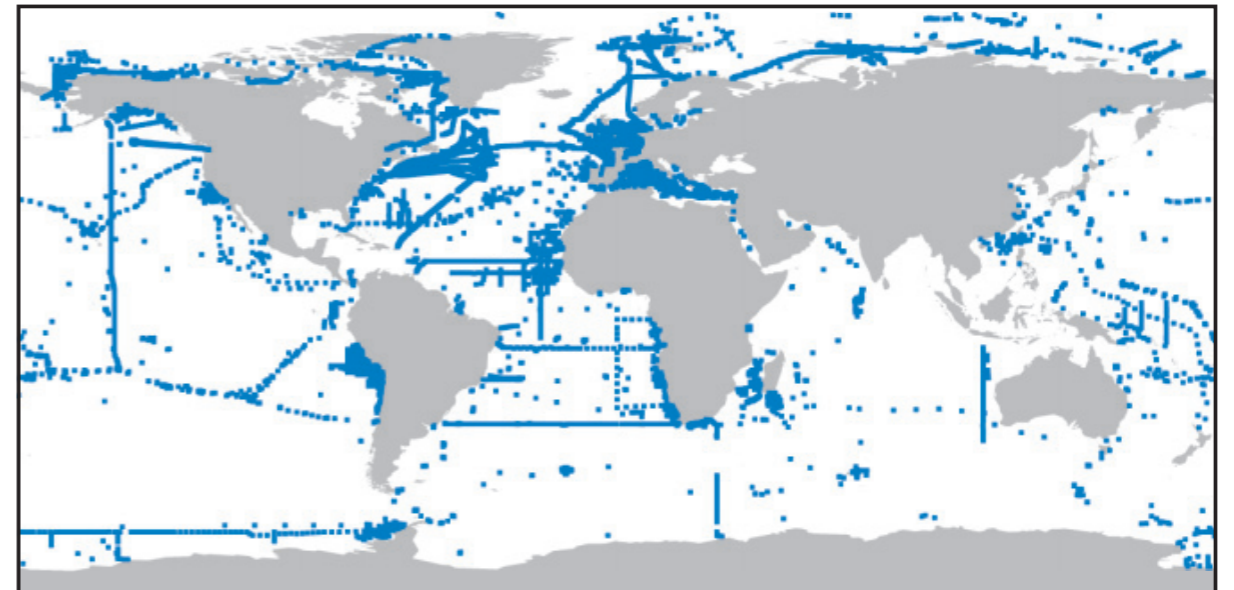
2008 (4,000 samples)



2012 (17,000 samples)



2016 (56,000 samples)



2020 (91,000 samples)

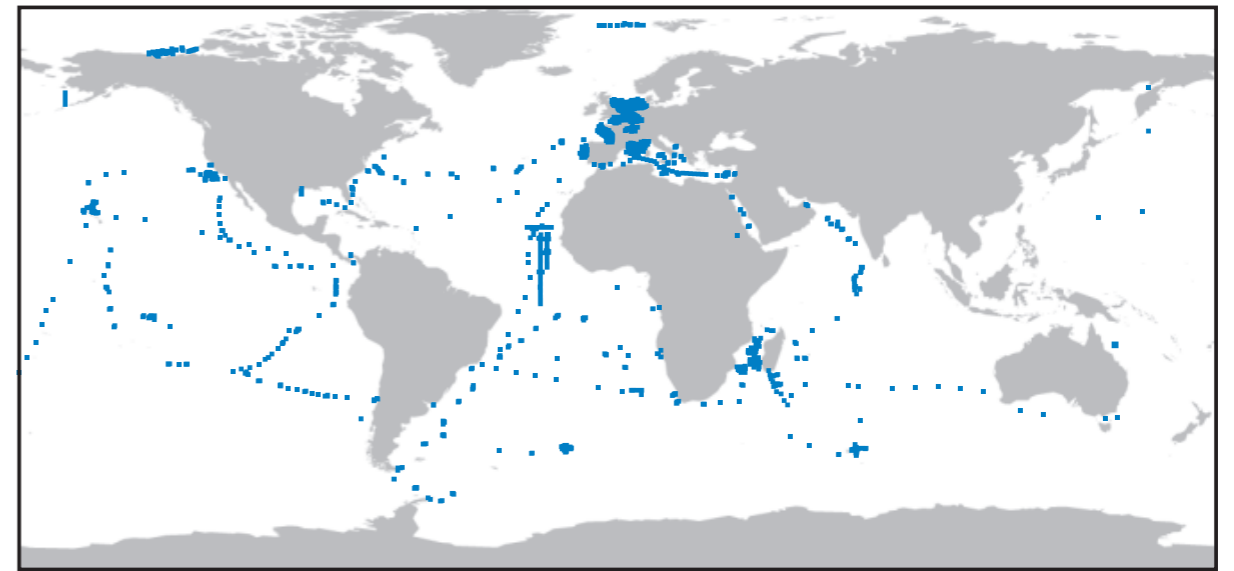


020

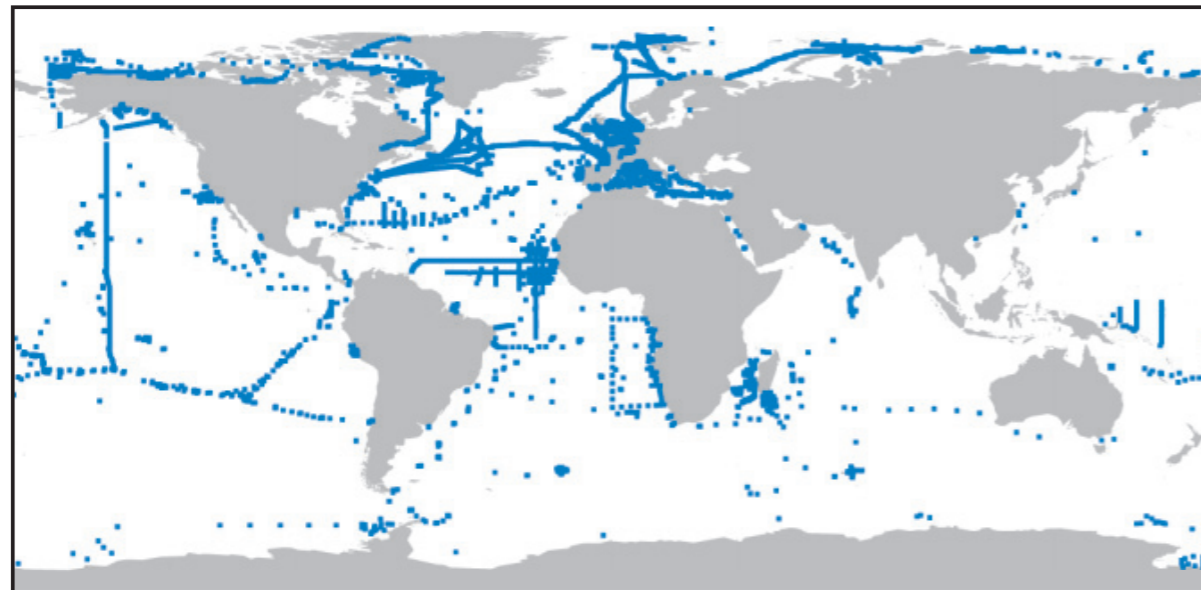
Steep growth in data acquisition



2008 (4,000 samples)



2012 (17,000 samples)



2016 (56,000 samples)



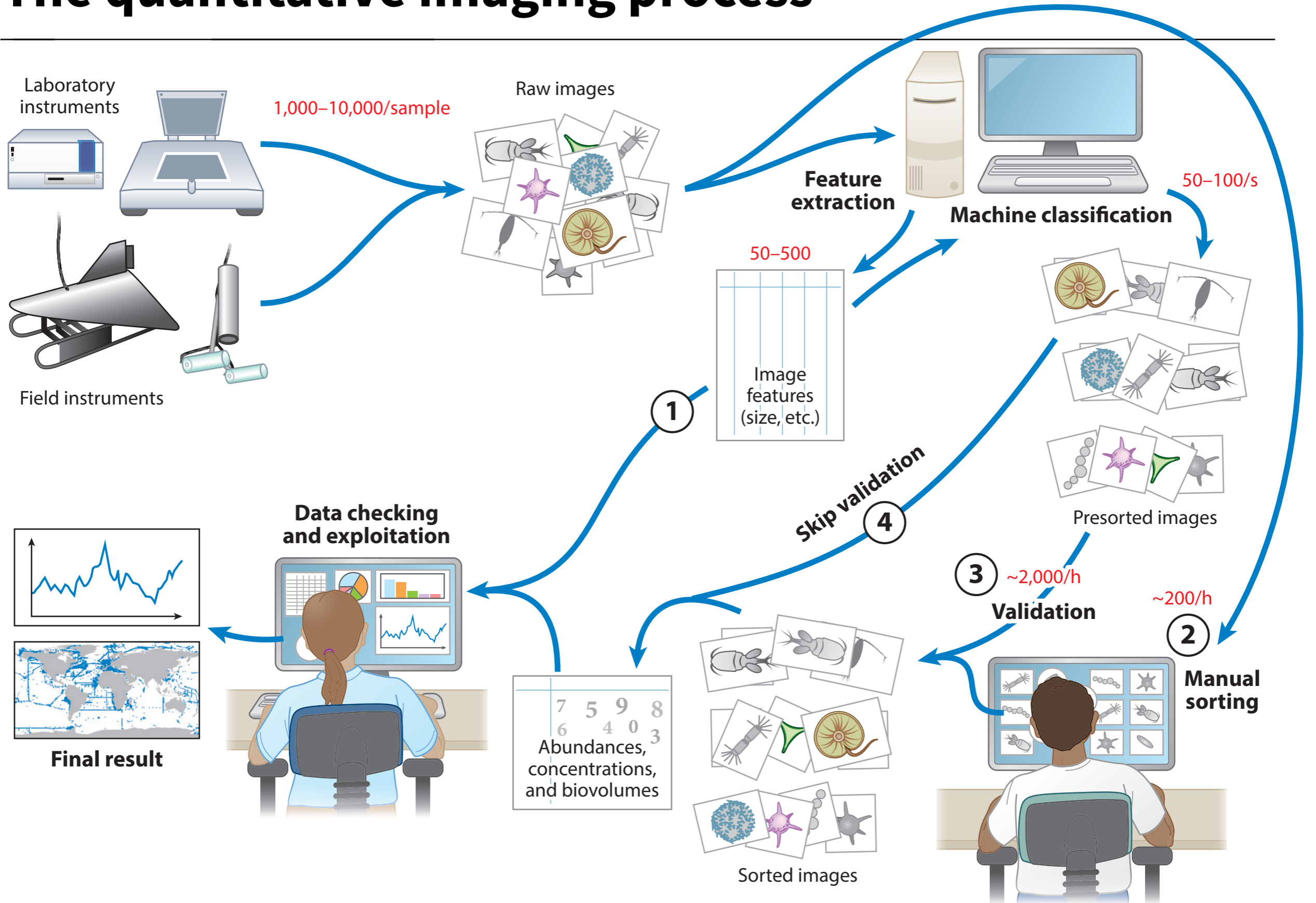
2020

How can we process this amount of data?



020

The quantitative imaging process



Web-based infrastructure for the classification of ecological images

ecotaxa.obs-vlfr.fr/prj/434

Zooscan AMP WP2 200 2013 2015 (1666, 1229, 0, 0 / 2895)

Jean-Olivier Irisson (log out)

Project Filtered Filter: Taxo= Oikopleuridae

Update view & apply filter Select all Score Display Status All 50 30

Taxonomy filter Other filters

Appen

Appendicularia

Appendiceirospora

Appendiceirospora graminicola

Appendicospora

Appendicospora hongkongensis

Appendicospora sp.

Appendicularia X

Appendicularia X sp.

Oikopleuridae 1666 1229

tail < Appendicularia 1503 944

Bryozoa 0

cyphonaute 525 810

Chaetognatha 4413 162

head < Chaetognatha 2

tail < Chaetognatha 849 148

Cnidaria < Metazoa 0

Hydrozoa 641 454

Aglaura 28

Leptothecata 14

Aequorea 3

Obelia 139 190

Sertulariidae 23

Siphonophorae 0

Abylidae 1

Oikopleuridae Score : 0.92

Oikopleuridae Score : 0.91

Oikopleuridae Score : 0.91

Oikopleuridae Score : 0.91

Oikopleuridae Score : 0.90

Oikopleuridae Score : -

Oikopleuridae Score : -

Oikopleuridae Score : 0.89

Oikopleuridae Score : 0.88

Oikopleuridae Score : 0.88

Oikopleuridae Score : 0.88

Oikopleuridae Score : 0.87

Oikopleuridae Score : 0.86

Oikopleuridae Score : 0.86

Oikopleuridae Score : 0.86

Oikopleuridae Score : -

Oikopleuridae Score : 0.85

Oikopleuridae Score : -

Oikopleuridae Score : 0.85

Oikopleuridae Score : 0.85

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.84

Oikopleuridae Score : 0.83

Oikopleuridae Score : 0.83

Oikopleuridae Score : -

Oikopleuridae Score : 0.83

Oikopleuridae Score : 0.83

Oikopleuridae Score : 0.82

Oikopleuridae Score : 0.82

Oikopleuridae Score : 0.82

Oikopleuridae Score : 0.82

Oikopleuridae Score : 0.82

Oikopleuridae Score : -

Oikopleuridae Score : -

Oikopleuridae Score : 0.80

Oikopleuridae Score : 0.80

Oikopleuridae Score : 0.81

Oikopleuridae Score : 0.81

Oikopleuridae Score : 0.81

Oikopleuridae Score : -

Oikopleuridae Score : 0.80

Oikopleuridae Score : 0.80

Oikopleuridae Score : -

Oikopleuridae Score : -

Oikopleuridae Score : 0.80

Oikopleuridae Score : 0.80

Oikopleuridae Score : -

Oikopleuridae Score : -

Oikopleuridae Score : -

Oikopleuridae Score : 0.79

Save pending changes [CTRL+S] Validate all and move to next page Validate Selection [CTRL+L] Set Selection Dubious Undo

Page 28 / 58

« 1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 »

Web-based infrastructure for the classification of ecological images

ecotaxa.obs-vlfr.fr/prj/859

EcoTaxa Demo Zooscan JO low taxo res

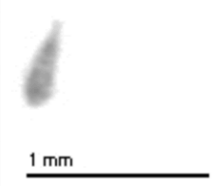
Demo Zooscan JO low taxo res (21, 1376, 0, 0 / 1397) Jean-Olivier Irisson (log out)

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

Project: [Demo Zooscan JO low taxo res](#) (managed by : [Jean-Olivier Irisson](#))
To report a mistake, contact [Jean-Olivier Irisson](#)

Classification :
Copepoda
Copepoda < Maxillopoda < Crustacea < Arthropoda < Metazoa < Holozoa < Opisthokonta < Eukaryota < living (id=25828)

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](#)

Sample details :

| | | | | | | | |
|------------------|----------------|--------------------|---------|--------------|---------|-------------|------|
| Original ID | wp220180205 | longitude | 7.32 | latitude | 43.69 | | |
| scan_operator | corinne_desnos | ship | sagitta | program | zooscan | stationid | ptb |
| bottomdepth | 85 | ctdrosettefilename | | other_ref | | tow_nb | 2 |
| tow_type | 3 | net_type | wp2 | net_mesh | 200 | net_surf | 0.25 |
| zmax | 75 | zmin | 0 | tot_vol | 37.5 | comment | no |
| tot_vol_qc | | depth_qc | | sample_qc | | barcode | |
| duration | | ship_speed | | cable_length | | cable_angle | |
| cable_speed | | nb_jar | | open | | | |
| Dataportal Desc. | | | | | | | |

Annotator / Free filters Advanced Clear

Annotator



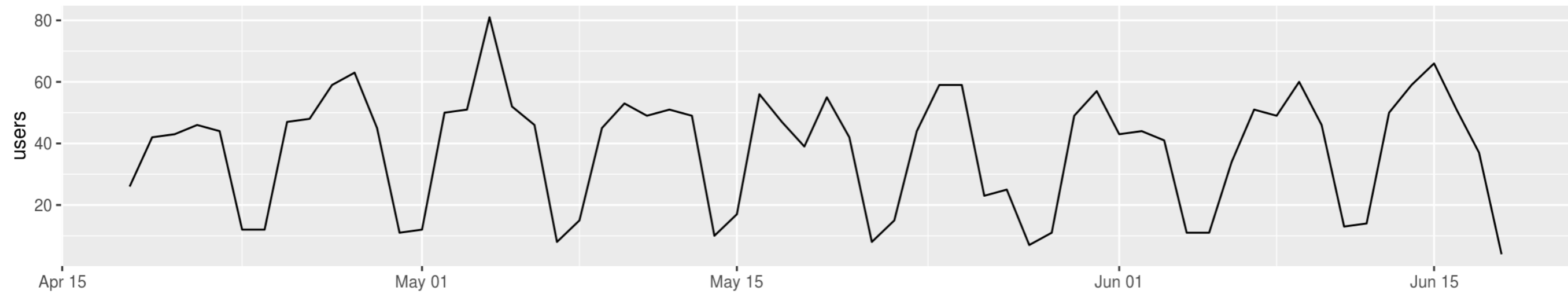
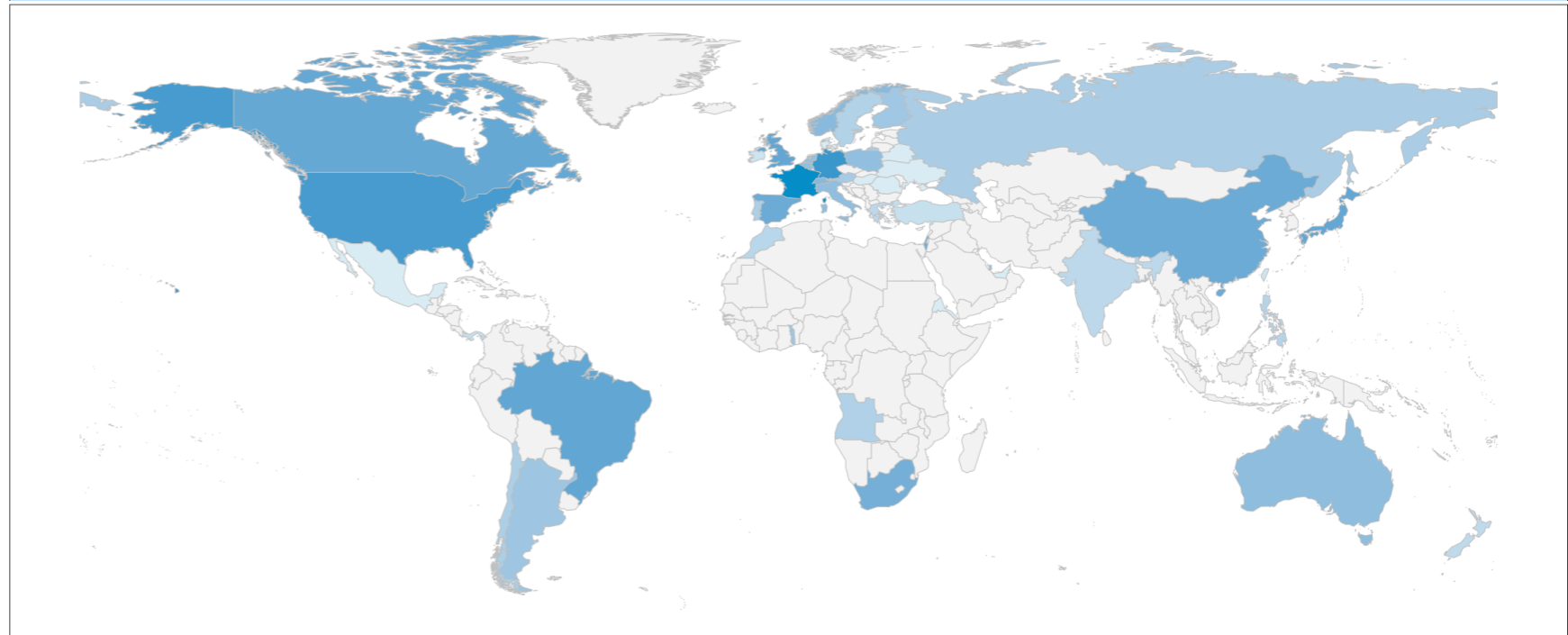
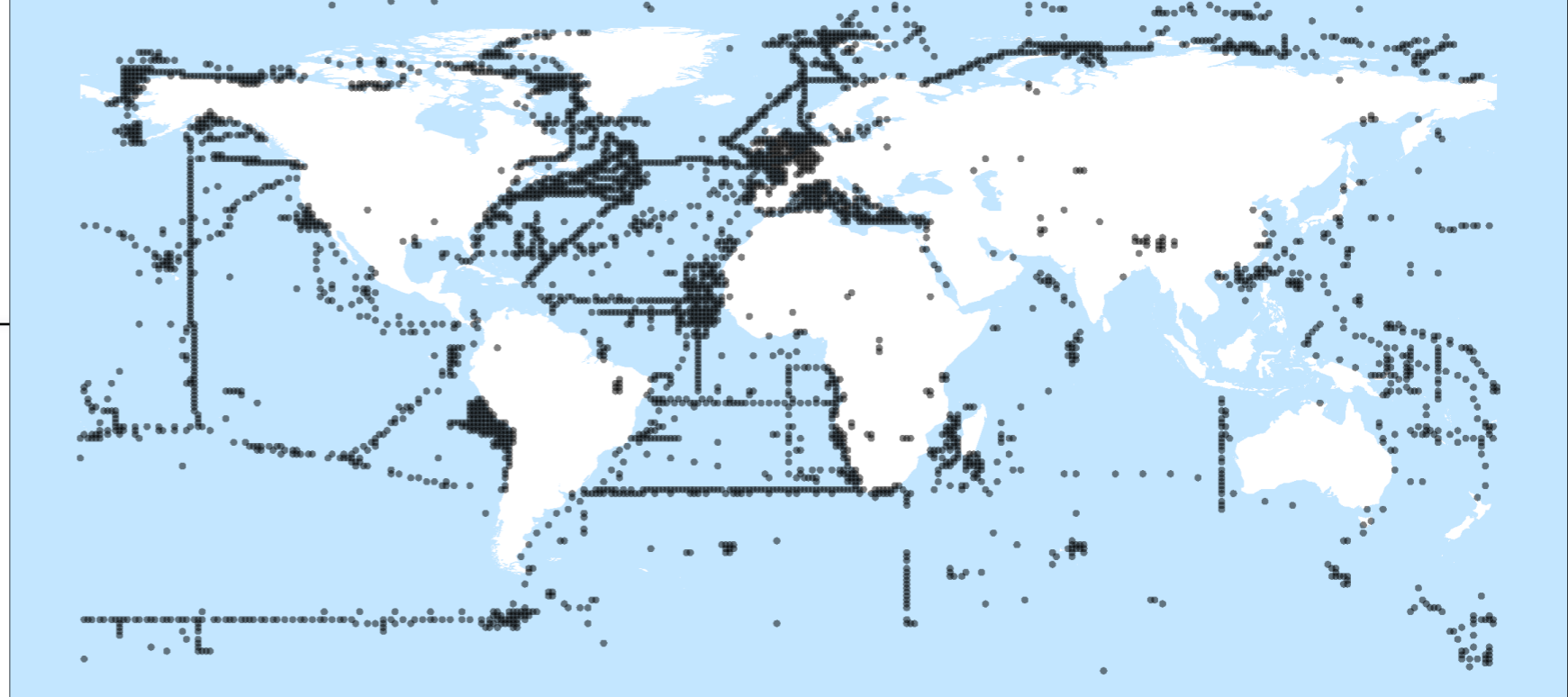
EcoTaxa today

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

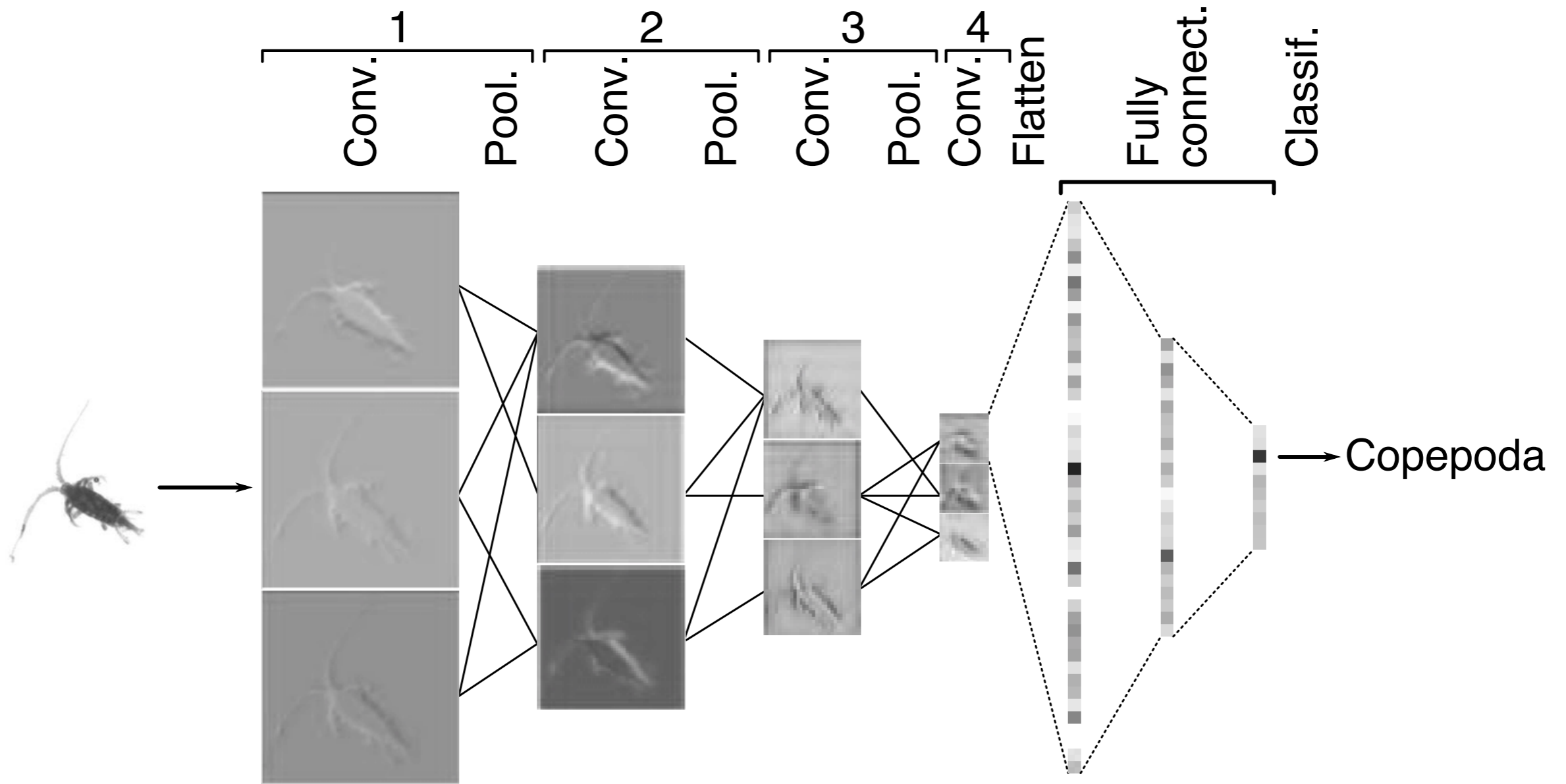
>100k locations in the world's oceans

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

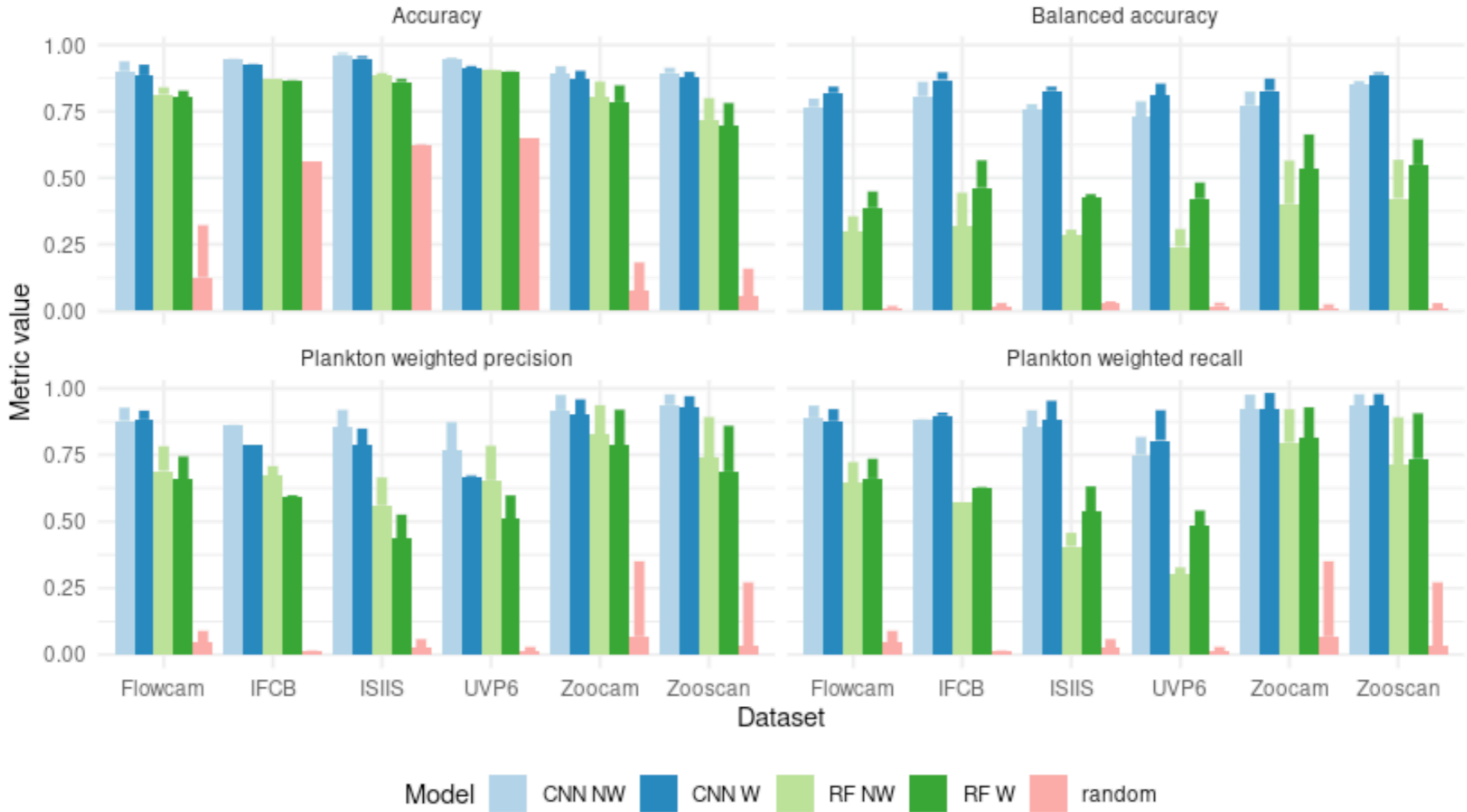
50 concurrent users at all times



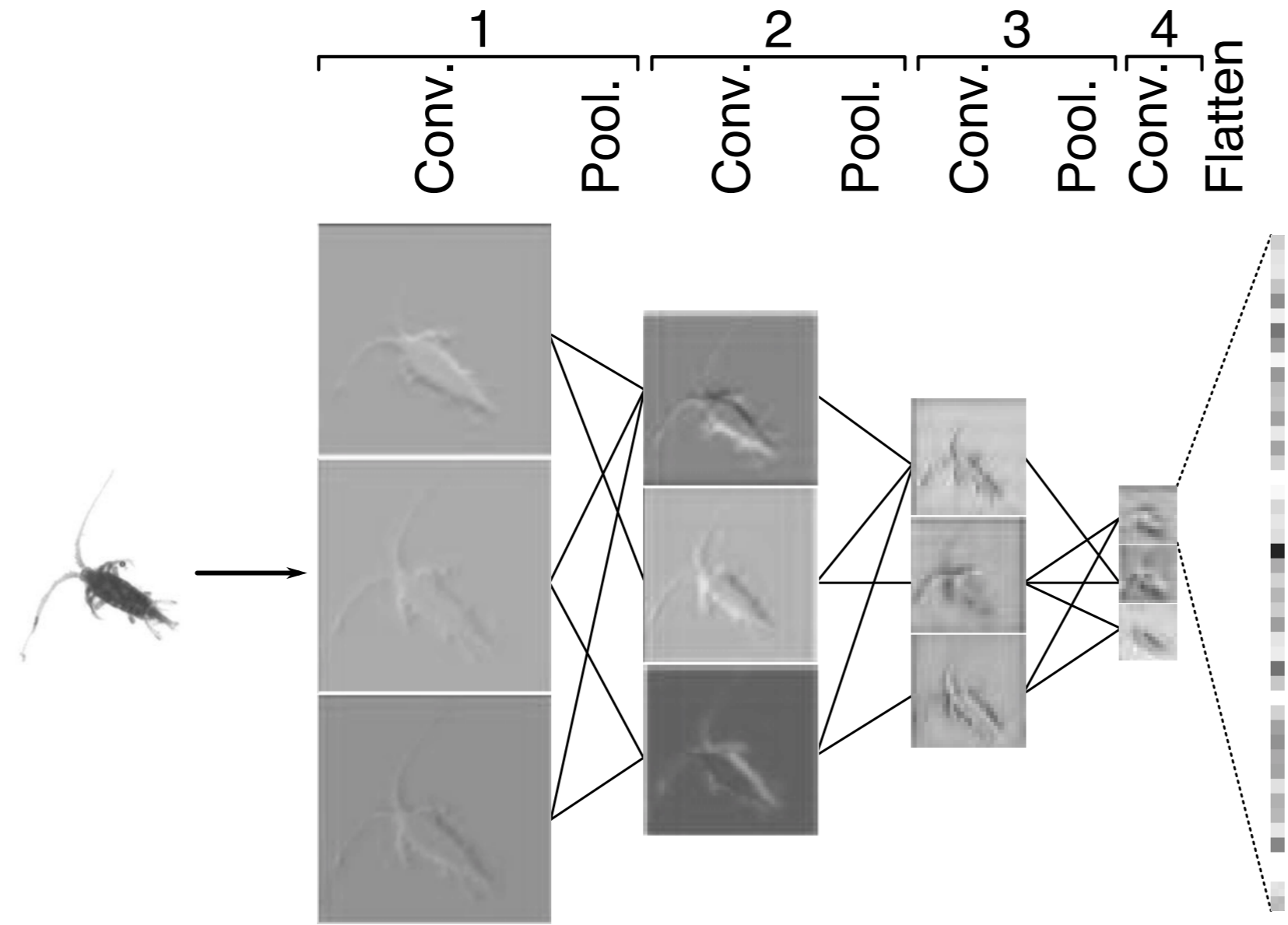
Improved Machine Learning backbone



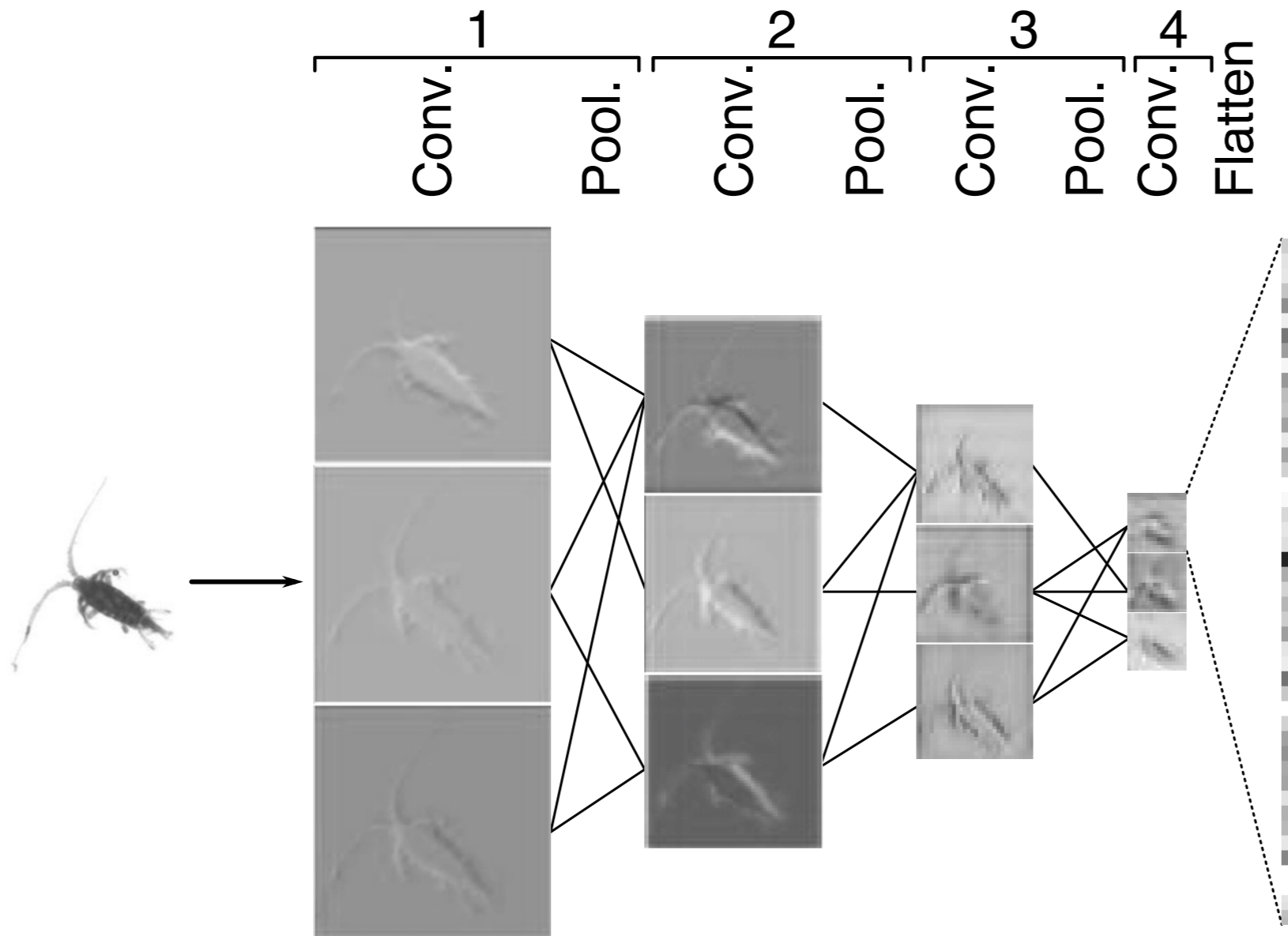
Improved Machine Learning backbone



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

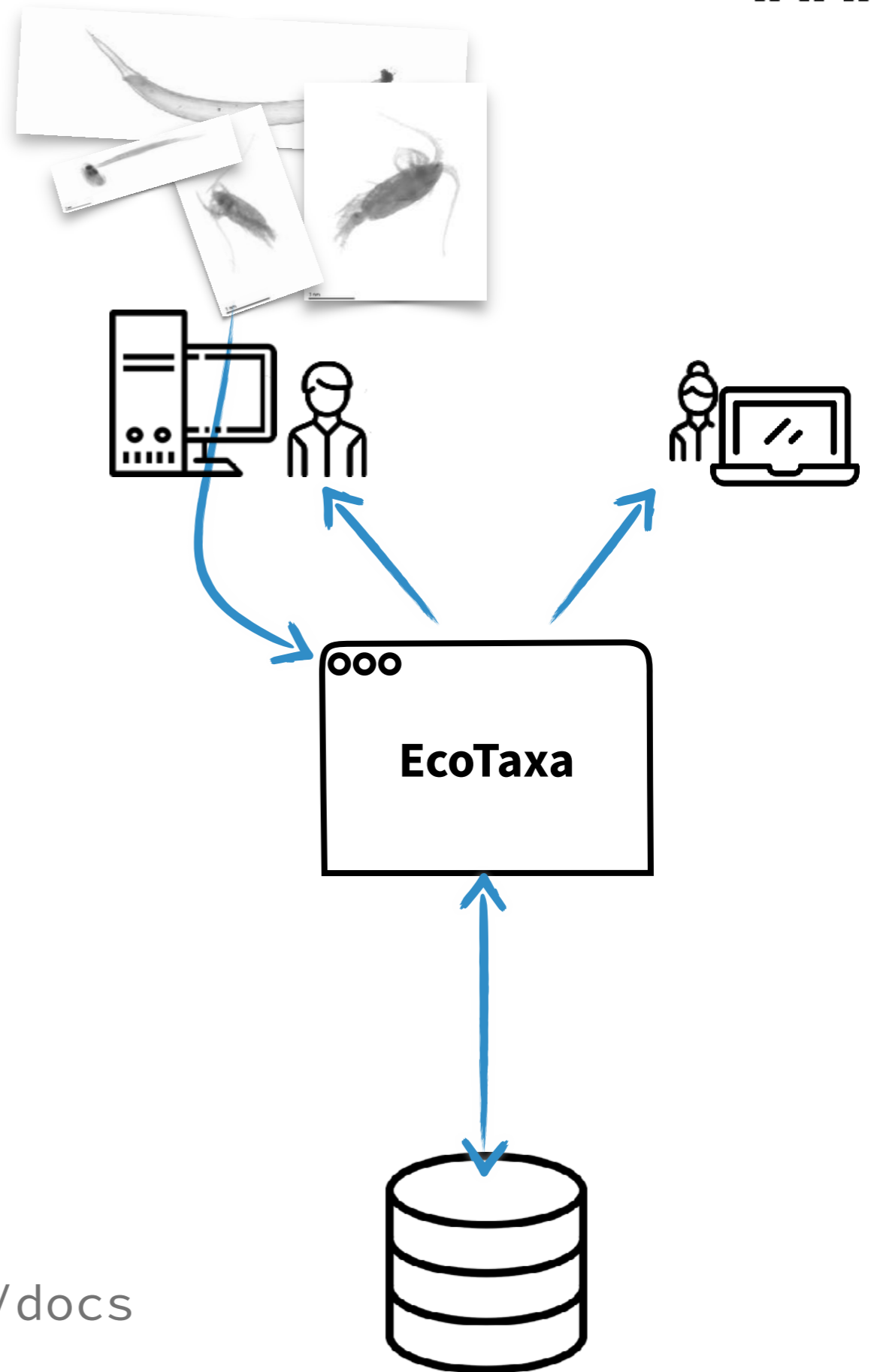
import data

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

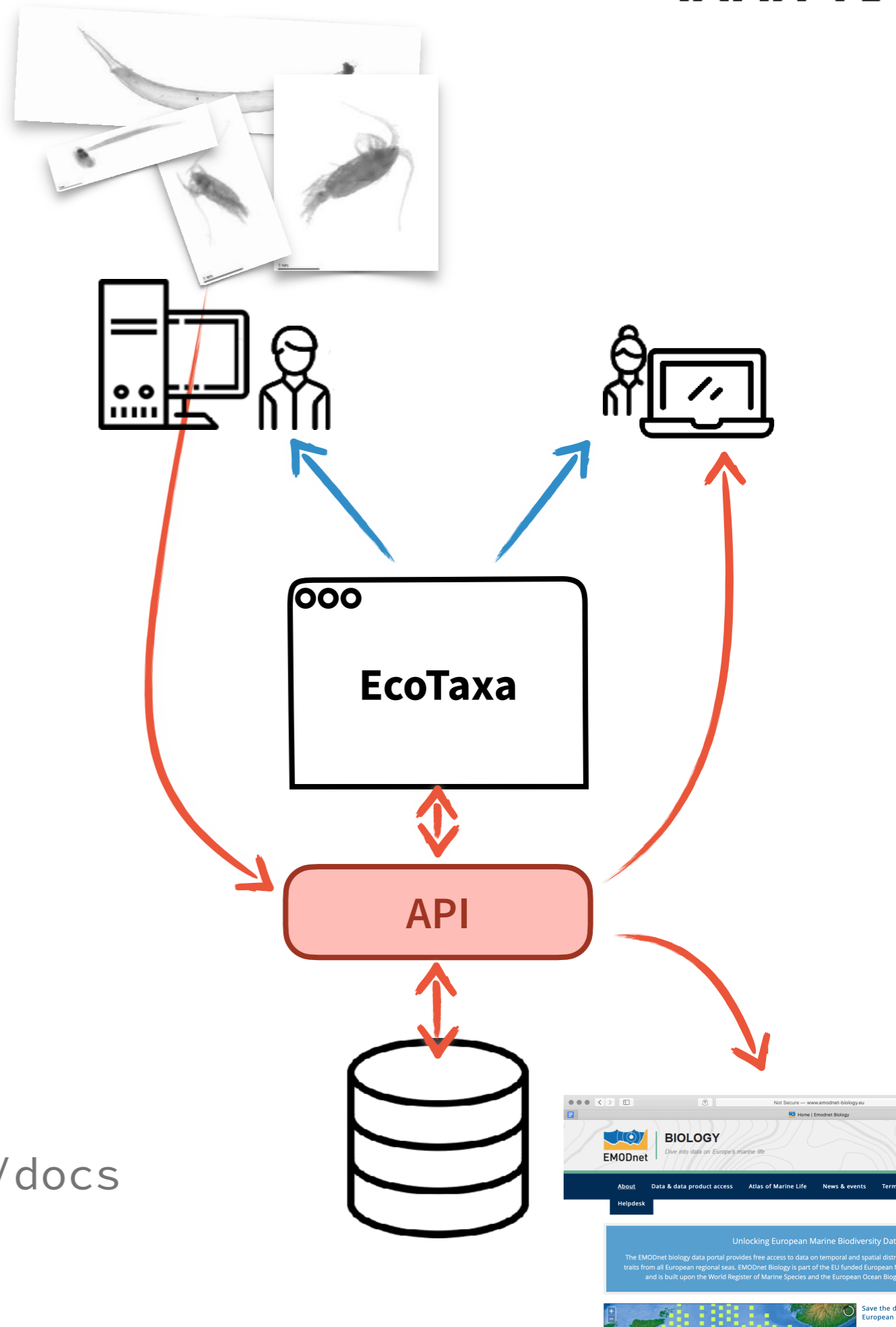
import data

query subsets of data

classify images automatically

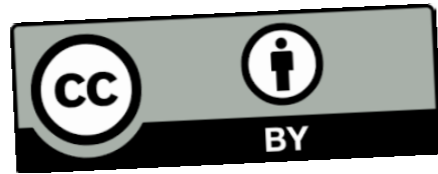
export datasets, in particular to external databases

<https://ecotaxa.obs-vlfr.fr/api/docs>



Export open data as DarwinCore Archive

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



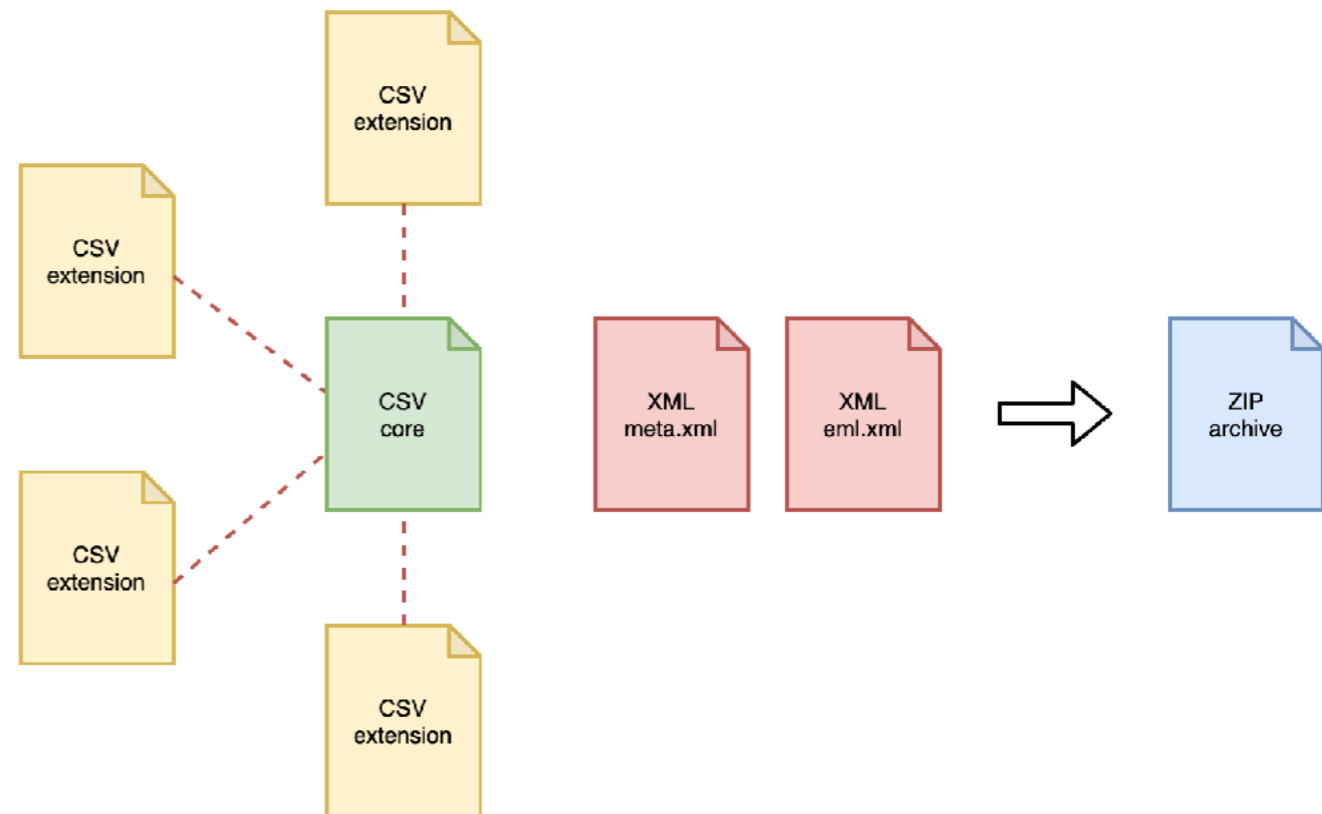
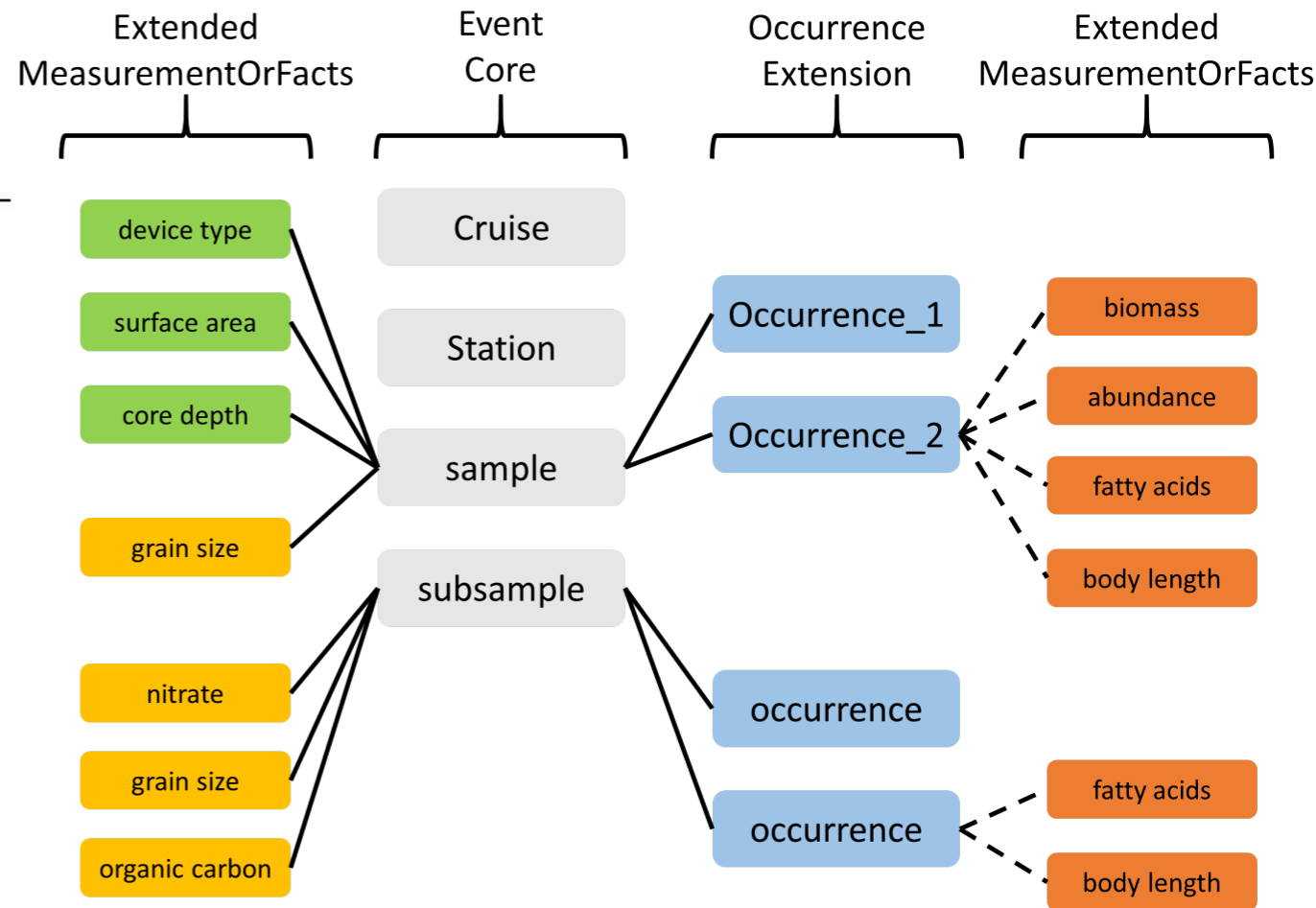
DarwinCore Archive is

The **standard** format for biodiversity information

Organised **hierarchically**, like EcoTaxa

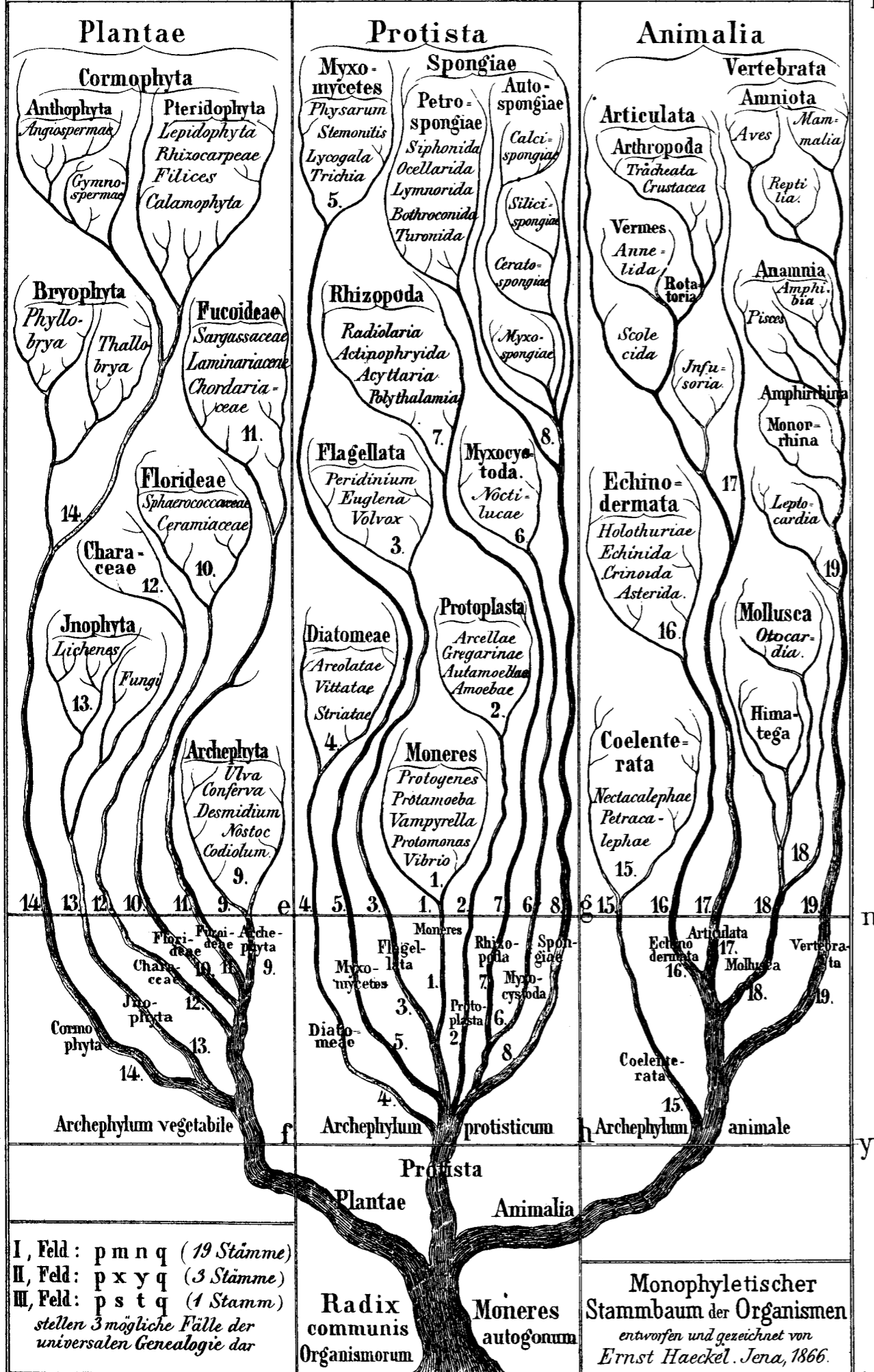
Very **descriptive** but **complex** to produce

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



Difficulties

Choosing a universal taxonomy

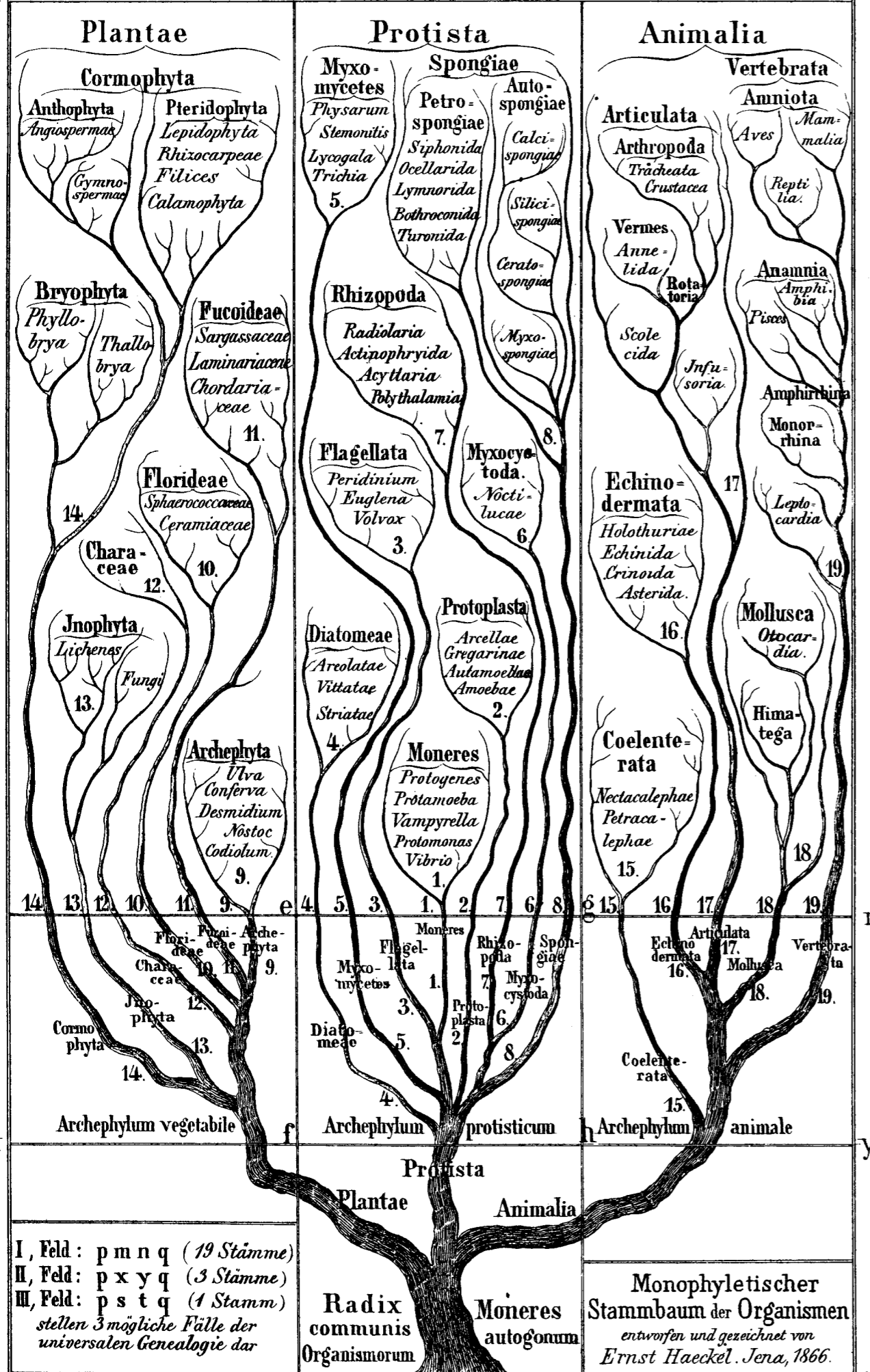


I, Feld: p m n q (19 Stämme)
 II, Feld: p x y q (3 Stämme)
 III, Feld: p s t q (1 Stamm)
 stellen 3 mögliche Fälle der
 universalen Genealogie dar

Difficulties

Choosing a **universal taxonomy**

Training the community to use the machine learning suggestions wisely



I, Feld: p m n q (19 Stämme)
 II, Feld: p x y q (3 Stämme)
 III, Feld: p s t q (1 Stamm)
 stellen 3 mögliche Fälle der
 universalen Genealogie dar



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**

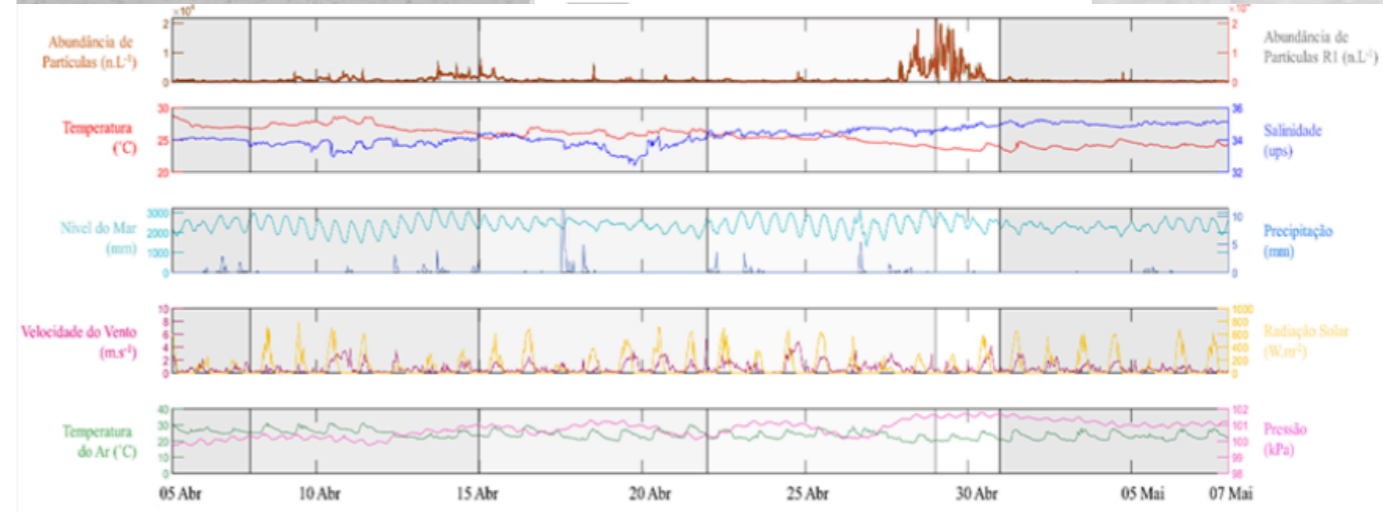
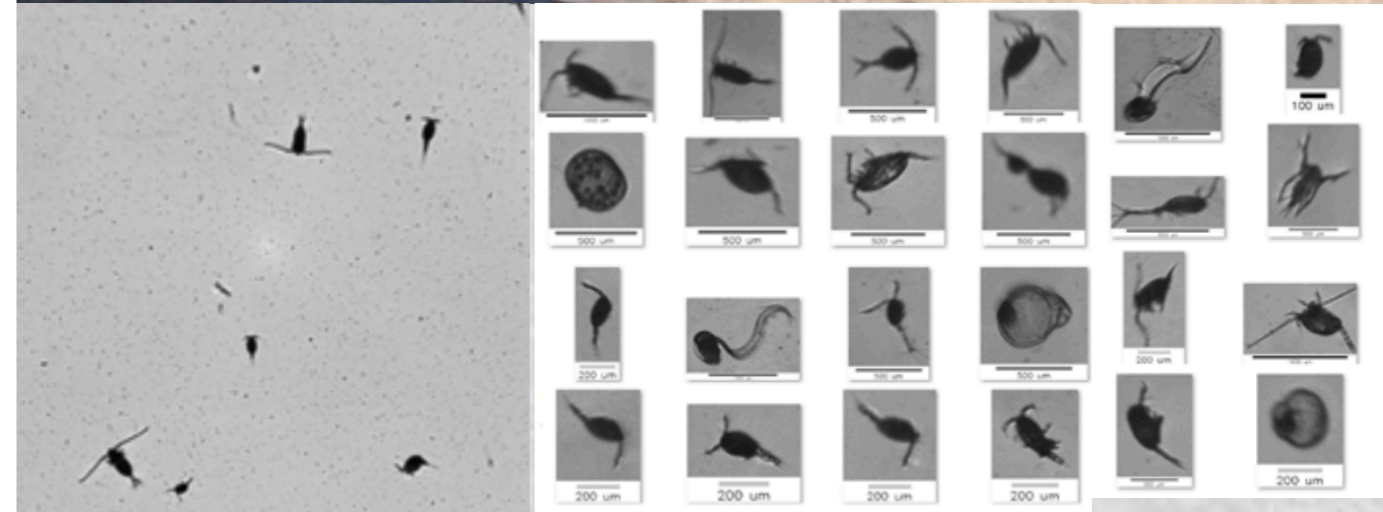
Finding (good) **programmers** !

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



Use 1: Environmental monitoring

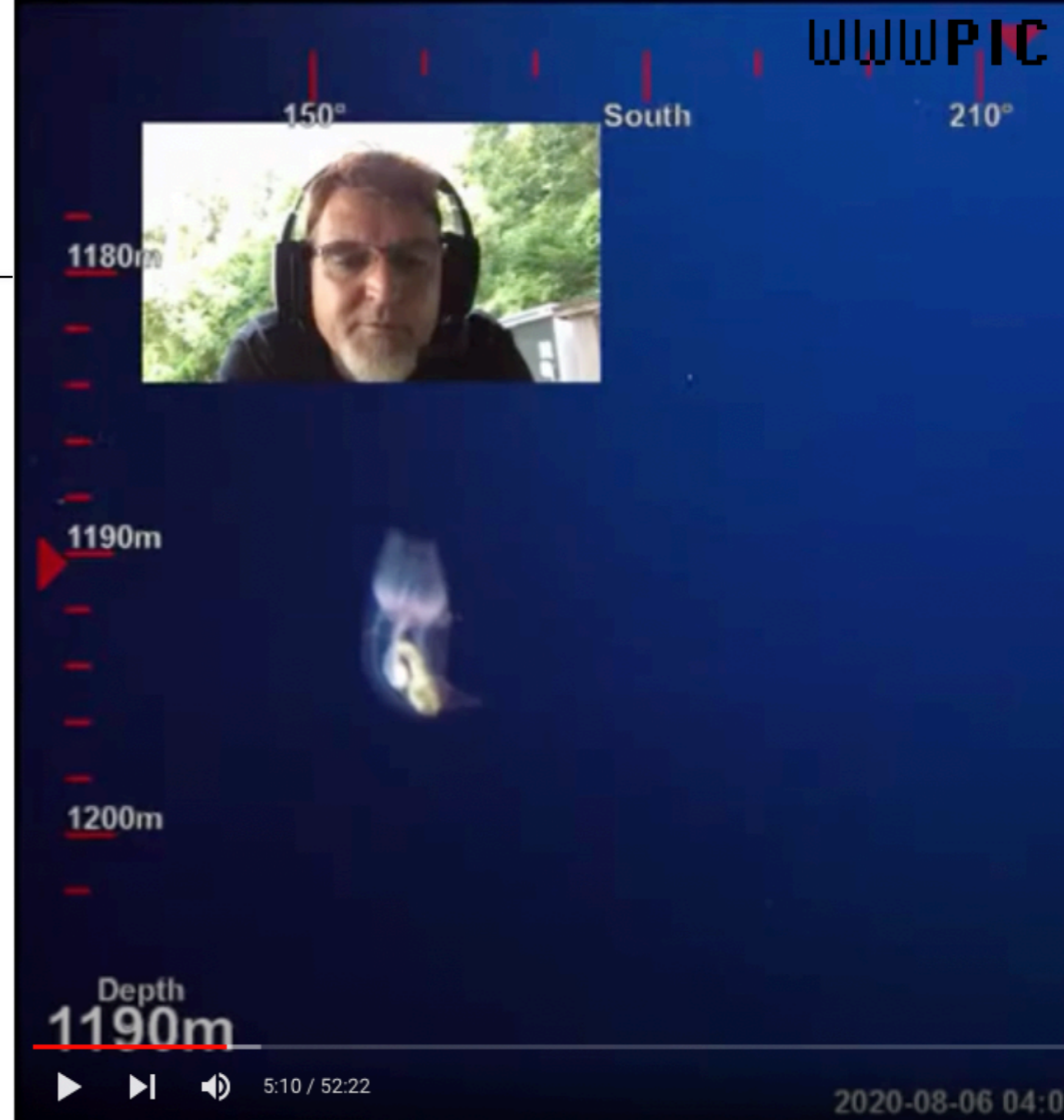
High resolution **time series** at coastal sites, to detect **invasive** or **harmful** species



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



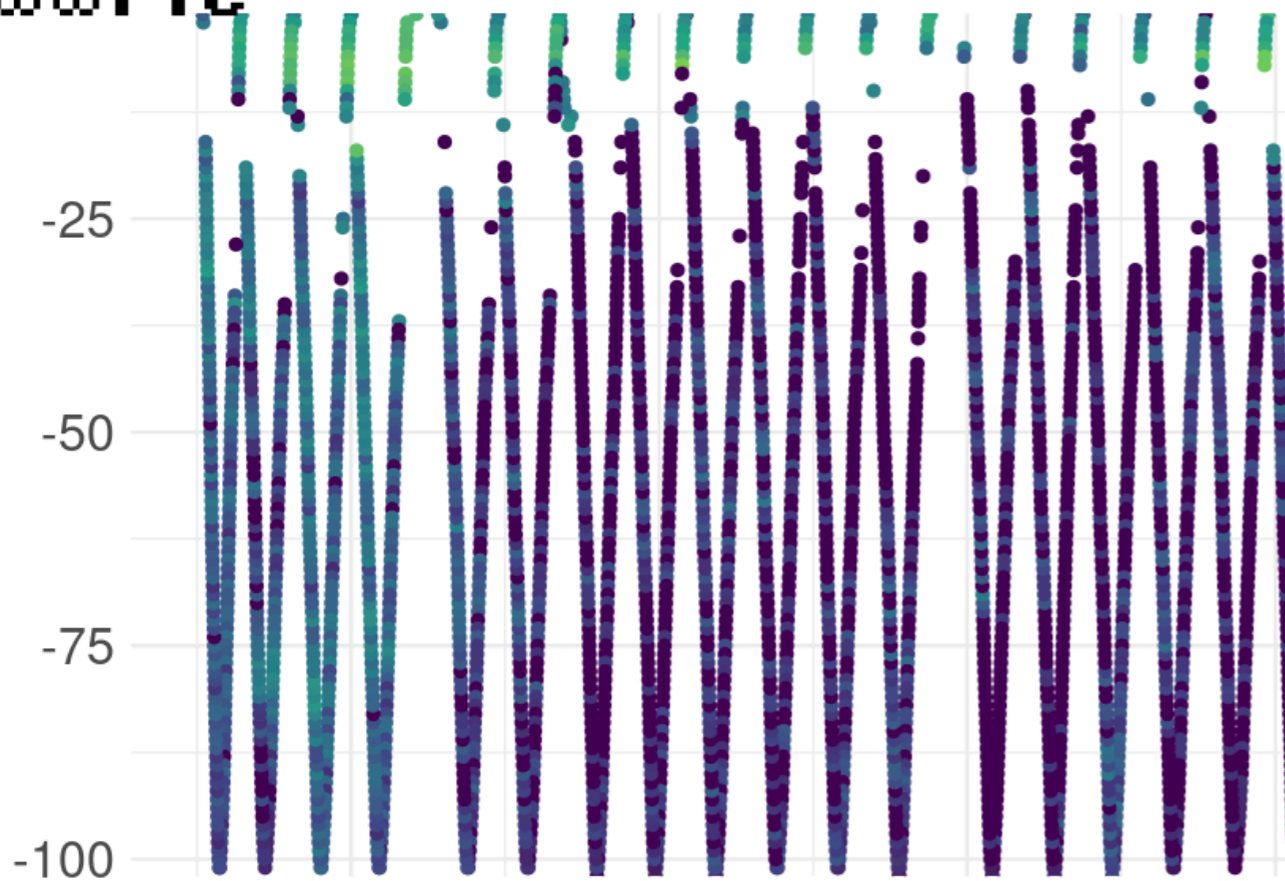
Use 2: Biology studies at *submesocale*

Imager deployed on automated
underwater **glider**

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

WWPIC

Acantharea

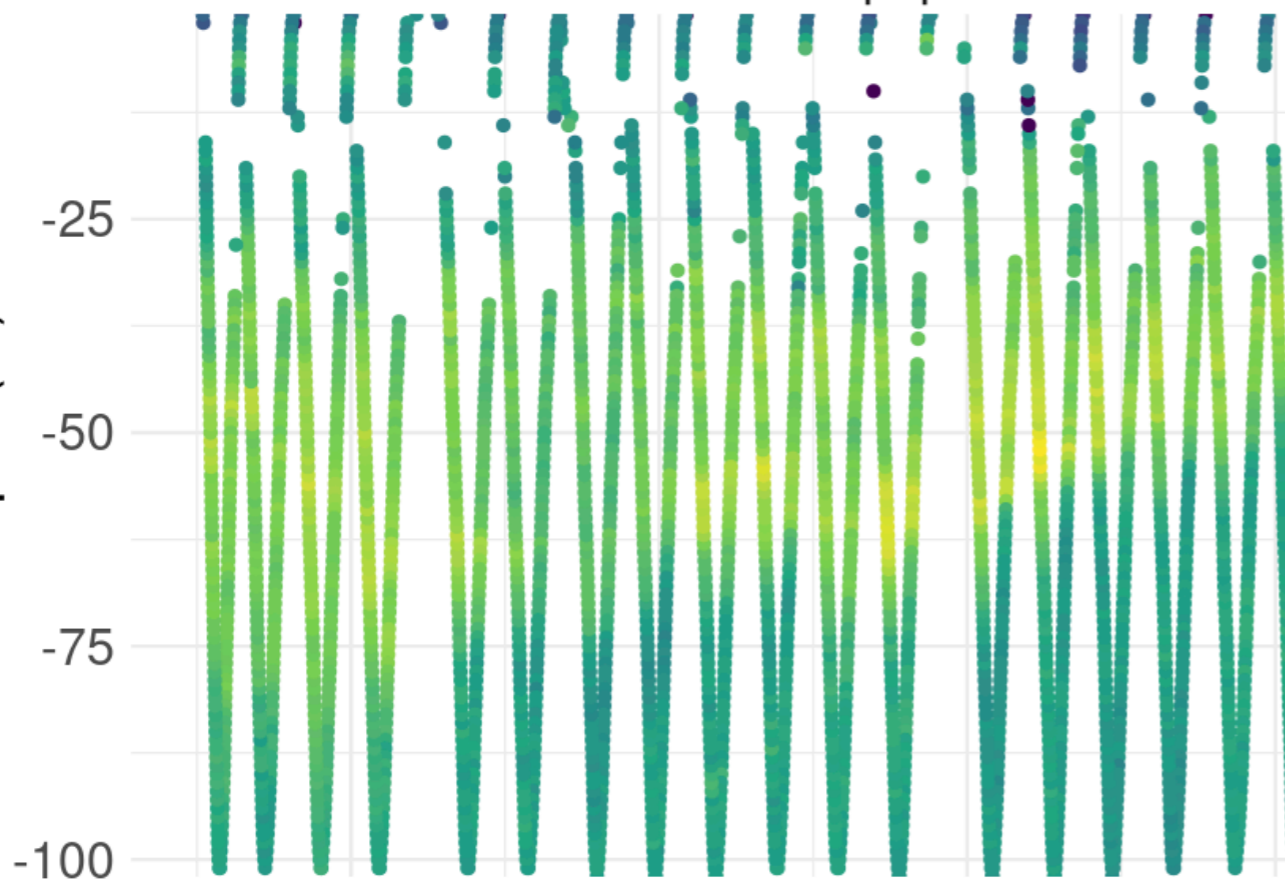


Use 2: Biology studies at *submesocale*

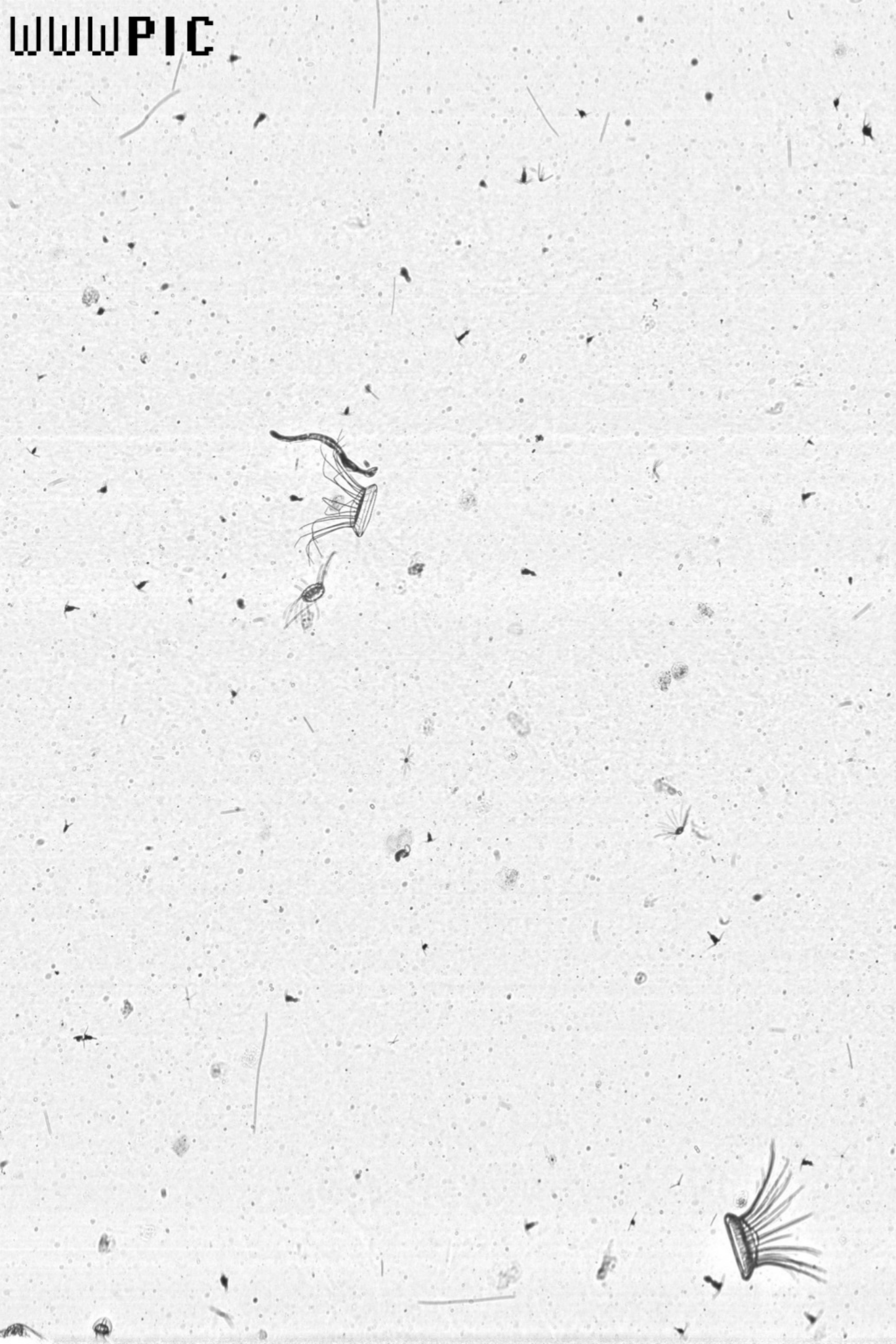
Imager deployed on automated
underwater **glider**

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

Copepoda



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



Use 2: Biology studies at *submesocale*

Imager deployed on automated
underwater **glider**

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

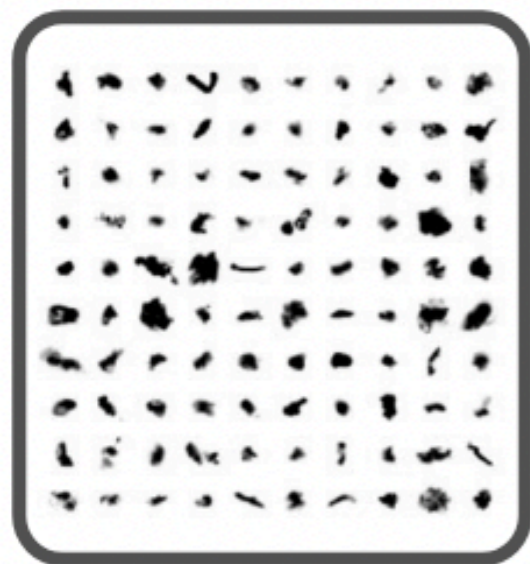
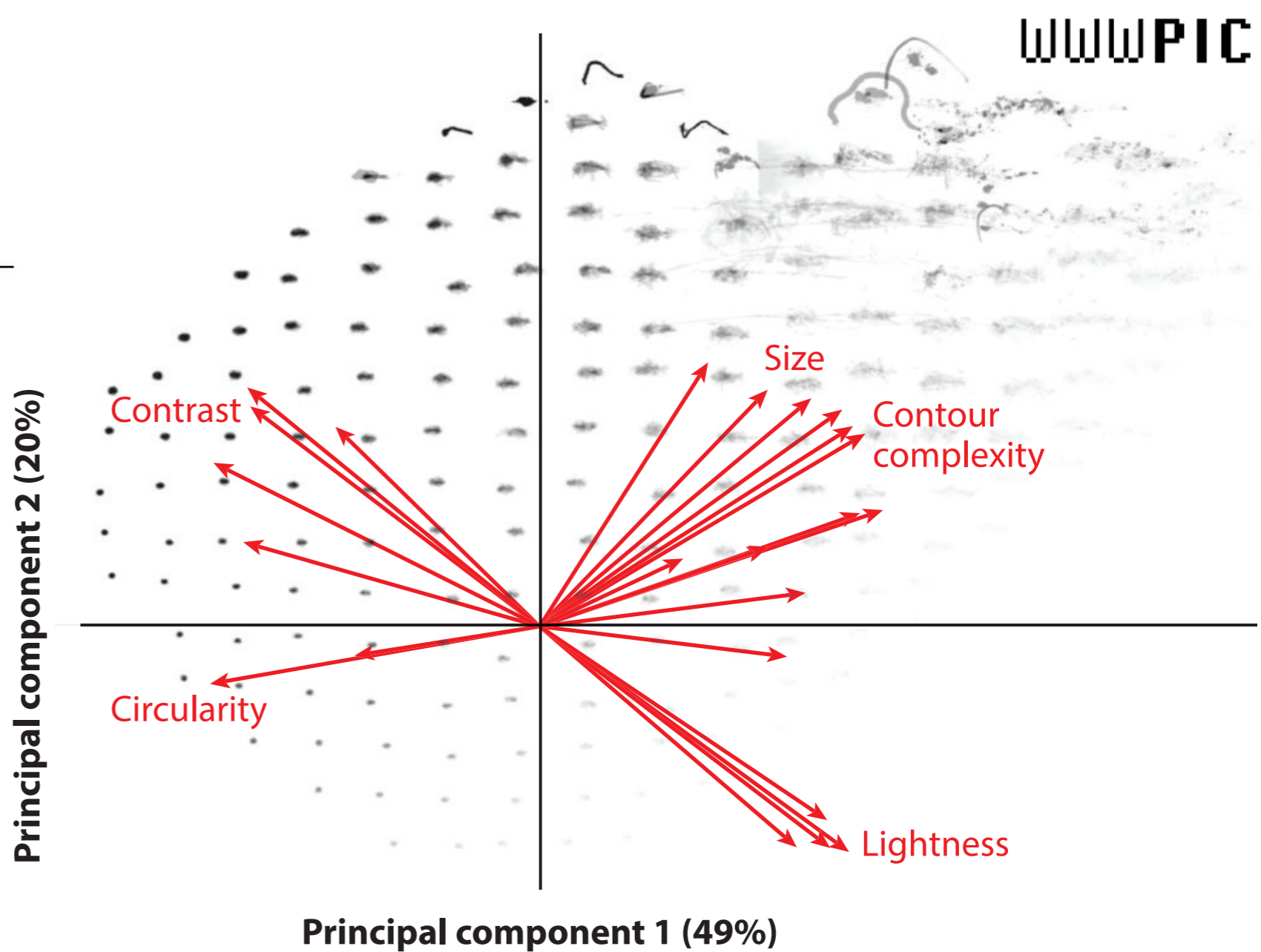
Describe the **biology** on a front at the
same resolution as the physics and
chemistry

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**



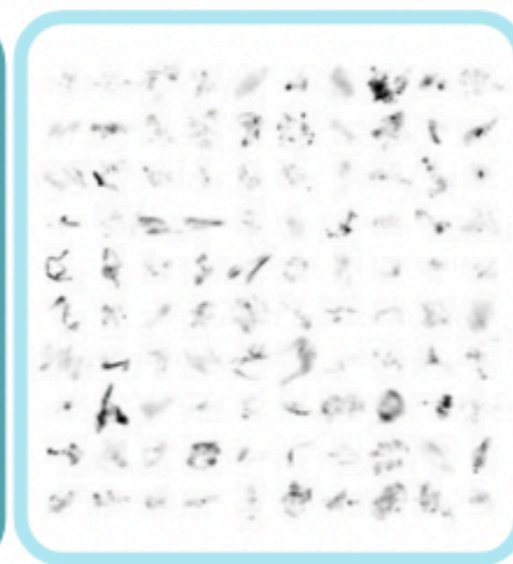
dark



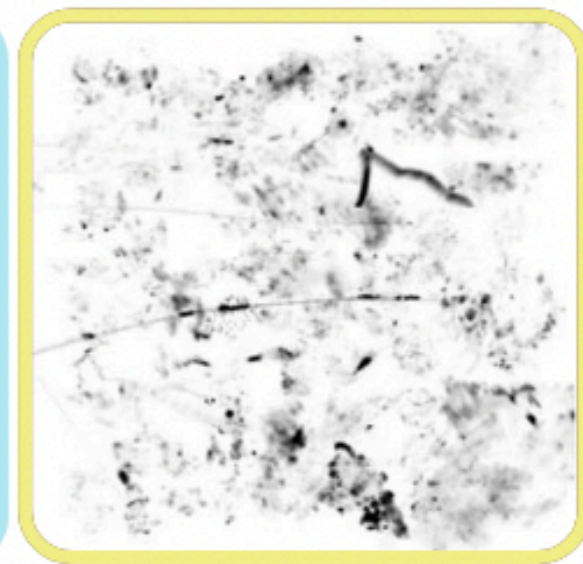
elongated



flakes



fluffy



agglomerated

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/>



Merci