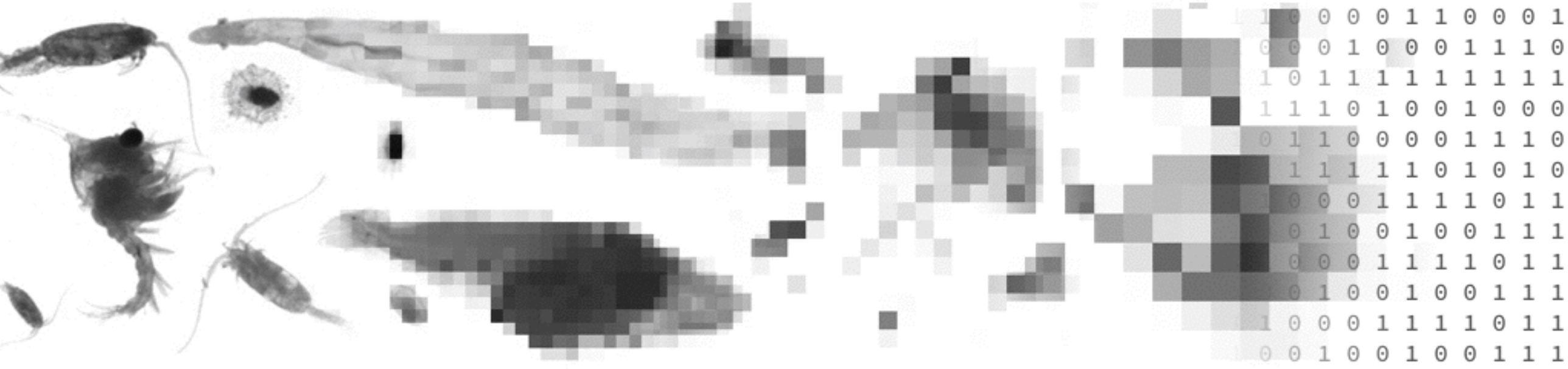
*Atelier Zooplancton RESOMAR-MOOSE, 2016-12-01*
*Jean-Olivier Irisson*

# Traitement, sauvegarde et diffusion des données

## Banques de données et démarche qualité



# Traitement des données

**taxa 1 + env**

modèle de niche (NPPEN, MaxEnt, etc.), diversité taxinomique, phylogénétique

**taxa 1/0 + env**

relations probabilité de présence - environnement (régression logistique, BRT, etc.)

**taxa n (± temps)**

composition des communautés, indices de diversité

**taxa n + env (± temps)**

relations écologiques abondance - environnement (PCA, CA, GLM, GAM, BRT, etc.)

**taille, volume + env (± taxa)**

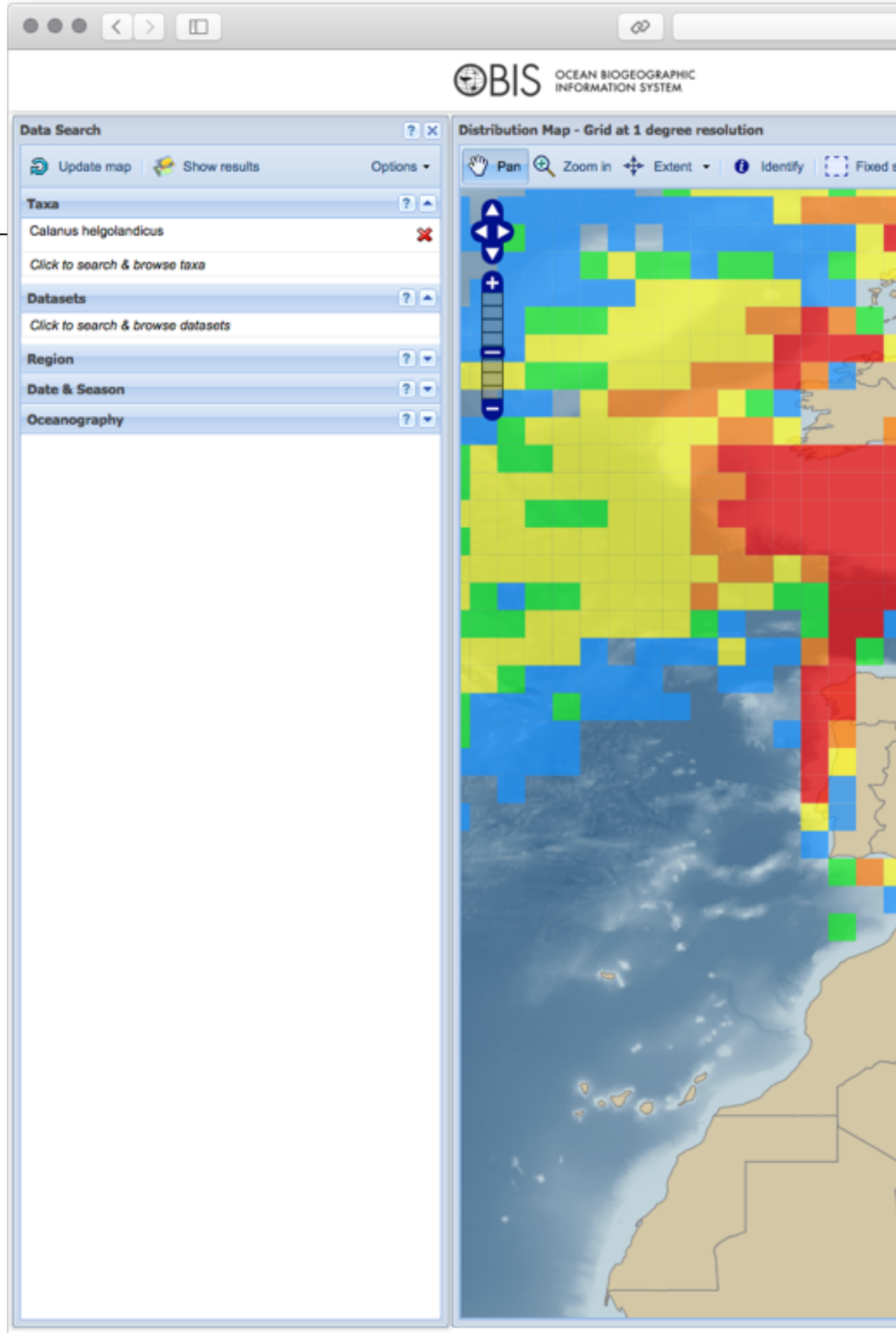
approche par traits fonctionnels, diversité fonctionnelle

# Bases de données

**Générales, internationales**
http://www.iobis.org
http://www.iobis.org/mapper/
1 (n, env)

**Zooplancton, internationales**
http://www.st.nmfs.noaa.gov/copepod/
http://www.st.nmfs.noaa.gov/copepod/time-series/
n, volume, env mais ancien

**Zooplancton, spécifiques**
CPR https://www.sahfos.ac.uk
Villefranche http://ecotaxa.obs-vlfr.fr
n (forme)



# Bases de données

**Générales, internationales**

http://www.iobis.org

http://www.iobis.org/mapper/

1 (n, env)

**Zooplancton, internationales**

http://www.st.nmfs.noaa.gov/copepod/

http://www.st.nmfs.noaa.gov/copepod/time-series/

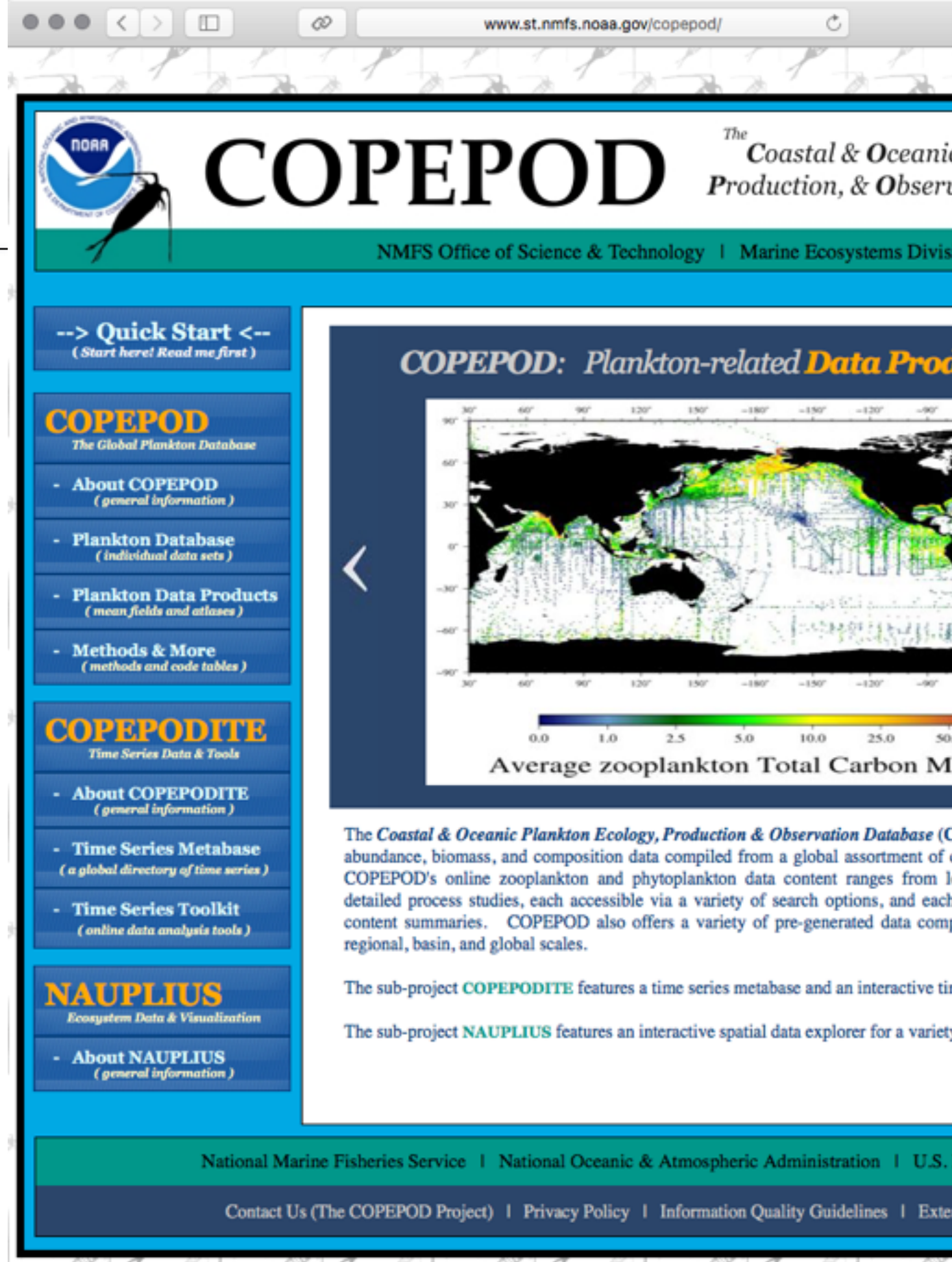
n, volume, env mais ancien

**Zooplancton, spécifiques**

CPR https://www.sahfos.ac.uk

Villefranche http://ecotaxa.obs-vlfr.fr

n (forme)

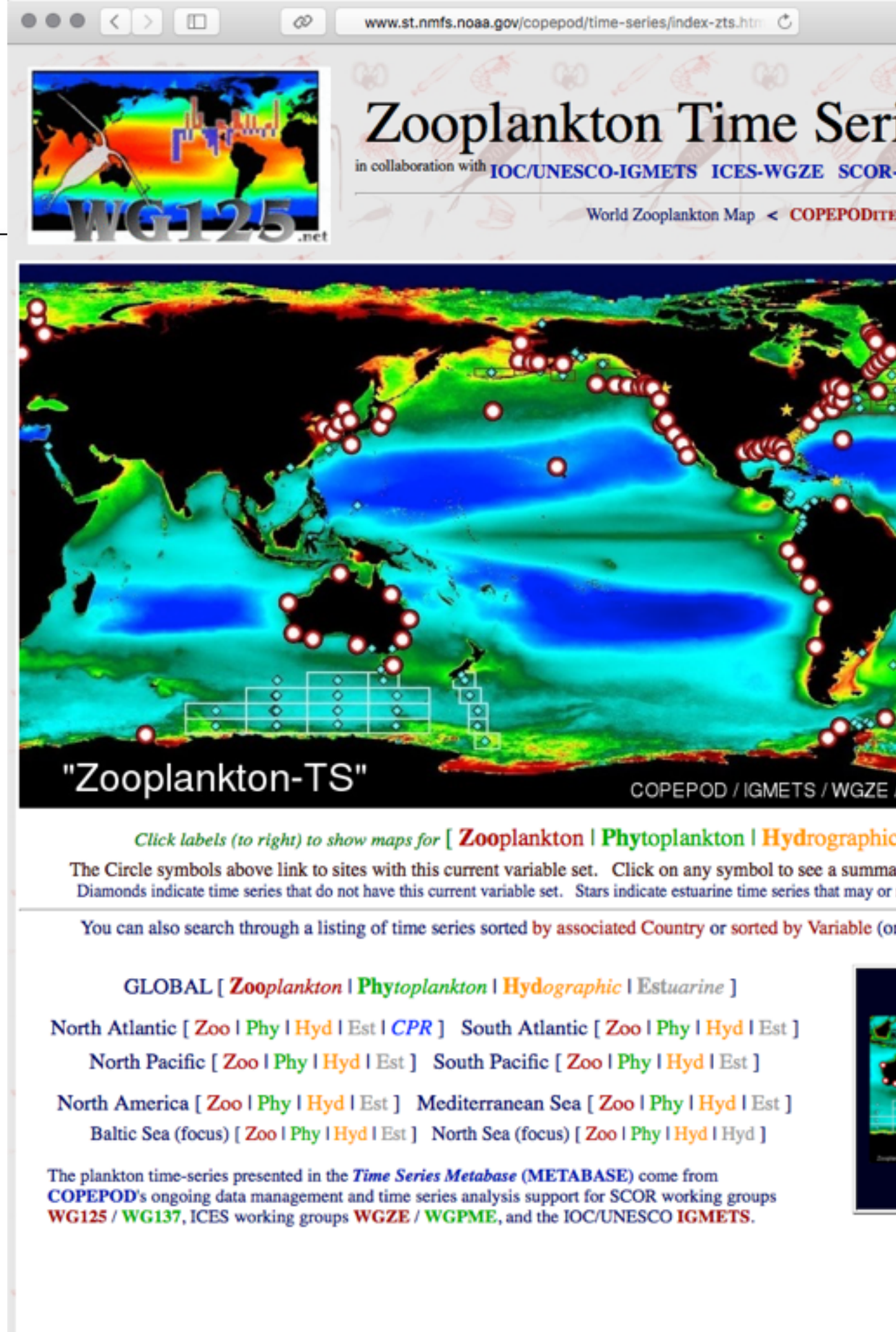


# Bases de données

**Générales, internationales**
http://www.iobis.org
http://www.iobis.org/mapper/
1 (n, env)

**Zooplancton, internationales**
http://www.st.nmfs.noaa.gov/copepod/
http://www.st.nmfs.noaa.gov/copepod/time-series/
n, volume, env mais ancien

**Zooplancton, spécifiques**
CPR https://www.sahfos.ac.uk
Villefranche http://ecotaxa.obs-vlfr.fr
n (forme)



# Bases de fichiers, campagnes

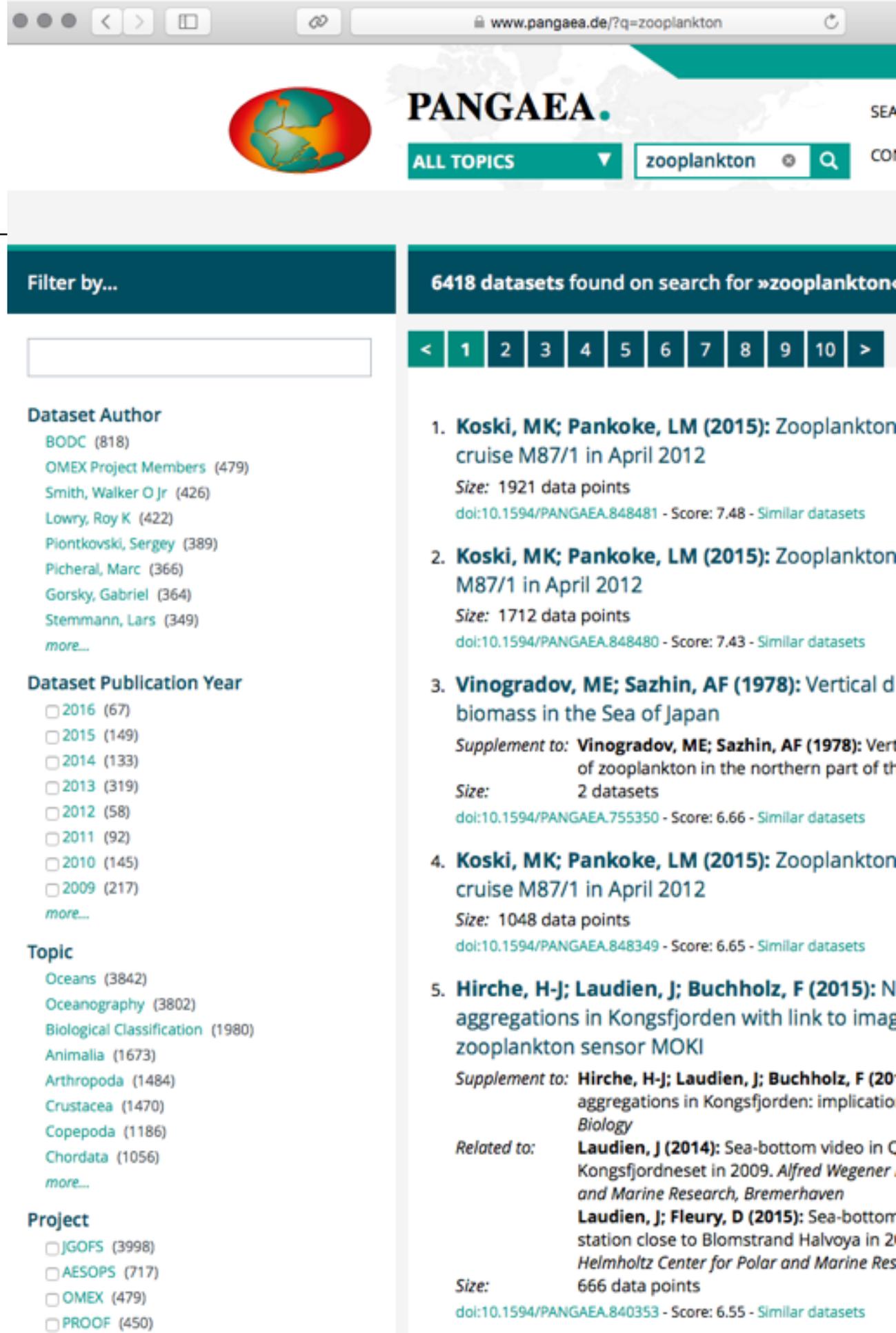
https://www.pangaea.de

https://www.pangaea.de/?q=zooplankton

https://www.pangaea.de/?q=maredat

http://www.seanoe.org/search

http://www.seanoe.org/data/00326/43749/



# Bases de fichiers, campagnes

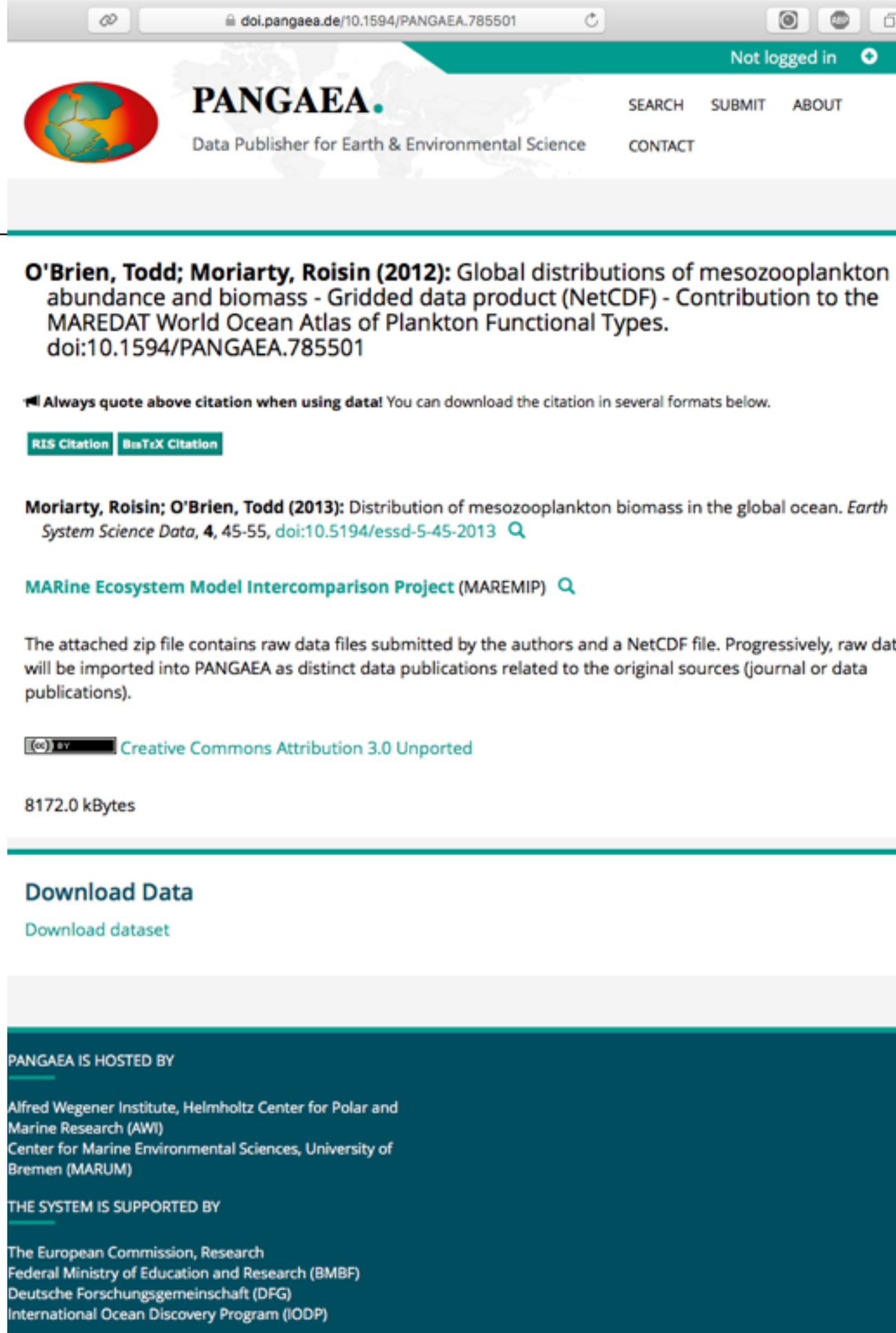
https://www.pangaea.de

https://www.pangaea.de/?q=zooplankton

https://www.pangaea.de/?q=maredat

http://www.seanoe.org/search

http://www.seanoe.org/data/00326/43749/



# Bases de fichiers, campagnes

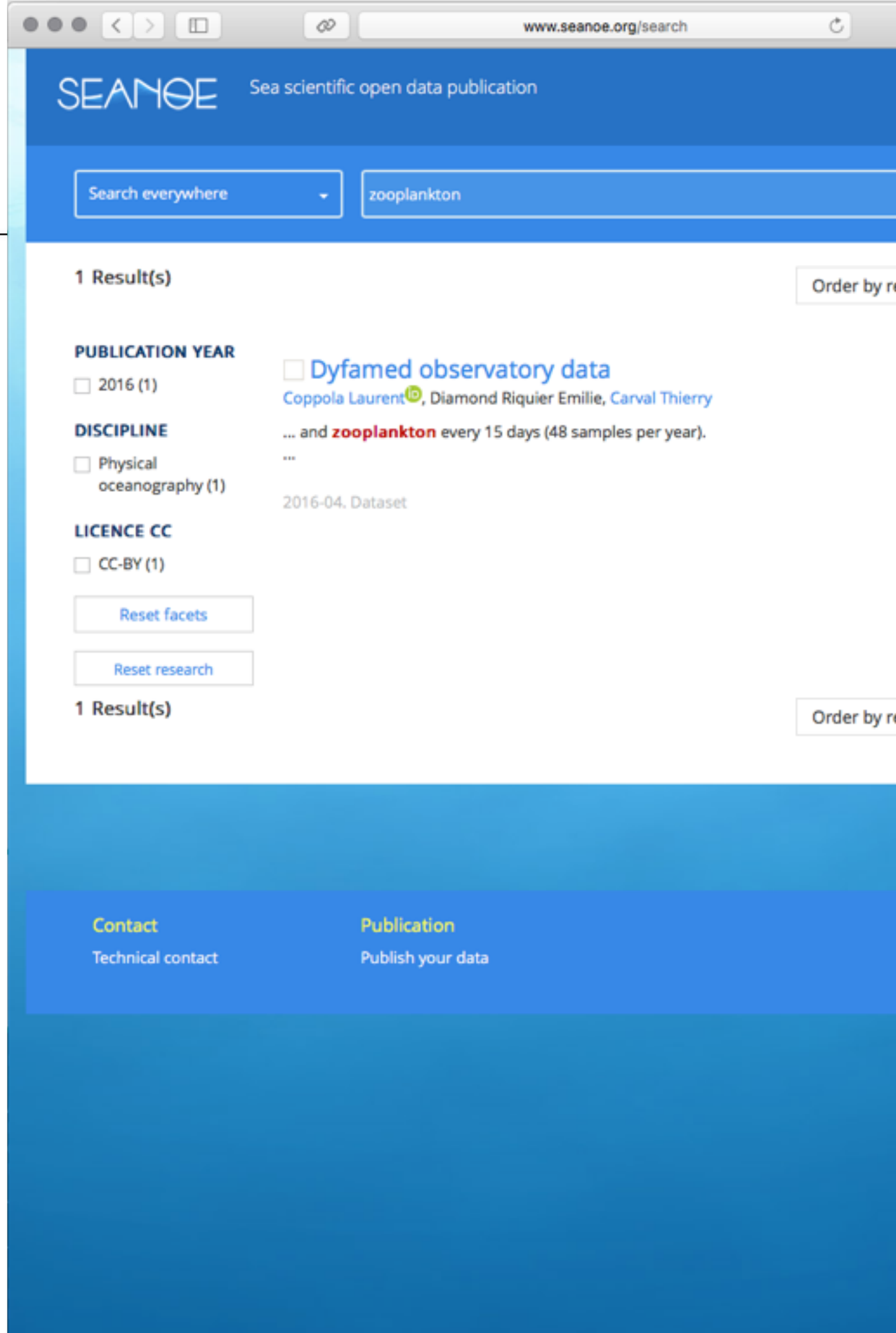
https://www.pangaea.de

https://www.pangaea.de/?q=zooplankton

https://www.pangaea.de/?q=maredat

http://www.seanoe.org/search

http://www.seanoe.org/data/00326/43749/



# Bases de fichiers, campagnes

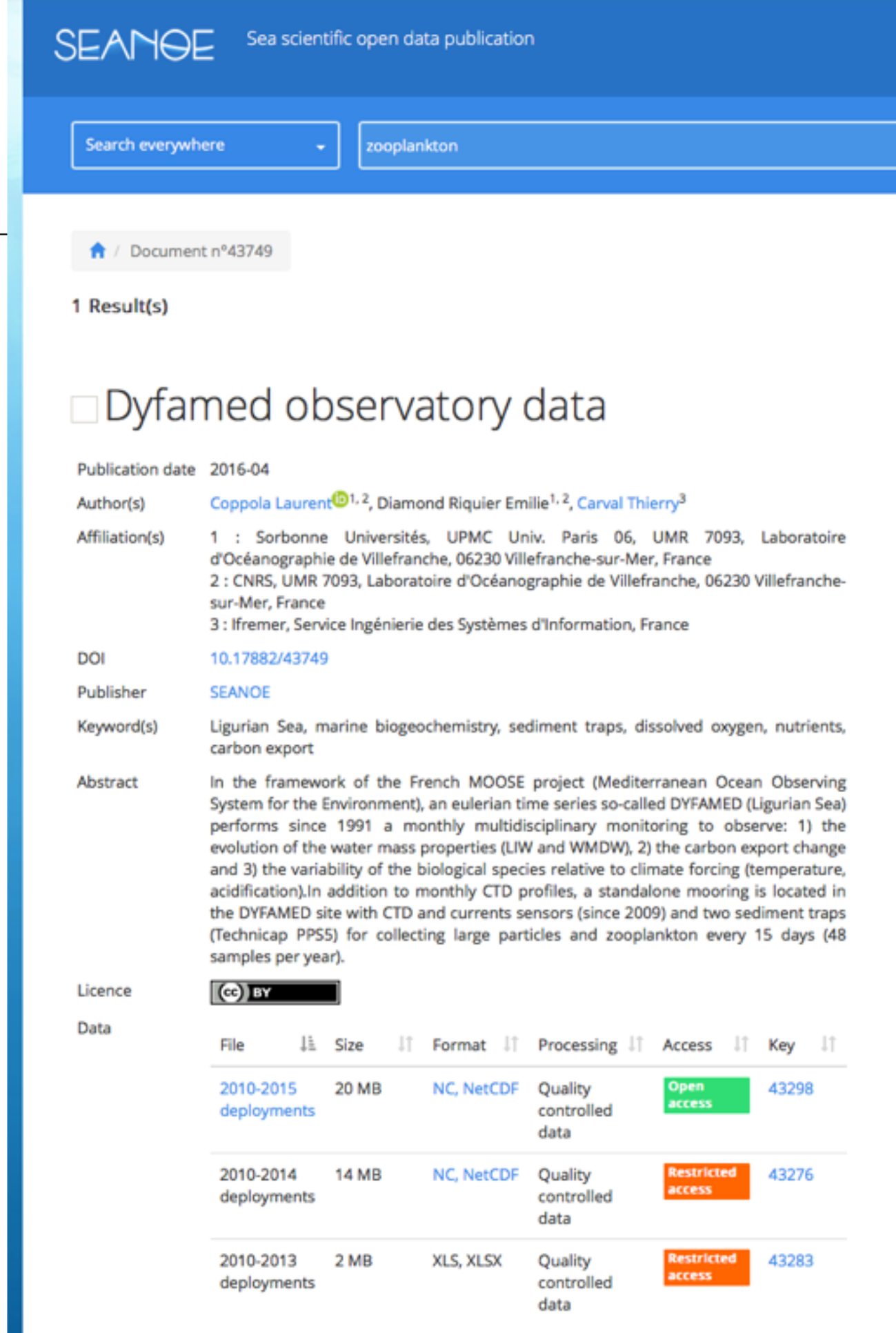
https://www.pangaea.de

https://www.pangaea.de/?q=zooplankton

https://www.pangaea.de/?q=maredat

http://www.seanoe.org/search

http://www.seanoe.org/data/00326/43749/



# Pôles et bases de données nationaux

Pôle ODATIS http://www.odatis-ocean.fr contient

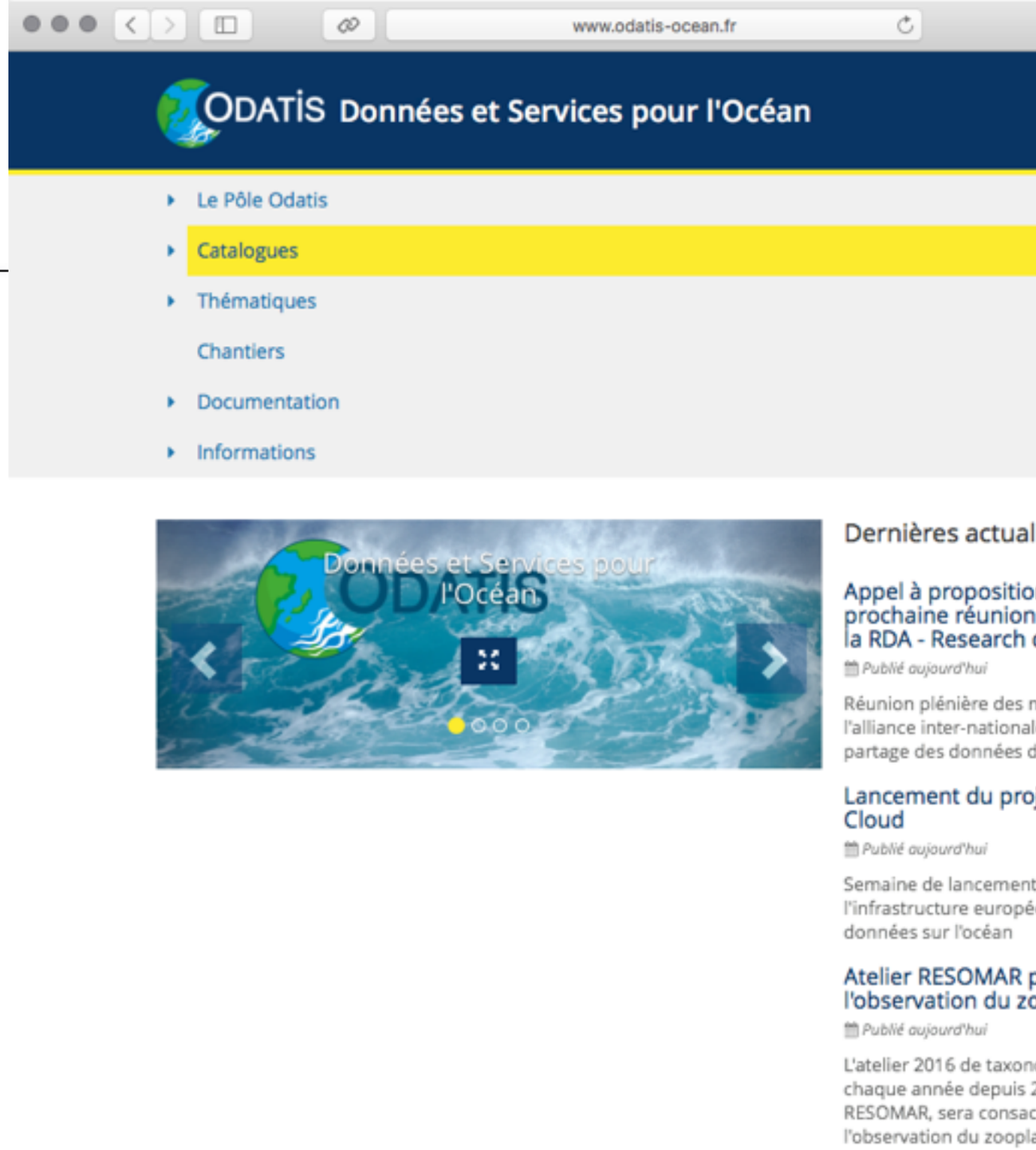
SEANOE, OASU (Bordeaux), OMP/ SEDOO (Toulouse), SISMER, UPMC (OOV, SBR, OOV, LOCEAN), LEFE-CYBER http://www.obs-vlfr.fr/proof/cruises.php

Pelagos et Benthos de RESOMAR (mais non consultable actuellement)

Quid de:

Quadrige[2] (IFREMER)

Autres?




# Pôles et bases de données nationaux

Pôle ODATIS http://www.odatis-ocean.fr contient

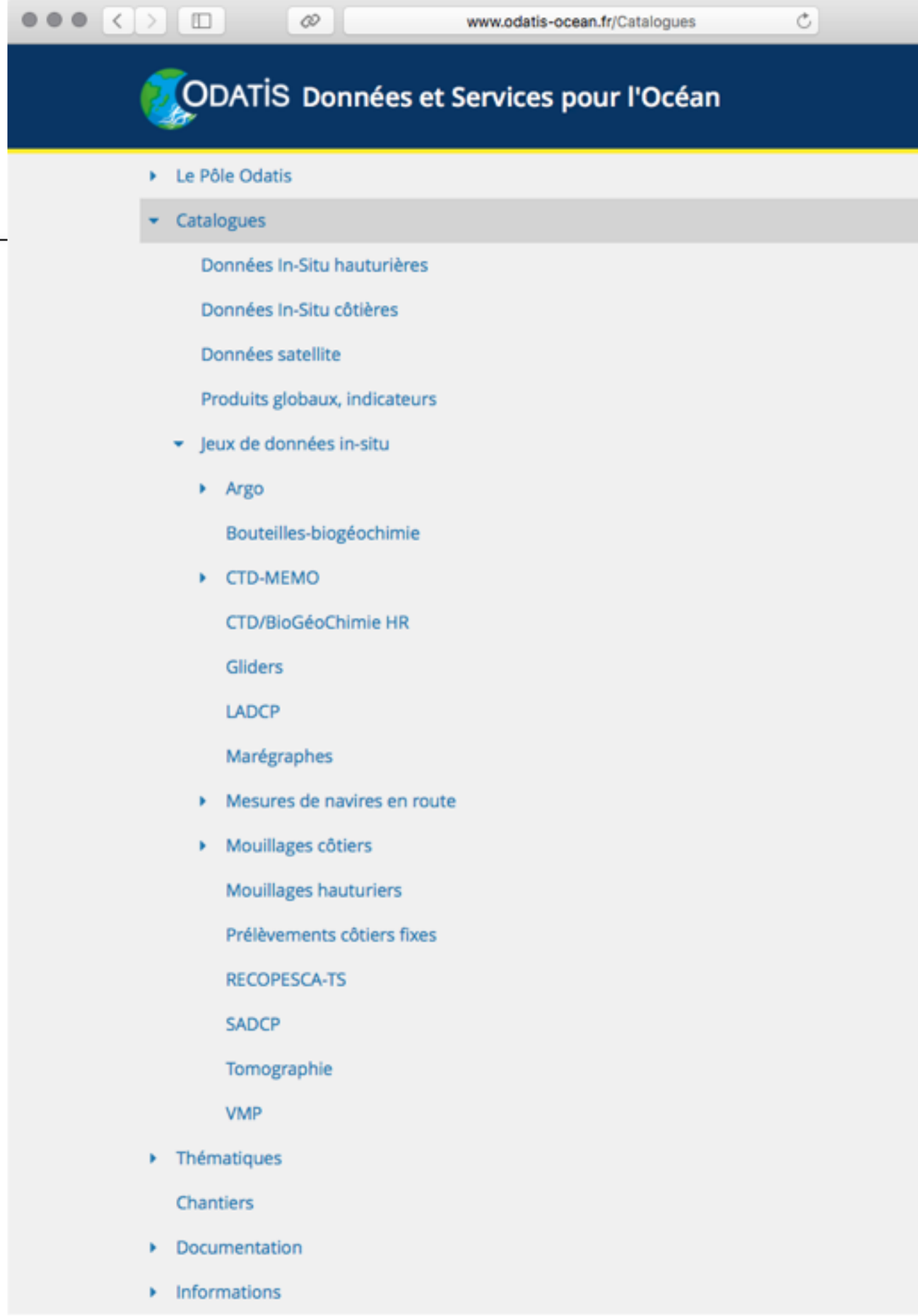
SEANOE, OASU (Bordeaux), OMP/ SEDOO (Toulouse), SISMER, UPMC (OOV, SBR, OOV, LOCEAN), LEFE-CYBER http://www.obs-vlfr.fr/proof/cruises.php

Pelagos et Benthos de RESOMAR (mais non consultable actuellement)

Quid de:

Quadrige² (IFREMER)

Autres?



# Pôles et bases de données nationaux

Pôle ODATIS http://www.odatis-ocean.fr contient

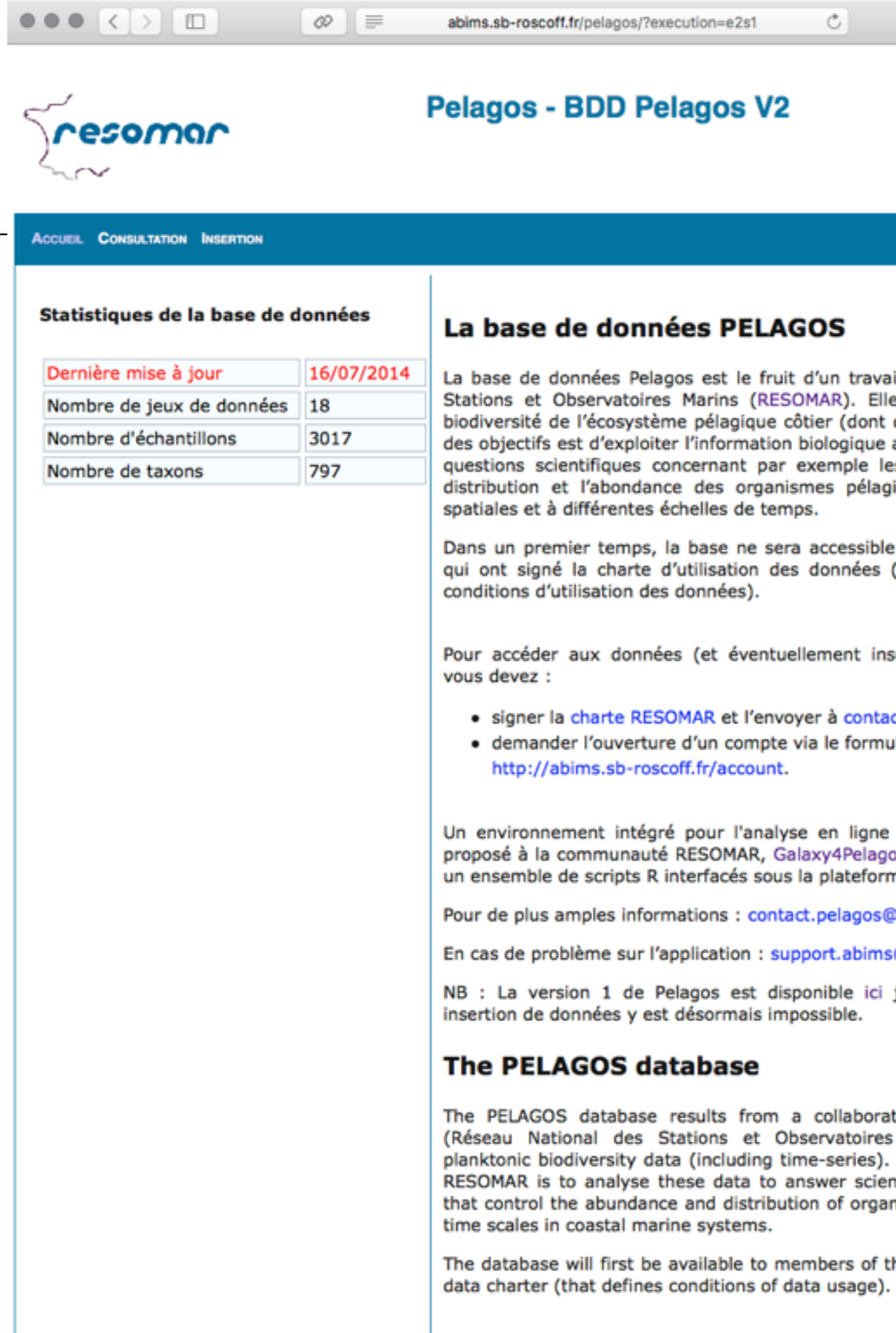
SEANOE, OASU (Bordeaux), OMP/ SEDOO (Toulouse), SISMER, UPMC (OOV, SBR, OOV, LOCEAN), LEFE-CYBER http://www.obs-vlfr.fr/proof/cruises.php

Pelagos et Benthos de RESOMAR (mais non consultable actuellement)

Quid de:

Quadrige² (IFREMER)

Autres?

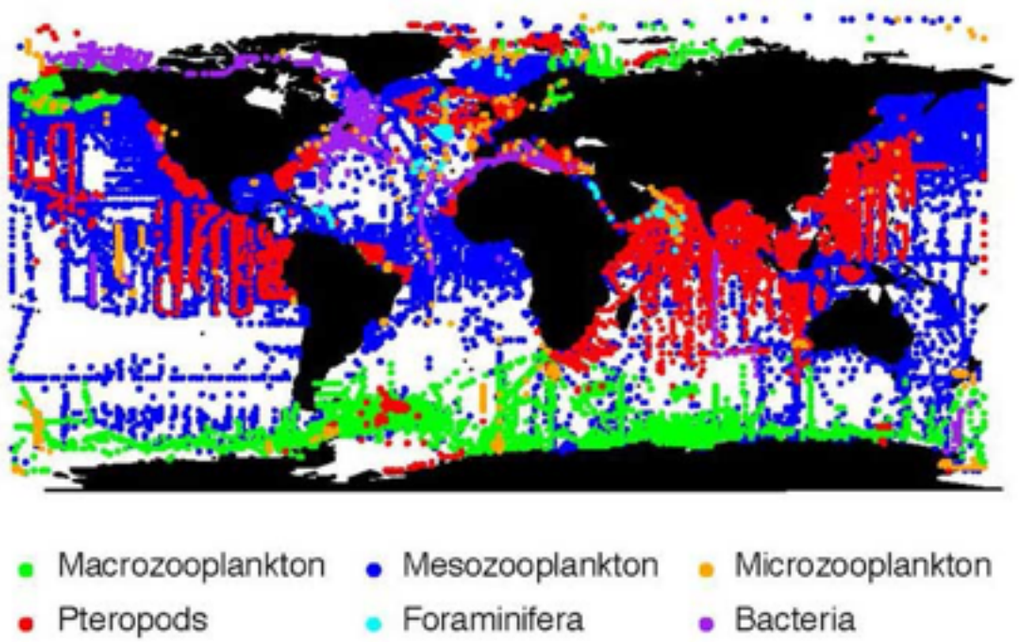


# Jugement personnel des entrepôts de données

Pour présences: OBIS

Pour abondances: MAREDAT

Pour suivi des données: SEANOE



# Format des données

Divers mais SeaDataNet semble consensuel

http://www.seadatanet.org/Standards-Software/Data-Transport-Formats

Minimum

| date | espèce | conc. | | *lat* | *lon* | *prof.* |
|---|---|---|---|---|---|---|
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

# SeaDataNet mode ODV

| Colonne | Contenu | Co |
|---|---|---|
| Cruise | 001-P-015 - Point 1 Dunkerque | 001-P-015 - P |
| Station | 1001022_BIO_Surf | 100102 |
| Type | * | |
| yyyy-mm-ddThh:mm:ss.sss | 1992-01-07T14:00:00.000 | 1992-02-0 |
| Longitude [degrees_east] | 2.3334994 | 2.3 |
| Latitude [degrees_north] | 51.0686493 | 51. |
| LOCAL_CDI_ID | 1001022_BIO_Surf | 100102 |
| EDMO_code | 486 | |
| Bot. Depth [m] | 0 | |
| MinimumObservationDepth [m] | 0.5 | |
| QV:SEADATANET | 1 | |

| | | |
|---|---|---|
| LOCAL_CDI_ID | 1001022_BIO_Surf | 100102 |
| EDMO_code | 486 | |
| Bot. Depth [m] | 0 | |
| MinimumObservationDepth [m] | 0.5 | |
| QV:SEADATANET | 1 | |
| MaximumObservationDepth [m] | 0.5 | |
| QV:SEADATANET | 1 | |
| SampleID:INDEXED_TEXT | 50021 | 5 |
| QV:SEADATANET | 1 | |
| SamplingEffort:INDEXED_TEXT | | |
| QV:SEADATANET | 9 | |
| SubsampleID:INDEXED_TEXT | 50021 | 5 |
| QV:SEADATANET | 1 | |
| SubSamplingCoefficient:INDEXED_TEXT | 1 | |
| QV:SEADATANET | 1 | |

| | | |
|---|---|---|
| SubSamplingCoefficient:INDEXED_TEXT | 1 | |
| QV:SEADATANET | 1 | |
| ScientificName:INDEXED_TEXT | Prorocentrum | Gym |
| QV:SEADATANET | 2 | |
| ScientificNameID:INDEXED_TEXT | urn:lsid:marinespecies.org:taxname:109566 | urn:lsid:marine<br>1 |
| QV:SEADATANET | 2 | |
| Sex:INDEXED_TEXT | Not Specified | Not |
| QV:SEADATANET | 9 | |
| LifeStage:INDEXED_TEXT | Not Specified | Not |
| QV:SEADATANET | 9 | |
| ObservedIndividualCount [#] | | |
| QV:SEADATANET | 9 | |
| IndividualCountperLiter [#/l] | 300 | |
| QV:SEADATANET | 2 | |

# Contrôle qualité

## Sur les identifications

Culverhouse, 2003 (dinoflagellés)

- 67 à 83% pour personnel entrainé
- 84 à 95% pour experts

Tara Zooscan

- subset aléatoire de 5%, vérification collégiale $\Longrightarrow$ taux d'erreur ~2%

## Sur les abondances

COPEPOD

- range check (lat, lon)
- par taxon/group: n $\notin$ mean ± 5 SD → investigation

GBIF

- Field error expected to be 1 to 5%

# Contrôle qualité

Chapman, A. D. 2005. Principles of Data Quality, version 1.0. Report GBIF.

Chapman, A. D. 2005. Principles and Methods of Data Cleaning – Primary Species and Species- Occurrence Data, version 1.0. Report GBIF http://www.gbif.org/resource/80528

## Error sources

spatial (temporal)

nomenclatural (taxa names)

data duplication

## Principles

Plan a data cleaning policy and strategy

Prevention is better than cure

Prioritise, do not duplicate

Everyone is responsible (collector, curator, user)

Accountability, transparency and documentation

# Flags SeaDataNet

| Flag | Signification | Explication |
|---|---|---|
| 0 | no quality control | |
| 1 | good | |
| 2 | probably good | Probably consistent but cannot be checked; associated with defect but small |
| 3 | probably bad | Probably inconsistent |
| 4 | bad | Obviously erroneous |
| 5 | changed | Adjusted during quality control |
| 6 | below detection | |
| 7 | excess | Too large to be measured by technique |
| 8 | interpolated | |
| 9 | missing | |
| A | ID uncertain | Uncertainty in identification |

# Visualisation des données

https://www.sahfos.ac.uk/data/data-charts/

https://www.sahfos.ac.uk/data/map-data/

http://www.obs-vlfr.fr/data/view/zoo/b/wp2/

Mise à disposition:
attention à ne pas faire la science à la place des gens!