

DATA  
TERRA



ODATIS

# Facilitating access to data and analytics services: the strategy of ODATIS, the gateway to French open ocean and coastal data

Sabine Schmidt<sup>1</sup>, Gilbert Maudire<sup>2</sup>, Valérie Harscoat<sup>2</sup>, Cécile Nys<sup>2</sup>, Joël Sudre<sup>3</sup>, Caroline Mercier<sup>4</sup>, Gérald Dibarboure<sup>5</sup>, Frédéric Huynh<sup>3</sup>

<sup>1</sup> CNRS, UMR5805 EPOC, Pessac, France

<sup>2</sup> Ifremer, Centre de Bretagne, Plouzané, France

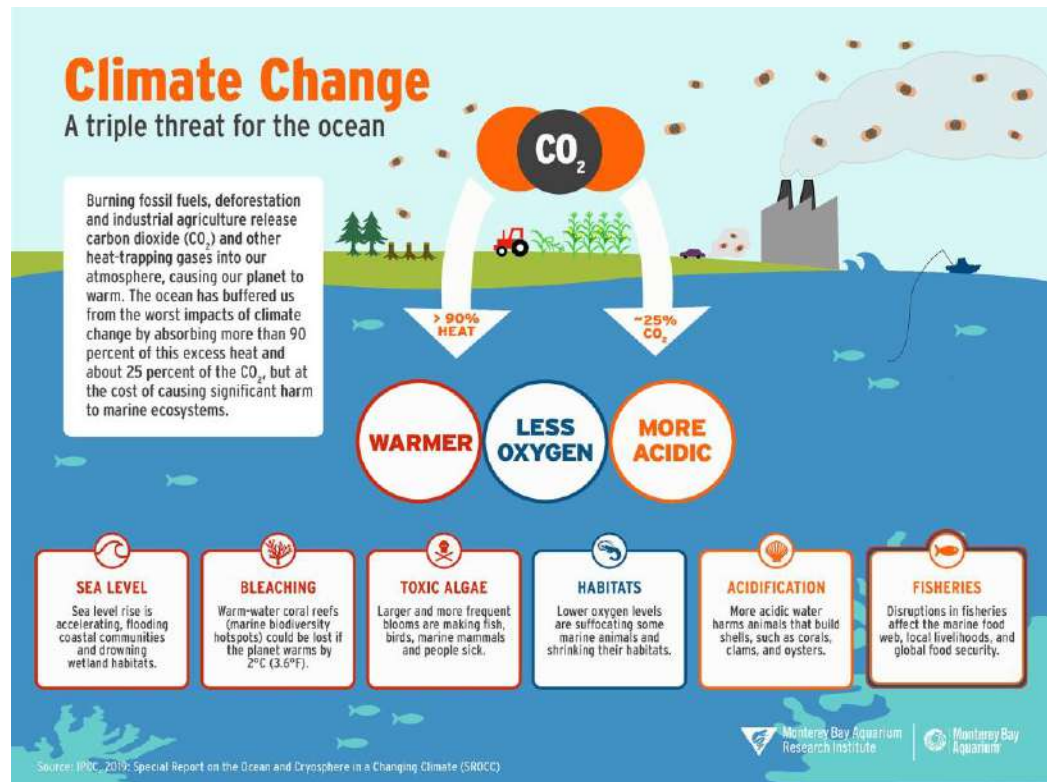
<sup>3</sup> UMS CPST, Montpellier, France

<sup>4</sup> Akka, Toulouse, France; <sup>5</sup> CNES, Toulouse, France

# Global change

Since the industrial revolution, the imprints of human activities on the global environment have intensified, leading to the assignation of the term “Anthropocene” to the present.

The ongoing and expected consequences of the global change on the ocean are multiple.



# A critical need to better understand and forecast the impacts of global change

Observations are needed at all the stages of the scientific process:  
description, understanding, modelling and forecasting

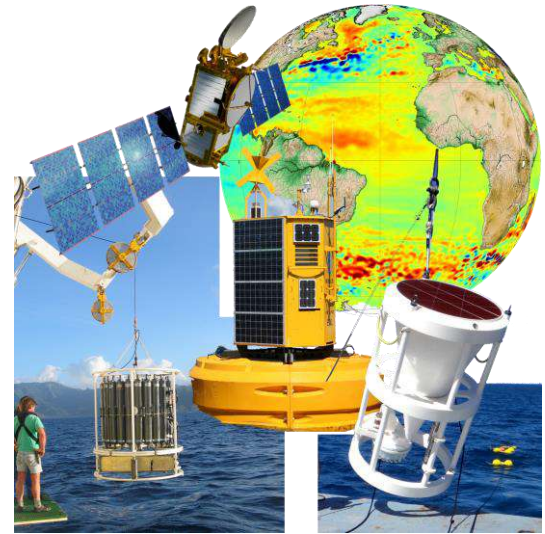
Considering that:

- the acquisition of marine data is difficult and expensive :  
Indeed it requires access to remote sites and many technical tools (research vessels, instrumented sites, gliders ...)
- without a professional archiving, > 30% of the data are lost (somewhere on a labtop...) or unusable ten years after their acquisition (source: Ifremer).

➡ The preservation of marine observation is a major issue for marine research

# Challenges for marine data access

The past few decades have seen a marked acceleration in the number of marine observations, both by using in situ and remote sensing measurements



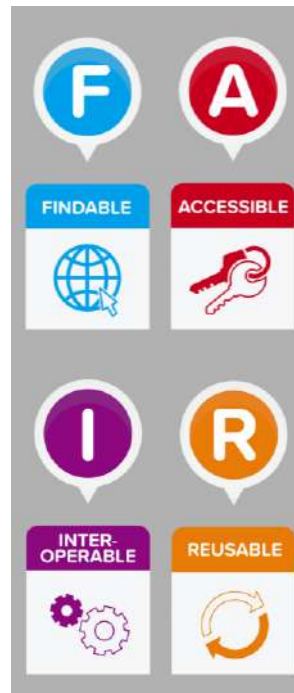
In order to make the most of this unprecedented flow of data for the benefit of knowledge and society, the procedures and policies of all data centers need to be harmonize.

# Challenges for marine data access

## Challenge 1: the data quality (FAIR principles)

Metadata and data must be easy to find and (re)use (describe your data, apply persistent identifiers)

To be integrated with other dataset (ie workflows for analysis/processing. (open format, consistent vocabulary, metadata standards)



Consider what will be shared and how it can be accessed

Data must then be Reusable, with well-described metadata and appropriate licence

# Challenges for marine data access

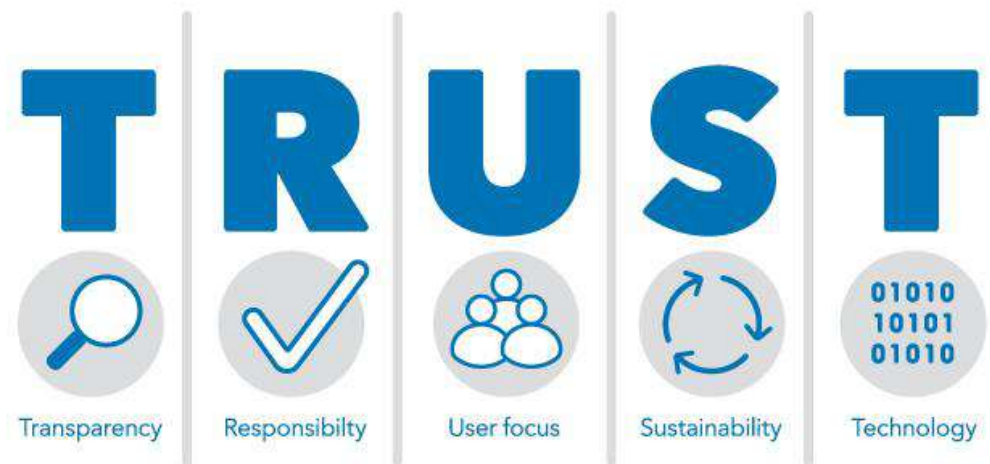
## Challenge 2: to guide data repositories toward a certification process

in order to reach the Core Trustworthy Data Repositories requirements of the RDA Core Trust Seal which provides a common framework to implement and maintain digital repositories

Certification is important to ensuring:

- ✓ the reliability and durability of data repositories,
- ✓ the potential for sharing data over a long period of time

for both their users and their funders



# Challenges for marine data access

## Challenge 3: to get people to share data

The International Council for Science (ICSU, 2011) already promoted “*full and open access to scientific data, especially when the research is publicly funded. Scientists should carry out research and disseminate their results with integrity and openness to maximise the benefits and minimise the possible harms of science for present and future generations*”

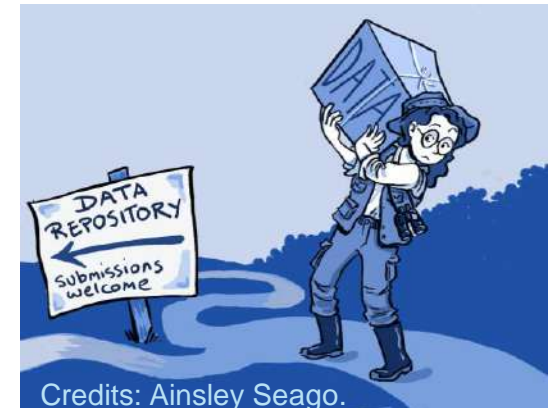
Researchers can be reluctant to share their data publicly because of real and/or perceived individual costs

feeling of

- loss of control over data,
- constraints that don't give back any value,

inadequate

- IT/human resources
- training.



**The data repositories need to make data sharing easier for the data owner too, not just the user/analyst.**

# The need of interoperable infrastructures

- In order to accelerate the collect, dissemination and intelligent use of **data**, there is a need of **interoperable infrastructures** for **helping**:
  - the **producers** to archive and share their data,
  - the **users** to get relatively easy access to data
- To be coordinated at least at the national, even international, level.





# The need of interoperable infrastructures

- In order to accelerate the collect, dissemination and intelligent use of **data**, there is a need of **interoperable infrastructures** for **helping**:
  - the **producers** to archive and share their data,
  - the **users** to get relatively easy access to data
- To be coordinated at least at the national, even international, level.



Here, we illustrate a national initiative of a portal dedicated to French marine data:  
**Ocean DATA Information and Services**



# ODATIS: Ocean DATA Information and Services

ODATIS is the ocean hub of the French Research Infrastructure for Earth data (Data Terra).

Launched in December 2017

[www.odatis-ocean.fr](http://www.odatis-ocean.fr)  
[www.odatis-ocean.fr/en/](http://www.odatis-ocean.fr/en/)

ODATIS has the ambition to become the unique entry point to access all the French open ocean and coastal observation data for the benefit of knowledge and society.

**DATA TERRA** RESEARCH INFRASTRUCTURE  
data-terra.org

**ODATIS**  
OCEAN Hub  
odatis-ocean.fr/en

**AERIS**  
ATMOSPHERE Hub  
en.aeris-data.fr

**ForM@Ter**  
SOLID Earth Hub  
en.poleterresolide.fr/

**Theia**  
LAND SURFACE Hub  
theia-land.fr/en

# ODATIS roadmap

- to offer a global view on both in-situ and satellite observations and their derived products;
- to facilitate access, through an single portal, to fully described and qualified databases, in agreement with the current scientific standards;
- to ensure the long-time preservation of datasets;
- to cross the space, time and discipline frontiers by ensuring the interoperability of datasets;
- to promote combined uses of data from different nature (in-situ/satellite) or origin (operational networks/scientific experiments);
- to enable the extraction of information from the databases by proposing exploration, extraction and analysis tools, as well as computing facilities.

# ODATIS structuration

## STRATEGIC LEVEL

### Inter-institution Steering Committee



### Data Terra Executive Board

Defines a common strategy for all the thematic clusters to provide an unified access to data (interoperability, vocabulary), products, softwares, tools and services.

### Scientific Council

Experts on marine and coastal sciences.  
Represent the user scientific community, express the needs and recommend scientific consortia.

## EXECUTIVE BOARD

Management team



Representative of each DSC

Director  
Technical director  
Scientific director  
Assistants

## Data and Service Centres (DSCs)

Catalogue: web & client: CNES  
server & catalogue: Ifremer

### Data:

#### Satellite (CDS-SAT)

CERSAT, AVISO

#### In situ (CDS-in situ)

Coriolis, SHOM, SISMER

OMP, OASU, IMEV. IMEV, SBR

## Technical Workshop

Proposes technical guidelines (interoperability, services, ...), Organizes practical sessions to test technics and exchange experiences among CDSs.

## Scientific Consortia

Propose and instruct new thematic data compilation (ex: EOVS) or new product on the request of the scientific community.

## OPERATIONNAL LEVEL



# ODATIS data catalog

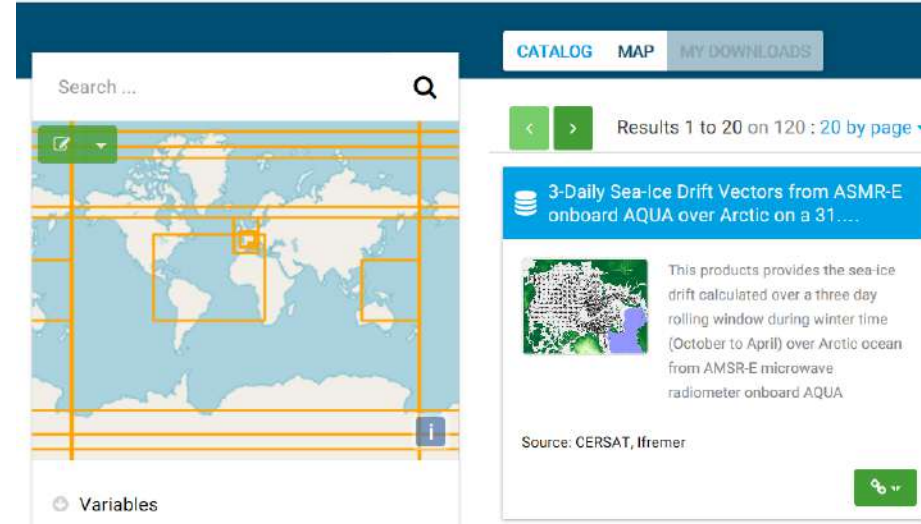
The catalogue includes the variables of:

all the marine disciplines

physic, chemistry, biology...

whatever the technique used

satellites, in-situ observatories, field cruises, analyses in lab).



The catalogue offers data access though different services:

- search with selection filters,
- data description (Preview or Complete),
- visualization,
- the possibility to download data (directly or via the local partner portals).

# SEANOE for orphan data

## SEANOE

SEA scieNtific Open data Edition, [www.seanoe.org](http://www.seanoe.org) )  
a publisher of scientific data in the field of marine sciences.








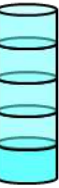
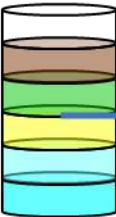


SEANOE provides to each published dataset a DOI which can be cited in a publication in a reliable and sustainable way.

Data are published in free access in SEANOE under the Creative Commons licenses.

Possibility of an embargo period (max. 2 years) to restrict access to data of a publication under review, for example .

# The need to develop a typology of data centers

The implementation of such e-infrastructure requires to define and organize a typology of data centers in a network in order to optimize the required IT and human resources

SERVICE	data repository	production	on demand
Involved structures			 
USER	data repositories doi, licences reporting on data use	combination of different marine dataset ( <i>in situ</i> / satellite) from the same thematic or area.	data analyses and interpretation cross analyses of different data from all Earth compartments
BACK OFFICE	<p>Data Assembling Centres</p> <ul style="list-style-type: none"> <li> Close to the producer</li> <li> Common catalogue and vocabulary servers</li> <li> Long-term archive</li> </ul>	<p>Data &amp; Service Centres</p> <p>National data hub Aggregates large collections at the national minimum level</p> 	<p>Virtual Research Environments (VRE)</p> <p>data lake</p>  <p>or</p>  <p>temporary personal storage</p>  <div style="border: 1px solid black; padding: 5px; width: fit-content;"> <p>TOOL BOX</p> <p>softwares, machine learning, ...</p> </div>



**DATA  
TERRA**



**ODATIS**

Thanks  
for your attention



14/10/2020

[contact@odatis-ocean.fr](mailto:contact@odatis-ocean.fr) | [www.odatis-ocean.fr](http://www.odatis-ocean.fr)