



**HAL**  
open science

## Streamlining Data and Service Centers for Easier Access to Data and Analytical Services: The Strategy of ODATIS as the Gateway to French Marine Data

Sabine Schmidt, Gilbert Maudire, Cécile Nys, Joël Sudre, Valérie Harscoat, Gérald Dibarboure, Frédéric Huynh

### ► To cite this version:

Sabine Schmidt, Gilbert Maudire, Cécile Nys, Joël Sudre, Valérie Harscoat, et al.. Streamlining Data and Service Centers for Easier Access to Data and Analytical Services: The Strategy of ODATIS as the Gateway to French Marine Data. *Frontiers in Marine Science*, 2020, 7, 10.3389/fmars.2020.548126 . hal-03087995

**HAL Id: hal-03087995**

**<https://hal.science/hal-03087995v1>**

Submitted on 24 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Streamlining Data and Service Centers for Easier Access to Data and Analytical Services: The Strategy of ODATIS as the Gateway to French Marine Data

Sabine Schmidt<sup>1\*</sup>, Gilbert Maudire<sup>2</sup>, Cécile Nys<sup>2</sup>, Joël Sudre<sup>3</sup>, Valérie Harscoat<sup>2</sup>, Gérald Dibarboure<sup>4</sup> and Frédéric Huynh<sup>3</sup>

## OPEN ACCESS

### Edited by:

Carol Robinson,  
University of East Anglia,  
United Kingdom

### Reviewed by:

Gwenaëlle Moncoiffé,  
British Oceanographic Data Centre  
(BODC), United Kingdom  
Toste Tanhua,  
GEOMAR Helmholtz Center for Ocean  
Research Kiel, Germany

### \*Correspondence:

Sabine Schmidt  
sabine.schmidt@u-bordeaux.fr

### Specialty section:

This article was submitted to  
Marine Biogeochemistry,  
a section of the journal  
Frontiers in Marine Science

**Received:** 01 April 2020

**Accepted:** 16 November 2020

**Published:** 11 December 2020

### Citation:

Schmidt S, Maudire G, Nys C,  
Sudre J, Harscoat V, Dibarboure G  
and Huynh F (2020) Streamlining Data  
and Service Centers for Easier  
Access to Data and Analytical  
Services: The Strategy of ODATIS as  
the Gateway to French Marine Data.  
*Front. Mar. Sci.* 7:548126.  
doi: 10.3389/fmars.2020.548126

<sup>1</sup> CNRS, Univ. Bordeaux, EPOC, EPHE, UMR 5805, Pessac, France, <sup>2</sup> Ifremer, Centre de Bretagne, Plouzané, France, <sup>3</sup> UMS CPST, CNRS 2013, IRD 1S26300, Montpellier, France, <sup>4</sup> CNES, Toulouse, France

The past few decades have seen a marked acceleration in the amount of marine observation data derived using both *in situ* and remote sensing measurements. For example, high-frequency monitoring of key physical-chemical parameters has become an essential tool for assessing natural and human-induced changes in coastal waters as well as their consequences on society. The number and variety of data acquisition techniques require efficient methods of improving data availability. The challenge is to make ocean data available via interoperable portals, which facilitate data sharing according to Findable, Accessible, Interoperable, and Reusable (FAIR) principles for producers and users. Ocean Data Information and Services (ODATIS) aims to become a unique gateway to all French marine data, regardless of the discipline (e.g., physics, chemistry, biogeochemistry, biology, sedimentology). ODATIS is the ocean cluster of the Data Terra research infrastructure for Earth data, which relies on a network of data and service centers (DSC) supported by the major French oceanic research organizations (CNRS, CNES, Ifremer, IRD, SHOM; Marine Universities). ODATIS, through its components, is involved in European and international initiatives such as Copernicus, SeaDataCloud, and EMODnet. The first challenge of ODATIS is to catalog all open ocean and coastal data and facilitate data collection and access (discovery, visualization, extraction) through its web portal. A specific task is to develop tools for handling large amounts of data and generate products for policymakers, practitioners, and academics. This study presents the strategy used by ODATIS to implement the FAIR and CoreTrustSeal requirements in each of its DSCs and promote adherence within the scientific community (the main data producer) regarding the upload and/or use of data and suggestion of new products. A second challenge is to cover the end-user needs ranging from proximity to the producer to cross-analysis of data from all

Earth compartments. This involves defining and organizing a classification of DSCs in the network, which will be developed within the framework of the French Data Terra research infrastructure, the only framework capable of providing the necessary IT and human resources.

**Keywords:** ocean, data repository, interoperability, FAIR, data and service center

## INTRODUCTION

The signs of global change are undeniable, and there is a critical need to better understand and forecast the impacts for Earth and its inhabitants. Since the industrial revolution, the impacts of human activities on the global environment have intensified, leading to use of the term “Anthropocene” for the present geologic time period (Crutzen, 2002; Steffen et al., 2015). To answer the questions that people ask about their environment, the research community needs to address the “Earth system” as a whole, from the Earth’s core to the limits of the atmosphere, taking into account the interactions of each of its components and exploring all aspects ranging from geophysics to the biosphere (Future Earth, 2020). The ocean is the largest habitable compartment and plays a key role in regulating the Earth’s climate. The ongoing and expected consequences of global change on the ocean are numerous: rising temperatures and sea levels, stronger storms, acidification, marine heatwaves, deoxygenation, and impacts on ecosystems (e.g., Levin and Le Bris, 2015; Breitburg et al., 2018; Smale et al., 2019). However, identifying these impacts and changes is still difficult because of the large variability of the environment and limited availability of data. The past few decades have seen a marked acceleration in the number of open ocean and coastal observations derived using both *in situ* and remote sensing measurements (Charria et al., 2016; Le Reste et al., 2016; Rode et al., 2016; Tyler et al., 2016). For example, high-frequency monitoring of physical-chemical parameters (temperature, salinity, fluorescence, dissolved oxygen, etc.) has become an essential tool for assessing the natural and human-induced evolution of coastal waters, as well as their societal and management implications (Schmidt et al., 2017; Nichols et al., 2019). The number and variety of data acquisition techniques require effective tools to ensure that such large volumes of data are available to and usable by the research community and stakeholders.

Observations are required at all stages of the scientific process: description, understanding, modeling, and forecasting. Technological progress provides us with increasing capabilities to generate richer datasets (Buck et al., 2019). However, accessing, using, or combining these datasets can be complicated by the large variety of data types, their volume and format, the complexity of their underlying processing, their distribution, and their location. 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 harmonized. This is the objective of the findable, accessible, interoperable, and reusable (FAIR) data principles (Wilkinson et al., 2016; GO-FAIR, 2020). Appropriate approaches must also be defined to

acquire, process, archive, and distribute the validated data and products issued from Earth observations. The implementation of such an approach must be coordinated and implemented at least at national and international levels (Miguez et al., 2019; Tanhua et al., 2019; Braud et al., 2020). Here, we illustrate the national initiative of a portal dedicated to French marine data, named ODATIS (Ocean DATA Information and Services). We briefly present the challenges to be considered when developing an operational system that not only meets the needs of data producers and users but also the FAIR requirements. Created in 2017, ODATIS is based on existing data and service centers (DSCs), which is extremely complex in terms of its managed databases and services provided. After an introduction to the challenges involved in assessing marine data, we detail the structure of ODATIS before presenting the strategy implemented to help DSCs apply FAIR principles and guide them toward certification. Finally, we define a classification system of DSCs to optimize human and IT costs with respect to the services offered to users.

## BRIEF ASSESSMENT OF GUIDELINES FOR DEVELOPING MARINE DATA REPOSITORIES

Observing, understanding, and predicting the status, function, and evolution of the entire Earth system under global change is a fundamental research issue and crucial for the implementation of the Sustainable Development Goals adopted by all UN Member States in 2015 (UN SDGs). The acquisition of marine data is difficult and expensive because it requires access to remote sites and many technical resources (e.g., research vessels, instrumented sites, gliders). In addition, the ocean is a highly variable environment, with very short (e.g., waves, tides) to very long (e.g., climate change, species evolution, geology of the seafloor) periodicities. Therefore, the preservation of marine observations is a scientific challenge for the generation of time series data to illustrate this variability. Without professional archiving of observations, more than 30% of the data are lost or unusable 10 years after their acquisition (source: Ifremer). The “International Oceanographic Data and Information Exchange” (IODE) program of the “Intergovernmental Oceanographic Commission” (IOC) of UNESCO was established in 1961 to “enhance marine research, exploitation and development, by facilitating the exchange of oceanographic data and information between participating Member States, and by meeting the needs of users for data and information products.” Renowned oceanographic institutions

have had established data management units for several decades (e.g., Woods Hole Oceanographic Institution and the Biological & Chemical Oceanography Data Management Office)<sup>1</sup>. However, such an approach is not yet generalized in all countries. Large multi-national and multi-fieldwork programs have also suffered from a lack of centralized coordination (and funding) for data management. For example, the Joint Global Ocean Flux Study (JGOFS, 1987–2003) program was a pioneer in setting a data policy based on national JGOFS data managers. However, the final JGOFS International Data Collection (discrete datasets, volume 1, 1989–2000) was available on DVDs, whereas some countries published their data online. This limited the discovery of this important dataset regarding the fluxes of carbon between the atmosphere, surface ocean, and ocean interior, as well as their sensitivity to climate change. Fortunately, an initiative of the World Data Centre PANGAEA permitted the compilation and harmonization of JGOFS datasets in Pangaea (for example, see Schmidt et al., 2002)<sup>2</sup>. More recently, the initiative of the European Marine Observation and Data network (EMODnet) has gathered European marine data, metadata, and data products “from fragmented and hidden marine observations and data stored in a myriad of data systems and repositories scattered all over Europe” (Miguez et al., 2019). This followed the recommendations of Buck et al. (2019) “to move beyond data portals to service-based architectures that combine data provenance, persistence and security,” “allowing users to configure and apply varied yet compatible ocean data services to build their own knowledge systems.”

The first challenge was to develop efficient sharing of data by adopting common protocols. A broad consensus has now been reached in the marine domain regarding the procedures and technologies required to implement FAIR principles for scientific data (Wilkinson et al., 2016), which were recently outlined for marine data (Tanhua et al., 2019). The FAIR Guiding Principles aim to improve the findability, accessibility, interoperability, and reuse of scientific data, with zero or minimal human intervention due to their increasing volume and complexity. **Table 1** describes each principle. The first principle is findable, i.e., metadata and data must be easy to find and (re)use. The second is accessible, that is, users must know how it can be accessed. Third, the data must be interoperable, i.e., able to be integrated with other datasets, especially in workflows for analysis and processing. Fourth, the data must then be reusable with well-described metadata.

A second challenge is to encourage people to share data. Data producers are increasingly pressured, in particular by research funders, to release their data (Gutmann et al., 2008). In 2011, the International Council for Science (ICSU) already promoted “full and open access to scientific data, especially when the research was publicly funded. Scientists should carry out research and disseminate their results with integrity and openness to maximize the benefits and minimize the possible harms of science for present and future generations.” However, the reasons why data owners do not share are multiple: loss of control over data, the

notion of constraints that do not provide any value, inadequate IT/human resources, and/or training. Therefore, it is necessary to explain to data producers the personal benefits of archiving data, that is, to make it easier for producers to reuse their own data, recognition through increased citation, and potential co-authorship. In addition to the obvious benefits of long-term data preservation through data archiving, the professional benefits of data sharing should also be promoted as a contribution to ocean knowledge. The data repositories need to make data sharing easier for the data owner, not just the user/analyst. Decisions to archive data are still often made by individual researchers, who may renounce the decision because of individual costs (Roche et al., 2014). To ensure sufficient quality to meet specific user needs, it is necessary to popularize already existing guidelines and resources, such as tools to clean and normalize data, and to include metadata that clearly describe the datasets’ content, prior to archiving, to enable faster and easier data sharing. Data owners also need feedback on data use, for example, through relevant metrics (e.g., number of downloads, citation).

In brief, in order to accelerate the collection, analysis, dissemination, and intelligent use of data, from national and international observations and across disciplines and Earth compartments, there is a need for interoperable infrastructures that help producers preserve and share their data and allow users to obtain relatively easy access to data. In the following section, we describe the implementation of FAIR and TRUST principles within the French Ocean cluster, which can serve as guidelines for further national or thematic data repositories.

## THE FRENCH NATIONAL INITIATIVE FOR EARTH OBSERVATION DATA

The French data and services hub for the Earth system, named Data Terra<sup>3</sup>, is the French response to the need for a research infrastructure for Earth observation data management and processing. The Data Terra research infrastructure is a priority of the national roadmap of the French Ministry for Higher Education and Research (MESRI). Data Terra includes four thematic clusters corresponding to each main Earth compartment: Ocean (ODATIS), Atmosphere (AERIS), Continental Surfaces (THEIA), and Solid Earth (FORM@TER) (**Figure 1**). The objectives of Data Terra, through its clusters, is to provide wide and unified access to data, products, software, tools, and/or services on the Earth system produced by the French scientific community. The terms of reference of Data Terra facilitate data access, improve data interoperability, and integrate data and knowledge. The targeted users are not only the academic community but also socio-economic actors or stakeholders responsible for implementing or evaluating public policies. Indeed, in addition to promoting a better understanding of the Earth system, environmental data also have a significant socio-economic impact in many fields, such as protection against natural hazards, water quality, and management of mineral resources or living resources. Data

<sup>1</sup> [www.bco-dmo.org](http://www.bco-dmo.org)

<sup>2</sup> <https://www.pangaea.de>

<sup>3</sup> [www.data-terra.org](http://www.data-terra.org)

**TABLE 1** | The FAIR principles for scientific data repository.

Principle	Description
Findable	F1. (Meta)data are assigned a globally unique and persistent identifier F2. Data are described with rich metadata (defined by R1 below) F3. Metadata clearly and explicitly include the identifier of the data they describe F4. (Meta)data are registered or indexed in a searchable resource
Accessible	A1. (Meta)data are retrievable by their identifier using a standardized communications protocol A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorization procedure, where necessary A2. Metadata are accessible, even when the data are no longer available
Interoperable	I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data
Reusable	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards

From GO-FAIR (2020).



**FIGURE 1** | Data Terra research infrastructure for Earth data and its four thematic clusters.

Terra also serves the international community through satellite missions, international monitoring networks, and partnerships for development.

ODATIS is the “Ocean” cluster of Data Terra and the new national gateway to access all French ocean data (Figure 1). Its aim is to promote and facilitate the use of all marine observation data collected by *in situ* and remote sensing measurements in open and coastal waters. The data managed by ODATIS include variables from all marine disciplines (physics, chemistry, biology,

etc.) measured from the coastal environment to the deep ocean by any technique (e.g., satellites, *in situ* observatories, field cruises, lab analyses). The objective of ODATIS is to describe, quantify, and understand the entire ocean, including offshore and coastal environments, from the perspective of processes such as thermohaline circulation, distribution of chemical species including carbon, biogeochemical cycles, ecosystem functioning, ocean evolution, ocean–climate relationships, and interactions with other components of the Earth system. To achieve this ambitious objective, ODATIS must be able to offer users the following:

- long-term preservation of datasets;
- easier access through a single portal to fully described FAIR databases: completeness and easy comprehension of data and metadata are required to ensure that sufficient information is available for end users to assess the quality of data according to current scientific standards;
- a global overview of *in situ* and remotely sensed observations and their derived products, in order to permit the user to fully discover data;
- interoperability of datasets, regardless of location, time, discipline, and compartment;
- the possibility of combining data of different type (*in situ*/satellite) and origin (observation networks/scientific experiments);
- assistance with extracting information from the databases by proposing exploration, extraction, analysis tools, and computing facilities.

ODATIS has a dedicated website<sup>4</sup> launched in December 2017 that already includes more than 120 data collections. The portal and associated services have a mirror version in French and English, the latter of which allows data reuse and interoperability.

<sup>4</sup><http://www.odatis-ocean.fr/en>



The ODATIS portal already offers several data access tools: a search service with selection filters, a data discovery service (with two options: "Preview" and "Complete"), a visualization service, and the possibility to download data directly or via local partner portals.

## STRUCTURE OF THE FRENCH OCEAN CLUSTER

### Governance of the French Ocean Cluster

To fulfill its mission, ODATIS is organized to efficiently interact with partner institutions and the overarching Research Infrastructure (Data Terra) regarding strategic aspects, and with its DSCs and users regarding operational aspects (Figure 2). The ODATIS executive board is the interface between these two levels. It is composed of a management team and a representative of each DSC. The management team consists of the director, two assistant directors (scientific and technical), *in situ* and satellite technical officers, *ad hoc* project officers, and an editorial manager. This team has continuous informal interactions as well as three formal meetings a year of the executive board.

There are three strategic levels: the Scientific Council (SC), the Data Terra executive board, and the Inter-Institution Steering Committee (IISC). The Scientific Council, under the supervision of the scientific director, has the role of a scientific advisory board. Its members, appointed by the steering committee, are experts in marine sciences and representative of the user communities. The tasks of the SC are to help the executive board identify the scientific community's needs, check the scientific quality of the data and products generated by the DSCs, and assess the progress of the scientific consortia in order to propose a prioritization for the prototyping and production phases. If necessary, the SC formulates recommendations to the steering committee. The second level is the Data Terra executive board. Data Terra relies on thematic clusters that need to be harmonized to permit data interoperability and combination. It is essential that common editorial and technical choices are defined collegially. ODATIS is also placed under the responsibility of the Inter-Institution Steering Committee, which is composed of one representative from each of the partner institutions. Lastly, the main prerogatives of the IISC are to define the strategy of the ODATIS cluster in order to achieve its objectives, particularly in terms of scientific policy and European and international positioning, to ensure that the needs expressed by the user community are taken into account, and to mobilize the human and financial resources necessary to develop ODATIS.

At the operational level, ODATIS is based on a network of geographically distributed DSCs (see section "Data and Service Centres") operated by major French oceanic research organizations (CNRS, CNES, Ifremer, IRD, SHOM, Marine Universities). To coordinate the DSCs, ODATIS organizes two to three technical workshops every year to define the technical orientations of the cluster, or runs practical training sessions, for example, on handling and testing data visualization tools. These workshops are mandatory for DSCs but are open to anyone who requests them within the limit of a reasonable

number of participants, which is typically 15–20. Finally, ODATIS relies on scientific consortia in order to promote and develop innovative processing methods and products for space, airborne, or *in situ* observations of the ocean and its interfaces (atmosphere, coastline) with the other thematic clusters. A scientific consortium is a group of public or private experts that conduct research or develop innovative methods for mobilizing observation data, producing prototypes of products, or operating these prototypes to produce specific data on coastal and open ocean issues around thematic fields (physical, chemical, and biological processes, ecosystems, ocean/atmosphere exchange, global approaches, resources, etc.) to meet the societal and environmental challenges of our time. The first consortium, the Dissolved Oxygen Scientific Expertise Consortium<sup>5</sup>, was implemented in 2019 in order to network and integrate scientific stakeholders at national and even international scales around the theme of deoxygenation of the offshore and coastal ocean and establish an exhaustive national database on oceanic dissolved oxygen. This work will be France's contribution to the international effort led by the IOC-UNESCO GO2NE and IOCCP network. Scientific consortia concerning other essential oceanic variables (carbon, salinity, color) and techniques (flow cytometry) are currently being implemented.

### Data and Service Centers

At present, there are nine DSCs, two of which are dedicated to satellite data and seven of which are dedicated to *in situ* datasets (Figure 2). *In situ* DSCs collect data from observations made via ships (or measured on samples in laboratories after the cruise), drifting, or moored systems. The tasks of a DSC are to record data, process data, control data quality, and provide routine access to marine data. Members of the ODATIS executive board and DSCs are involved with the coordination of several initiatives for marine data management from regional to international scales, including:

- regional and national programs, in particular the long-term national observation services of the CNRS. This concerns, for example, coastal ocean and nearshore observations conducted under the umbrella of the French research Infrastructure for littoral and coastal observation services, named ILICO (Cocquempot et al., 2019);

- marine European data infrastructures, in particular:

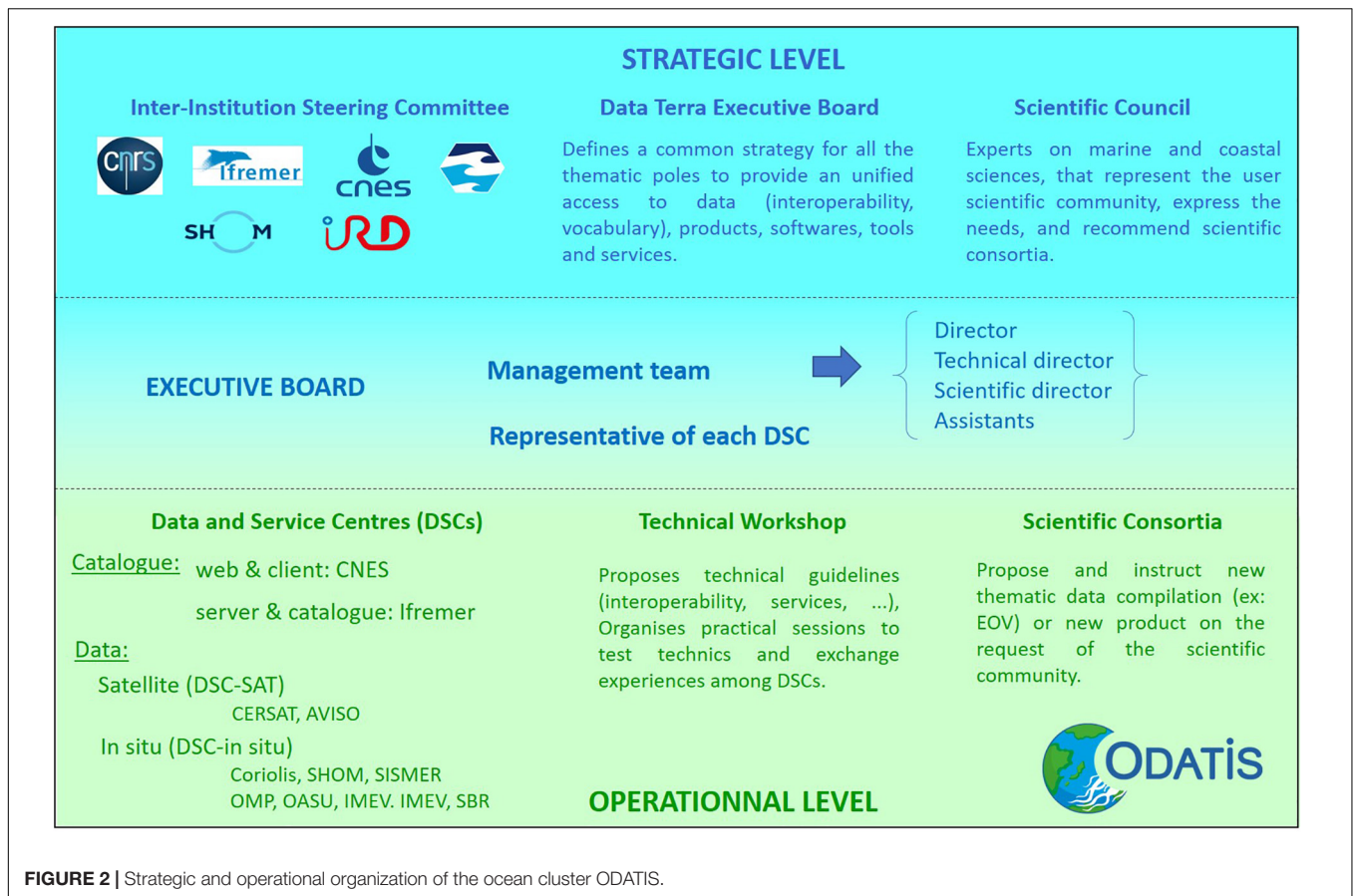
Copernicus Marine Environmental Monitoring Services (CMEMS), by coordinating the Marine *In Situ* Thematic Assemble Centre (INS-TAC) established via partners of the ODATIS Cluster<sup>6</sup>,

SeaDataNet infrastructure<sup>7</sup>, a network of marine data centers that defines, customizes, and implements marine data management procedures on a pan-European scale (120 pan-European data centers are now connected, including most ODATIS partners),

<sup>5</sup><https://www.odatis-ocean.fr/en/activities/scientific-expertise-consortium/ces-dissolved-oxygen>

<sup>6</sup><http://marine.copernicus.eu/>

<sup>7</sup><https://www.seadatanet.org/>



**FIGURE 2 |** Strategic and operational organization of the ocean cluster ODATIS.

EMODnet thematic lots (Miguez et al., 2019)<sup>8</sup>, the marine component of the EOSC-HUB (Marine Competence Centre).

- international programs such as ARGO (*in situ* observation by free-drifting profiling floats), GOSUD (Global Ocean Surface Underway Data; sea surface observation by ships of opportunity), IOC-IODE [International Oceanographic Data and Information Exchange of the Intergovernmental Oceanographic Commission], EMSO (European Multidisciplinary Seafloor and water column Observatory), OBIS (Ocean Biodiversity Information System), and interactions with the projects of the Space Agencies of Europe, the United States and China (ESA, NASA, CNSA).

Finally, ODATIS, through its data centers, contributes to the Environmental Research Infrastructures (ENVRI) community, which includes all European Earth Observation infrastructures, both for observation and data management. ENVRI defines guidelines to be implemented for compliance with FAIR principles, in collaboration with working groups of the Research Data Alliance (RDA) (e.g., metadata and catalogs, permanent identifiers such as DOI, FAIR controlled vocabularies). ODATIS benefits from all these collaborations. Common vocabularies (ontologies), metadata and data formats, and interoperability

protocols have been adopted and adapted to this thematic field and should be implemented by all DSCs. In addition, as marine data are environmental data with geo-references, used in public policies such as the EU Marine Strategy Framework Directive (MSFD), the technical requirements of the INSPIRE directive must be applied (INSPIRE, 2007). For example, the discovery metadata made available must conform to the European INSPIRE standard and appear in national geo-catalogs and geo-portals.

## IMPLEMENTATION OF FAIR AND TRUST PRINCIPLES

The implementation of FAIR principles is heterogeneous across DSCs. Several DSCs are responsible for datasets that contribute to key international data networks and supra-national infrastructures (e.g., ARGO, AVISO + (Archiving, Validation and Interpretation of Satellite Oceanographic data), SOCAT (Surface Ocean CO<sub>2</sub> Atlas), GOSUD (Global Ocean Surface Underway Data), SeaDataNet (a pan-European infrastructure to ease the access to marine data measured by the countries bordering the European seas). These DSCs already meet many of the FAIR criteria. However, the large majority of DSCs do not meet FAIR data criteria with regards to their data handling, formatting, and compliance. This is related to the genesis of ODATIS, which was built on existing DSCs and can be extremely

<sup>8</sup>[www.emodnet.eu/](http://www.emodnet.eu/)

complex in terms of managed databases and services provided. A diagnostic is that most observation datasets are archived and organized in databases deployed by the different DSCs; however, the coordination and harmonization of all ocean databases still needs to be implemented at the national level by ODATIS. At present, there is still huge heterogeneity in the databases between DSCs, mostly for those concerned with *in situ* data, which is a major obstacle for interoperability. In order to obtain support for upgrading *in situ* DSCs, the ODATIS management team replied to a recent national call for projects and was awarded the ANR COPiLOtE project (Toward the Certification of the Data and Service Centres of the Ocean Data Cluster 2020–2022), which began in April 2020. The main objective of COPiLOtE is to harmonize the implementation of FAIR principles within ODATIS *in situ* DSCs according to the tasks identified in **Table 2**. The first step is an individual self-evaluation of each DSC using a questionnaire, with the help of the executive office, to assess its level of implementation of FAIR principles. Then, the approach will be implemented in two volunteer *in situ* DSCs. At the end of the project, the implementation guidelines will be transposable to all *in situ* DSCs of ODATIS, that is, those not involved in the COPiLOtE project and future DSCs that will emerge from the community.

The second objective is to guide *in situ* DSCs toward certification in order to meet the Core Trustworthy Data Repositories requirements of the RDA CoreTrustSeal<sup>9</sup>. The TRUST Principles (i.e., transparency, responsibility, user focus, sustainability, and technology) provide a common framework to implement and maintain digital repositories (Lin et al., 2020). Certification is important for ensuring the reliability and durability of data repositories, and hence the potential for sharing data over a long period of time, for both users and their funders. We do not expect all DSCs to obtain certification at the end of the project. For illustration, there is currently only a single French marine data repository that is currently certified (IFREMER-SISMER, see the list of certified repositories at the link: [www.coretrustseal.org/why-certification/certified-repositories/](http://www.coretrustseal.org/why-certification/certified-repositories/)). However, our aim is to motivate DSCs to engage in the process, based on the premise that certification will soon become mandatory in order to obtain the support of funding institutions.

## DEVELOPING A CLASSIFICATION SYSTEM OF DATA CENTERS

As described above, ODATIS relies on a network of nine existing DSCs. These centers are not evenly distributed across the French territory and are mostly located in West France (Brittany, South-West). There is a rising demand from all French laboratories involved in marine sciences to have local data centers. For example, the Mediterranean Institute of Oceanography (MIO) on the Mediterranean coast is one of the most important French oceanography centers, which could justify having a DSC. However, is it reasonable to increase the number of existing

DSCs? There is a limit to the number of centers manageable by ODATIS, in terms of the efficiency of management of a distributed structure, staff requirements, and IT and software investments. Another difficulty involves the capacity to aggregate heterogeneous data (different volumes and storage centers) and therefore the ability to make data interoperable and FAIR at all levels. The challenge for ODATIS is now to develop a network that can cover the need to be close to the data producers, while providing cross-analytical tools for data from all Earth compartments for the end user. For this purpose, ODATIS and Data Terra are currently building a common strategy based on a classification system of data centers (**Figure 3**).

The first category is the data assembly center (DAC). DACs are small laboratory- or institution-based data units working closely with the data providers and focusing on making their data FAIR. It is indeed a critical point for data quality to have proximity with the producer. DACs will archive and enable access to the dataset through a persistent identifier (PI or PID), metadata, and common vocabulary. DAC repositories will be certified (CoreTrustSeal, ISO) to meet the repository criteria of scientific journals. DACs will also produce statistics on data usage. Such centers do not need large informatic resources and could be fairly widespread. However, there is a need to harmonize tools to optimize this service and ensure interoperability and Web harvesting.

The second category is the DSCs themselves, which are larger national data hubs that aggregate data for larger collections and offer services specialized around at least one type of thematic data. A DSC has different levels of activities, assembles data such as DAC, is located as close as possible to the producer, stores data at least at the national level by theme (by aggregating, for example, thematic data from different DACs), and is a multi-thematic data center. This implies higher storage and computing capacities, as well as dedicated data management and technical teams.

The third category corresponds to data centers offering data analysis and interpretation services (DIAS) on demand. DIAS are large IT facilities with the capacity for fast handling of large and complex data volumes. A DIAS data center will propose working environments to analyze and interpret data, in particular, large data sets that are difficult to download and require large computing capacities. Cross-thematic studies require efficient access to data from multiple sources (*in situ*, satellite) and different compartments. It is then necessary to have large aggregated data sets (data lake) to allow cross-analysis of data from multiple sources (*in situ*, satellite) and from different compartments (e.g., ocean, land, atmosphere, biosphere, solid earth). The objective is to offer virtual research environments (VRE) with dedicated statistical and geostatistical tools, cartography, machine learning, or development environments (e.g., Jupyter Notebooks, Pangeo suite). Such DIAS centers require substantial resources (e.g., informatics, team), which are only conceivable at the level of the Data Terra research infrastructure.

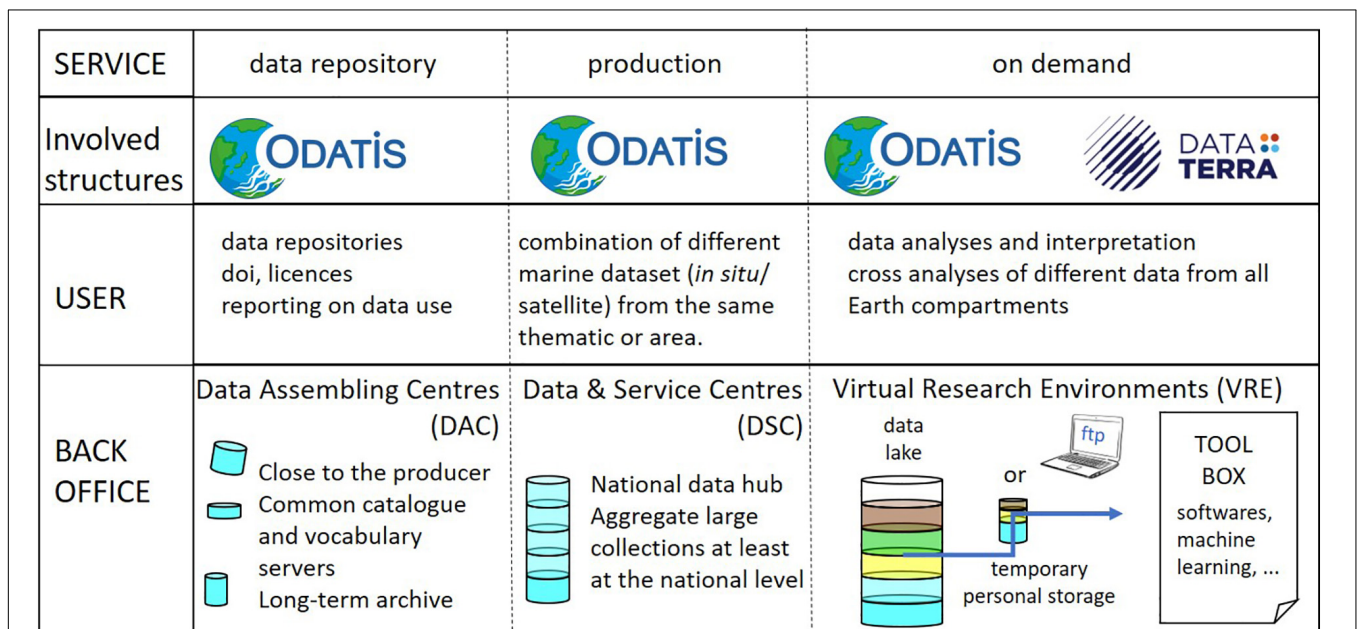
To date, most data available through the ODATIS portal are related to long-term observation networks, which have an automatic data workflow to acquire data directly from

<sup>9</sup>[www.coretrustseal.org/](http://www.coretrustseal.org/)



**TABLE 2** | Tasks identified to implement FAIR and CoreTrustSeal requirements to guide ODATIS data and service centers toward certification.

Structure level	Executive office	Data and services centers
<b>Step 1</b> Tasks	<b>Diagnosis of the degree of implementation of DSCs</b> <ul style="list-style-type: none"> <li>• Produce a document that simply introduces FAIR and certification requirements</li> <li>• Define criteria to evaluate the organization of the DSC</li> <li>• Create a questionnaire for an auto-evaluation of each DSC</li> </ul>	<ul style="list-style-type: none"> <li>• Individual auto-evaluation of each DSC using the questionnaire established by the executive office to assess the degree of implementation of FAIR principles and certification</li> <li>• Identify possible improvements</li> </ul>
<b>Step 2</b> Tasks	<b>Implementation of the approach to volunteer DSCs</b> <ul style="list-style-type: none"> <li>• Production of a roadmap and individual recommendations for each DSC to achieve FAIR principles</li> <li>• Assistance by providing technical support to facilitate the standardization, harmonization, and upgrade of the DSC: <ul style="list-style-type: none"> <li><i>Production of implementation guidelines and templates</i></li> <li><i>Organization of technical workshops</i></li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Application of the roadmap to volunteer DSCs <i>Allocated time: 9 months</i></li> <li>• Feedbacks on the experience to improve the guidelines</li> </ul>
<b>Step 3</b> Tasks	<b>Harmonization of DSCs</b> <ul style="list-style-type: none"> <li>• Transposition of the guidelines for all DSCs</li> </ul>	<ul style="list-style-type: none"> <li>• Individual auto-evaluation of each DSC to identify the required improvements</li> </ul>

**FIGURE 3** | Schematized typologies of data centers, from marine data assembling centers (DAC) and data and services centers (DCS) to the virtual research environment (VRE), in response to user needs. The structure indicates the implication in the management of the data center of ODATIS and its support research infrastructure Data Terra.

sensors (e.g., Argo). It is even mandatory for accredited French observation networks to share data and metadata (Cocquemot et al., 2019). On the contrary, it is still complicated to collect data from a specific project or a cruise because of the diversity of data and acquisition delays. Some data are measured directly on board, but most samples are analyzed back at the laboratory. The need to help data producers share their data has already been mentioned; this task must be performed in addition to data acquisition. To make this task easier, it is possible to access the SEA scieNtific Open data Edition (SEANOE)<sup>10</sup> service through

the ODATIS portal. SEANOE offers the possibility to publish scientific data in the field of marine sciences free of charge. Data are published as open access for a maximum of 2 years, for example, to restrict access to data of an article under review. The producer sets the terms and conditions for the use of the data by selecting one of seven Creative Commons licenses. SEANOE recommends the use of perennial data files (e.g., CSV instead of Excel) and requests a description of the dataset (metadata). If the dataset matches the quality criteria and the theme (marine science), SEANOE provides a DOI within two working days and will monitor future dataset citations in articles, with download statistics reported to authors once a year. Data published in

<sup>10</sup>[www.seanoe.org](http://www.seanoe.org)

SEANOE are also automatically duplicated to the EMODnet Data Ingestion portal and, depending on their interest, may also be published via European data portals such as SeaDataNet or a thematic EMODnet portal.

## CONCLUSION

The aim of this study was to provide information on the structure of ODATIS, the Data Terra cluster dedicated to French open ocean and coastal data, which could serve as a guideline for future national or thematic data repositories. The first challenge for ODATIS is to provide easier and wider access to FAIR marine data. This issue is crucial for addressing ongoing and future climate and environmental changes, particularly in coastal regions directly affected by human activities. Repositories already exist all over the world; some have abundant experience but many do not fully comply with FAIR data principles. A second challenge is to develop data analysis and interpretation services. Here, we have highlighted the different issues to consider when developing a strategy to guide DSCs toward certification and for defining a DSC classification system for optimizing human and IT costs with respect to the services offered to users. The implementation of such e-infrastructure requires significant resources and time to develop the tools necessary for their implementation. Substantial personnel and IT resources are also required. Assistance to data producers is essential because it is at this first stage that the metadata associated with the datasets are defined, ensuring the development of FAIR repositories. ODATIS and its reference infrastructure Data Terra should not only organize the data life cycle based on the classification of DSCs, but should also consider the users. Such access to data from multiple sources (*in situ*, satellite, post-sampling analysis) and from different compartments is fairly new, and most colleagues are not yet accustomed to searching and manipulating such datasets, let alone remotely. In parallel to the informati-

development of the ocean cluster of Data Terra, there is a need for further communication and training. The research community has to be informed of these new tools, which could be achieved via news on the cluster web. The management team is responsible for communication of information about ODATIS to the French community through participation in national meetings of scholarly associations and international conferences. A Tour de France was also initiated in 2019, which consists of visiting major French marine research centers to present ODATIS and encourage exchange with colleagues. Training can be performed through webinars, workshops, and technical assistance. In conclusion, our ambition is to develop marine hubs for French marine data, with the aim of promoting widespread use beyond the scientific community.

## AUTHOR CONTRIBUTIONS

SS, GM, CN, JS, VH, GD, and FH contributed to the conception or design of the infrastructure. SS and GM have written the original draft. All the co-authors commented on and approved the final manuscript.

## FUNDING

ODATIS was supported by funding from CNRS, CNES, Ifremer, IRD, SHOM, Marine Universities, AllEnvi, and MESRI, and by the French Agence Nationale de la Recherche (ANR) through the COPiLOtE project.

## ACKNOWLEDGMENTS

We warmly thank Caroline Mercier, who is mandated by the CNES, for helping with the catalog of the ocean cluster.

## REFERENCES

- Braud, I., Chaffard, V., Coussot, C. H., Galle, S., Juen, P., Alexandre, H., et al. (2020). Building the information system of the french critical zone observatories network: Theia/OZCAR-IS. *Hydrol. Sci. J.* doi: 10.1080/02626667.2020.1764568
- Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., et al. (2018). Declining oxygen in the global ocean and coastal waters. *Science* 359:eaam7240. doi: 10.1126/science.aam7240
- Buck, J. J. H., Bainbridge, S. J., Burger, E. F., Kraberg, A. C., Casari, M., Casey, S., et al. (2019). Ocean data product integration through innovation—the next level of data interoperability. *Front. Mar. Sci.* 6:32. doi: 10.3389/fmars.2019.00032
- Charria, G., Lamouroux, J., and De Mey, P. (2016). Optimizing observational networks combining gliders, moored buoys and FerryBox in the Bay of Biscay and English Channel. *J. Mar. Syst.* 162, 112–125. doi: 10.1016/j.jmarsys.2016.04.003
- Cocquemot, L., Delacourt, C., Paillet, J., Riou, P., Aucan, J., Castelle, B., et al. (2019). Coastal ocean and nearshore observation: a french case study. *Front. Mar. Sci.* 6:324. doi: 10.3389/fmars.2019.00324
- Crutzen, P. J. (2002). Geology of mankind. *Nature* 415:23.
- Future Earth (2020). *Our Future on Earth*. Available online at: <https://futureearth.org/publications/our-future-on-earth/> consulted the (Accessed March 26, 2020).
- GO-FAIR (2020). *FAIR Principles*. Available online at: <https://www.go-fair.org/fair-principles/> (Accessed March 26, 2020).
- Gutmann, M., Witkowski, K., Colyer, C., McFarland, O., Rourke, J., and McNally, J. (2008). Providing spatial data for secondary analysis: issues and current practices relating to confidentiality. *Popul. Res. Policy Rev.* 27, 639–665.
- INSPIRE (2007). *Directive 2007/2/EC The EU's Infrastructure for Spatial Information*. Available online at: <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=LEGISSUM%3A128195>, (Accessed July 7, 2020).
- Le Reste, S., Dutreuil, V., André, X., Thierry, V., Renaut, C., Le Traon, P.-Y., et al. (2016). “Deep-Arvor”: a new profiling float to extend the Argo observations down to 4000m depth. *J. Atmos. Ocean. Technol.* 33, 1039–1055. doi: 10.1175/JTECH-D-15-0214.1
- Levin, L. A., and Le Bris, N. (2015). The deep ocean under climate change. *Science* 350, 766–768. doi: 10.1126/science.aad0126
- Lin, D., Crabtree, J. C., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., et al. (2020). The TRUST principles for digital repositories. *Sci. Data* 7:144. doi: 10.1038/s41597-020-0486-7
- Miguez, B. M., Novellino, A., Vinci, M., Claus, S., Calewaert, J.-B., Vallius, H., et al. (2019). The european marine observation and data network (EMODnet): visions and roles of the gateway to marine data in Europe. *Front. Mar. Sci.* 6:313. doi: 10.3389/fmars.2019.00313

- Nichols, C. R., Wright, L. D., Bainbridge, S. J., Cosby, A., Hénaff, A., Loftis, J. D., et al. (2019). Collaborative science to enhance coastal resilience and adaptation. *Front. Mar. Sci.* 6:404. doi: 10.3389/fmars.2019.00404
- Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., et al. (2014). troubleshooting public data archiving: suggestions to increase participation. *PLoS Biol.* 12:e1001779. doi: 10.1371/journal.pbio.1001
- Rode, M., Wade, A. J., Cohen, M. J., Hensley, R. T., Bowes, M. J., Kirchner, J. W., et al. (2016). Sensors in the stream: the high-frequency wave of the prese. *Environ. Sci. Technol.* 50, 10297–10307. doi: 10.1021/acs.est.6b02155
- Schmidt, S., Andersen, V., Belviso, S., and Marty, J.-C. L. (2002). Dissolved and particulate thorium 234 concentrations at time series station DYFAMED. *PANGAEA*. doi: 10.1594/PANGAEA.737924
- Schmidt, S., Bernard, C., Escalier, J.-M., Etcheber, H., and Lamouroux, M. (2017). Assessing and managing the risks of hypoxia in transitional waters: a case study in the tidal Garonne River (South-West France). *Environ. Sci. Pollut. Res.* 24, 3251–3259. doi: 10.1007/s11356-016-7654-5
- Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., et al. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat. Clim. Chang.* 9, 306–312. doi: 10.1038/s41558-019-0412-1
- Steffen, W., Richardson, K., Rockstrom, J., Cornell, S. E., Fetzer, I., Bennett, E. M., et al. (2015). Planetary boundaries: guiding human development on a changing planet. *Science* 347:1259855. doi: 10.1126/science.1259855
- Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., de Bruin, T., et al. (2019). Ocean FAIR data services. *Front. Mar. Sci.* 6:440. doi: 10.3389/fmars.2019.00440
- Tyler, A. N., Hunter, P. D., Spyarakos, E., Groom, S., Constantinescu, A. M., and Kitchen, J. (2016). Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. *Sci. Total Environ.* 572, 1307–1321. doi: 10.1016/j.scitotenv.2016.01.020
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The reviewer TT declared a past collaboration with one of the authors VH to the handling editor.
- Copyright © 2020 Schmidt, Maudire, Nys, Sudre, Harscoat, Dibarboure and Huynh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.