

**In Research Data Sharing and Valorization - Developments, Tendencies, Models, Chap. 5, pp. 77-96**

2020

Eds Joachim Schöpfel, Violaine Rebouillat

ISBN 9781789450736

Online ISBN 978139416341

<https://doi.org/10.1002/9781394163410.ch5>

<https://archimer.ifremer.fr/doc/00787/89945/>

**Archimer**  
<https://archimer.ifremer.fr>

---

## SEANOE - a thematic repository

Merceur Frédéric <sup>1</sup>, Petit De La Villeon Loic <sup>1</sup>, Van Iseghem Sylvie <sup>1</sup>

<sup>1</sup> Ifremer, France

---

### Abstract :

The French Research Institute for Exploitation of the Sea has developed a repository for marine research data, called SEANOE. It is not an institutional repository, as it is open to all researchers in the field. SEANOE allows marine science datasets to be published easily, quickly and freely. The SEANOE project is managed by a project manager from the Sismar team. The Archimer solution had the advantage of offering a simple repository system, as well as a validation system, and of being very quickly operational. Datasets are accessible from the SEANOE front-office. The metadata are also stored in a structured way in schema.org format in the landing pages, which allows Google to index them in its Dataset Search tool. Reaching a critical mass of repositories, which is difficult to estimate, would ensure that SEANOE has sufficient visibility at the international level, which in turn would lead to new repositories.

**Keywords :** A rchimer solution, dataset search tool, institutional repository, SEANOE

## SCIENCES - Transformation Dynamics of Tools and Practices

*The French Research Institute for Exploitation of the Sea (Institut français de recherche pour l'exploitation de la mer, Ifremer) has developed a repository for marine research data, called SEANOE (Sea Scientific Open Data Edition). SEANOE is published in Ifremer's marine data portal. It is not an institutional repository, as it is open to all researchers in the field. What are the unique characteristics of SEANOE? What is it used for? Is it FAIR? What are the challenges and opportunities for development?*

How would you describe the general mission of the SEANOE repository?

SEANOE<sup>1</sup> allows marine science datasets to be published easily, quickly and freely. It is open to the whole international scientific community in the field of marine research. SEANOE is solely dedicated to data from different fields of marine science, such as physical oceanography, marine geology, marine biology, etc.

Each dataset published by SEANOE has a DOI (Digital Object Identifier used as a resource identification mechanism) in which the author (the producer of the data) is clearly identified. This allows for an accurate, reliable and sustainable citation.

SEANOE thus offers a solution tailored to journals requesting that the data used in an article be available online (e.g., PLoS ONE).

The data published by SEANOE can be used under the conditions of the Creative Commons license selected by the author of the data. An embargo limited to 2 years on a dataset is possible (for current scientific publications).

SEANOE accepts datasets smaller than 100 GB.

What is its administrative context?

SEANOE is managed by the team through the Simer data center<sup>2</sup>.

In 1971, the French government entrusted CNEXO (*Centre National pour l'Exploitation des Océans*) with the mission of representing France at the IOC (Intergovernmental Oceanographic Commission) of UNESCO. At that time, only the management of data from oceanographic campaigns was assured. When Ifremer was created in 1984, the result of a merger between the CNEXO and ISPTM organisations, this mission was passed on to it.

Ifremer then set up the Simer (*Systèmes d'Information Scientifique pour la Mer*) data center to manage the collection, processing and dissemination of these marine data. The scope of the data handled has been progressively extended since 1984. The Simer staff is now a team of 28 people (21

---

<sup>1</sup><https://www.seanoe.org>

<sup>2</sup><https://data.ifremer.fr/SISMER>

permanent and 6 fixed-term contracts). Simer manages and/or participates in a series of national and international thematic databases:

- The *Coriolis* International Physical Oceanography Base<sup>3</sup>;
- The national *Harmonie* database of the Fisheries Information System<sup>4</sup> (*Système d'information halieutique*), which contains fishing data;
- The national *Quadrige* database for environmental data;
- The catalogue of French oceanographic campaigns<sup>5</sup>;
- *CATDS* and *CERSAT* spatial oceanographic databases.

The Simer data center is supported by two other teams in its missions:

- The ISI team (14 permanent employees, 7 fixed-term contracts) manages the development of these databases in a subcontracting context;
- The RIC team (20 permanent staff, 4 temporary staff) maintains the technical infrastructures (e.g., servers) of Ifremer. About 40% of the RIC team's activity is dedicated to the management of the infrastructures of these databases.

In 2019, the total budget for these three teams was €5.3m. The budget is 100% funded by Ifremer.

At the national level, Simer is now a component of the Odatis<sup>6</sup> ocean data center, which federates data management activities and scientific expertise in oceanography at the national level.

How does the *SEANOE* team work?

*SEANOE* (Figure 1) is hosted on Ifremer's central shared servers which are administered by the RIC team (*Oracle*, *ElasticSearch*, *Tomcat*, file system, archiving system). The specific cost for Ifremer of the *SEANOE* project in terms of infrastructure and management of this infrastructure is therefore marginal as it is based on existing services and shared with a large number of services.

---

<sup>3</sup><http://www.coriolis.eu.org/>

<sup>4</sup><http://sih.ifremer.fr/>

<sup>5</sup><https://campagnes.flotteoceanographique.fr/>

<sup>6</sup><https://www.odatis-ocean.fr/>



The screenshot shows the SEANOE web interface. On the left, there are two menu items: 'SEANOE' and 'SEXTANT'. The main content area has a breadcrumb trail: 'Accueil / Déposer / Archiver des données / SEANOE'. Below this is the title 'SEANOE - SEA SCIENTIFIC OPEN DATA EDITION'. The text describes SEANOE as a solution for publishing marine scientific data, created by The French NODC. It allows scientists to publish data in open access, cite it reliably, and integrate it into national efforts for data management. A link 'Publier ses données via SEANOE' is provided. Below the link, it states 'L'Open Science, vers une recherche plus transparente et plus rapide' and 'La diffusion en libre accès des résultats de la recherche financée par des fonds publics est une demande sociétale,'.

*Figure 1: SEANOE web interface*

The *SEANOE* project is managed by a project manager from the Simer team. The evolutionary maintenance of the *SEANOE* code is currently subcontracted to a service company (Altran). In 2015, the adaptation of *Archimer* to implement *SEANOE* took 4 months of work. Since 2015, about one month of work is needed annually to ensure its evolutionary maintenance. The management of the subcontracting is ensured by someone from the ISI team who also ensures provides some support to the users.

The validation of the filings is currently carried out in turn by two people from the Simer team (including the project manager). In case of absence, the validation is ensured by the Simer team's helpdesk. A continuous service is thus ensured

We systematically contact depositors who have initiated a deposit but have not finalized it straight away, to offer them our help. This is often an opportunity to answer questions about versioning, for example.

We also have to help from time to time with uploading data files when the size of the data file exceeds 20-30 GB. In case of difficulty, we suggest that users make the data available on an FTP site of their organization so that we can download and upload them ourselves in *SEANOE* from the Ifremer network.

In total, between half a day and one day per week is required for:

- Project management;
- User support (e.g., answering questions about versioning, etc.);
- Checking and updating the uploaded data;
- Overseeing citations;
- Soliciting new deposits from authors of papers co-authored by Ifremer;
- Management of the outsourcing of computer code.

When was *SEANOE* launched, and what were the requirements?

*SEANOE* was launched in November 2015.

The European and then French regulatory context with the law on the digital republic enacted in 2016 recommends open access dissemination of public research data. In addition, publishers such as PLoS require that the data used in an article be freely accessible online and cited using a DOI; the simultaneous publication of data and an article can strengthen the credibility of the study.

However, back in 2015, the Simer data center was not organized in such a way that it could meet this need:

- Simer managed a set of databases that covered most marine science topics. The data that fed these databases came from automated devices (e.g.: Argo autonomous floats), national monitoring networks (e.g.: Rephy), data from oceanographic campaigns, etc.
- The ingestion of a dataset into one of these databases was often done manually, with a need to transform the format of the data and to qualify them before being able to upload them into one of the databases.
- Some of these databases were not open access (e.g., fishing data).
- The data uploaded onto these databases was anonymized and did not allow a particular author to be credited.

What were the initial specifications?

In the above context, Simer decided to develop a dataset publishing system that offers a simple web-based repository interface, a validation system and the possibility of assigning a DOI to these datasets. The model we had then was the *Pangaea* database developed and maintained in Germany<sup>7</sup>.

Scientists who publish a dataset to cite an article in it often do so in a hurry, sometimes after having been refused by the publisher because the data used was not accessible online (this is a little less true today, when some authors have taken on board this requirement on the part of publishers and are starting to anticipate the publication of their data before submitting an article). The system should therefore allow a DOI to be assigned quickly, ideally in less than 24 hours.

The data had to be freely accessible, after a possible time-limited embargo.

For Simer, the objective of this new service was also to broaden the scope of data collection and to capture new datasets that could feed its thematic databases.

During the validation process, Simer must simply ensure that the deposited dataset is indeed a dataset, that the theme is respected, and that it is sufficiently described. In fact, only two deposits have been refused so far: an article deposited by mistake in *SEANOE* and a dataset insufficiently described and whose author seemed unable to describe it correctly.

During validation, Simer may request additional metadata (e.g., geographic range, author affiliation, etc.) and recommend, where possible, that the data be converted to another format (e.g., Excel vs. csv) or that the data be better described (e.g., adding missing units in data columns).

The size of the accepted datasets should not exceed 100 GB. This limit has been set because, beyond that, it starts to become difficult for users to download the data. It is therefore preferable to select

---

<sup>7</sup>See <https://www.pangaea.de/>

systems backed by computing resources where the data can be used directly without having to download them.

These specifications have not been modified since the launch.

What technical resources were mobilized for *SEANOE*?

To set up this service, two approaches were considered: using the *Sextant*<sup>8</sup> infrastructure, a Geographic Information System (GIS) based on the *GeoNetWork* system, or adapting *Archimer*<sup>9</sup>, Ifremer's institutional archive. The *Archimer* solution had the advantage of offering a simple repository system as well as a validation system and of being very quickly operational. This is the solution that was chosen.

*SEANOE* is therefore a subset of *Archimer*, Ifremer's Institutional Archive. *Archimer* is an internal development of the documentation service based on the central servers of Ifremer (*Oracle*, *ElasticSearch*, *Tomcat*). *Archimer* has been adapted to create *SEANOE*. These are two versions of the same system. The code of *SEANOE* is more than 95% similar to that of *Archimer*. The vast majority of metadata describing a report in *Archimer* and a dataset in *SEANOE* are the same. Only a few specific metadata have been added to describe a dataset (e.g., temporal extent, level of data processing, geographical area, etc.). The file loading form also had to be adapted. Indeed, *SEANOE* accepts datasets up to 100 GB in size. To upload such large volumes of files online, special mechanisms must be implemented to manage uploads that can last several hours without the risk of http "TimeOut" errors.

The data are published under the new domain name *SEANOE.org*. We have taken care to make *SEANOE* anonymous to Ifremer so as not to give the impression that Ifremer wanted to appropriate the published data.

The interface and metadata are only available in English. Even to attract datasets from French teams, we have to present ourselves as an international service.

How do you promote *SEANOE* to researchers? How do you communicate with them?

We have presented *SEANOE* in all Ifremer research units. We have also presented it to several national data collection networks.

We have contacted the publishers PLoS, Elsevier and Nature to list *SEANOE* in their recommended data repositories.

Finally, since 2016, we have been sending a message to the authors of all papers co-authored by authors from the UMRs with which Ifremer is associated to offer them the opportunity to publish the data associated with their papers in *SEANOE*. The immediate positive response to these requests is less than 1%. On the other hand, when these authors are confronted with a request from a publisher that requires the publication of data to accept an article, they sometimes remember the service offered by *SEANOE*.

As we expected, and as confirmed by usage statistics, Google is the main source of access to datasets in *SEANOE*; in 2019, 50% of site visits came from Google. Improving the visibility of a dataset means above all working on its SEO and this is something we are working on in particular.

A good SEO rating cannot be obtained without quality metadata and among all the metadata, the title is particularly important. When an author submits a dataset, we therefore refuse, for example, titles

---

<sup>8</sup><https://sextant.ifremer.fr/>

<sup>9</sup><https://archimer.ifremer.fr/>

that are not explicit. An example: a dataset initially published with the title "*GULF\_IND\_PLOS\_One*" will ultimately be published under the title "*Output from a 1/12-degree Global experiment with the Hybrid Coordinate Ocean Model (HYCOM), forced with NCEP Reanalysis products - Data for the Persian Gulf and Strait of Hormuz*".

Technically, the HTML code of the dataset landing pages is then structured to optimize its visibility. For example, it contains metadata structured in JSON-LD schema.org format.

Lastly, we aim to enrich the network with automatic cross-links between datasets published in *SEANOE* and external resources. Google counts the number of links to a resource in its popularity calculation.

At this stage, a dataset published in *SEANOE* may offer cross-links to:

- The ORCID page of its authors;
- The CV of its Ifremer authors;
- Documents in *Archimer*;
- A batch of images in the Ifremer Ocean Library;
- Campaigns in the catalog of French oceanographic campaigns;
- Samples in the catalog of French oceanographic campaigns;
- Elsevier and Taylor & Francis articles via Scholix.

Finally, the SeaDataNet project and the Odatis data cluster, in which Ifremer is a driving force, have selected *SEANOE* as their DOI attribution service.

We also monitor the citations of datasets in international papers using alerts on the word *SEANOE* and the DOI prefix in Google Scholar and on the websites of several publishers (e.g., Elsevier). When we find an article that cites a dataset, we add its reference to the *SEANOE* record. Its record is then automatically updated with the DataCite DOI/article DOI pair. These pairs are then also pushed by DataCite to the Scholix project, which is used by several publishers.

When the authors of the article are not the authors of the dataset, we report the citation to the authors of the data.

Sometimes the dataset is misquoted. The most common error is the citation of the landing page URL instead of the DOI. We then report the error to the authors of the article so that they correct the citation if it is not too late.

Finally, for some major datasets that benefit from a name that is not too competitive on the Internet (e.g.: Argo, Rephy, etc.) we have also set up an alert on the name of the dataset. If we spot an article that cites a dataset without citing the corresponding DOI, we ask the authors of the article to add the citation of the *SEANOE* DOI. Otherwise, when it is too late, the authors usually promise to respect the citation instruction in their subsequent articles.

What has changed since the launch?

*SEANOE* has evolved relatively little since its launch. From the beginning, we noticed that some metadata fields were missing. For example, in the first repositories, many scientists recorded their acknowledgements at the end of the description field. We therefore added a specific "Acknowledgement" field.

in April 2016, to meet the specific needs of the Argo<sup>10</sup> project, we developed the possibility of managing several versions of a dataset within the same DOI and accessing them in a differentiated way using fragmentation (#). This possibility has since been opened up to all datasets published in *SEANOE*.

For the Argo data, a snapshot of the dataset is frozen and stored monthly. In the initial version, a master DOI was assigned to the Argo dataset and specific DOIs were assigned to each monthly snapshot.

At the request of the scientific direction of the Argo project, a unique DOI has been assigned by *SEANOE* to the Argo data. This unique DOI allows the global dataset or a specific snapshot using the same DOI to be cited. In this respect, each snapshot is uploaded to *SEANOE* which assigns it a URL and a key. For example, the snapshot of 2016-02-08 has been assigned the key 42350.

The citation of the global dataset is done by citing the new DOI without parameters. Example:

- *Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. <http://doi.org/10.17882/42182>*

The citation of a specific snapshot is done by adding its key preceded by the # character to the DOI:

- *Argo (2016). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of February 8<sup>th</sup>, 2016. SEANOE. <http://doi.org/10.17882/42182#42350>*

This ability to assign a unique DOI to an evolving dataset has many advantages, in particular:

- Citation instructions are simpler to give to users of the dataset. They have only one DOI to understand. This is an important point because even simple instructions are difficult to apply: for example, we see that scientists regularly ask for a DOI to cite their data in an article, and, in the end, they cite the URL of the landing page instead of the DOI.
- To improve the visibility of the DOI in search engines (e.g., Google), it is preferable to distribute a single landing page that concentrates a maximum number of citations and therefore links (backlink) rather than multiple landing pages of almost identical content.
- If a specific DOI is assigned to each version, search engines will not necessarily show the most recent version in their results lists. A user may therefore discover an obsolete version of the dataset and not realize that a more recent version exists.

Recently, we have implemented the capability of duplicating published *SEANOE* data in EMODnet Ingestion. EMODnet (European Marine Observation and Data Network) is a network of organisations supported by the EU Integrated Maritime Policy. The EMODnet project manages a set of thematic databases, including:

- *EMODnet Biology* (<https://www.emodnet-biology.eu>)
- *EMODnet Bathymetry* (<https://www.emodnet-bathymetry.eu>)
- *EMODnet Physics* (<https://www.emodnet-physics.eu>)

---

<sup>10</sup><http://www.argo.ucsd.edu/>



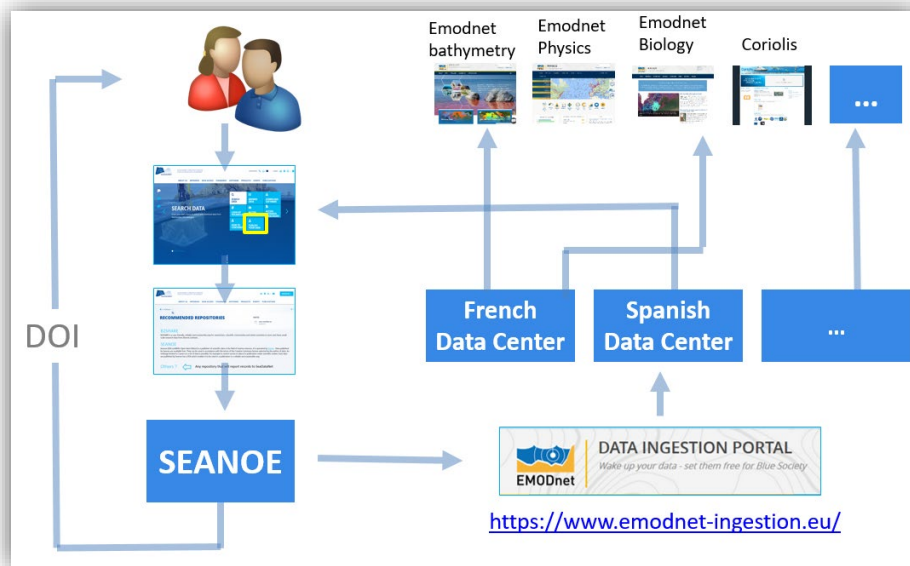


Figure 2: Duplication of SEANOE data in EMODnet

Datasets published in SEANOE can be automatically pushed into EMODnet Ingestion (see Figure 2). EMODnet Ingestion then assigns them to the national data center corresponding to the nationality of the "corresponding author". This national data center must then check whether the data can be ingested into one or other of these portals and, if so, format them and ingest them into the most appropriate thematic database. The Sismer team selects the datasets to be duplicated from SEANOE to EMODnet Ingestion: the datasets are duplicated when the data are qualified, when the selected distribution license allows it and when the dataset is not already exported from an international database.

Furthermore, the link with EMODnet is currently the only way in which SEANOE interconnects with another infrastructure.

What is the number of deposits to date?

As of May 15, 2020, 594 datasets were published in SEANOE. Since the end of 2018, two spontaneous filings are recorded every week on average. In more detail: in 2019, 118 datasets were published by SEANOE, of which 108 were spontaneous deposits by authors and 10 were deposits related to the EGO Gliders project. Spontaneous deposits are overwhelmingly related to article publications. Several publishers (e.g., PLoS ONE, Elsevier, etc.) now require that the data used in an article be made freely accessible online and cited with a DOI. Figure 3 shows the evolution of repositories.

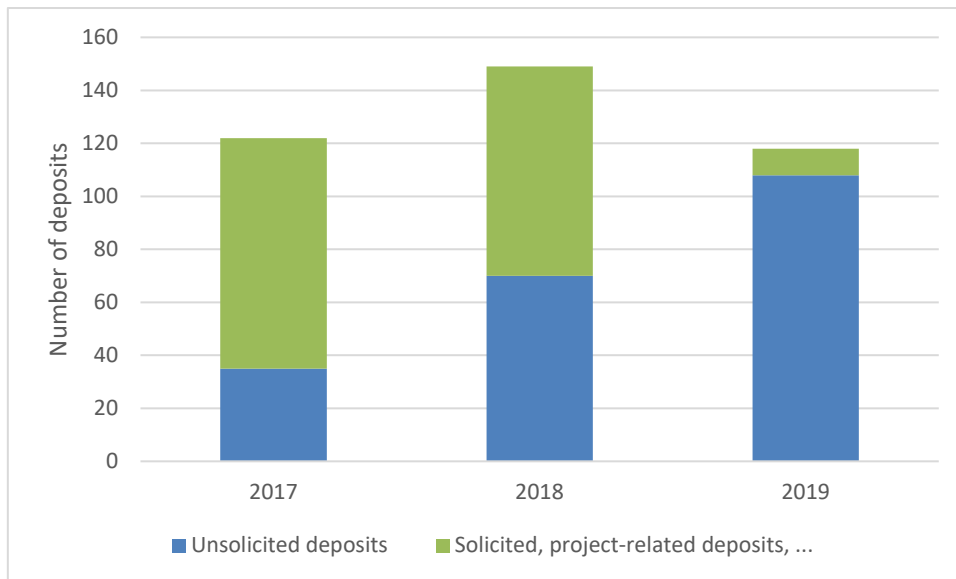


Figure 3: Number of filings in SEANOE 2017-2019

Figure 4 shows the topic of the datasets deposited in 2019. A deposit can be classified in several themes. In this graph, only datasets deposited spontaneously by their authors have been taken into account. The most representative areas are physical, biological and chemical oceanography.

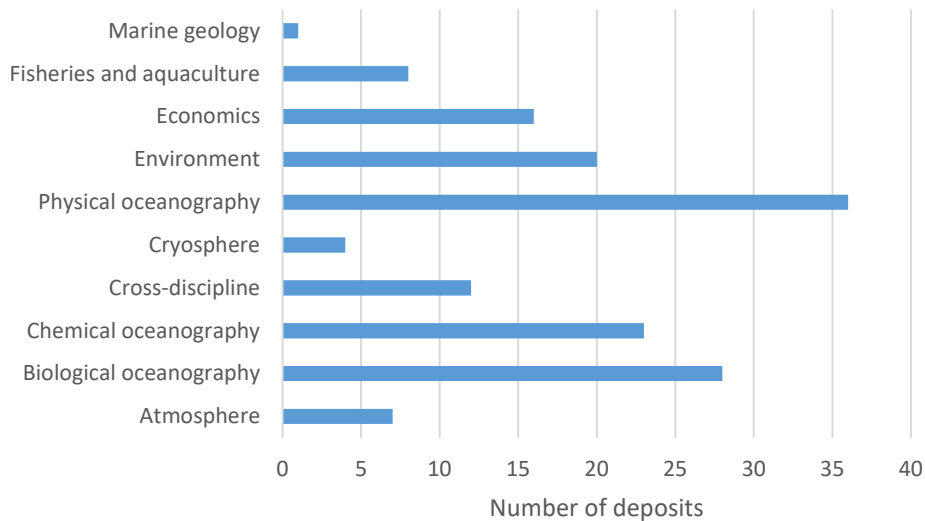


Figure 4: Topics of unsolicited filings in 2019 (108)

The total volume of data files deposited in 2019, if we include version updates (e.g., Argo, Cora, etc.) is 0.56 TB. The median size of the deposits is 10-100 MB. Figure 5 shows the distribution of data volume, DOI by DOI. If a dataset is published as several files, the size of the files is cumulated.

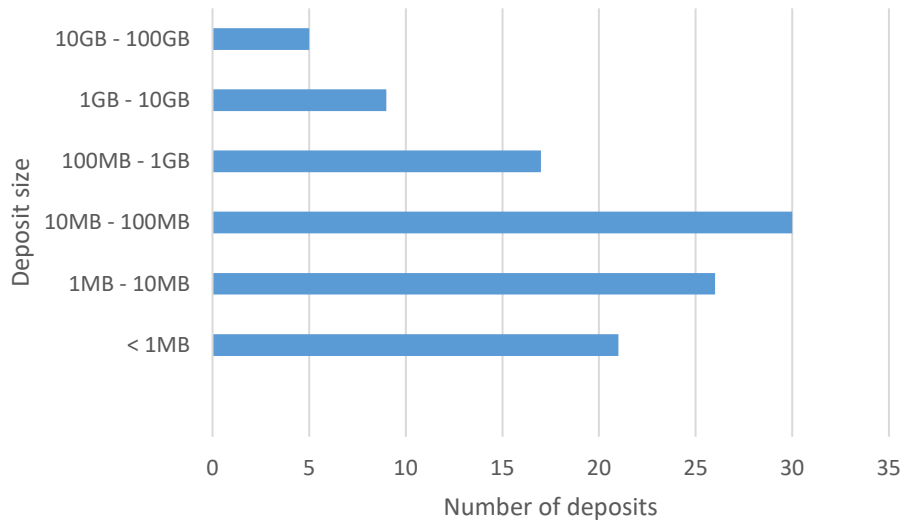


Figure 5: Data file volumes by DOI for 2019 repositories

As for the format of the repositories, NetCDF and CSV are the most commonly used formats. Files that are provided in Excel format are usually transformed into CSV before publication. However, this is not always possible when, for example, the Excel file contains presentation elements necessary to understand the data. PDF files usually contain information describing the data (format, acquisition). Some datasets are composed of a large number of files, sometimes of different formats, which are then zipped.

What are the statistics in terms of connections, user sessions, downloads and so on?

Here is some information taken from an analysis of Ifremer's Apache web server logs. In 2019, the landing pages were accessed 24,000 times from outside Ifremer (excluding robots). The data files were downloaded more than 9,700 times. The progression since 2016 has been significant; between 2017 and 2019, page visits and downloads have more than doubled (Figure 6).

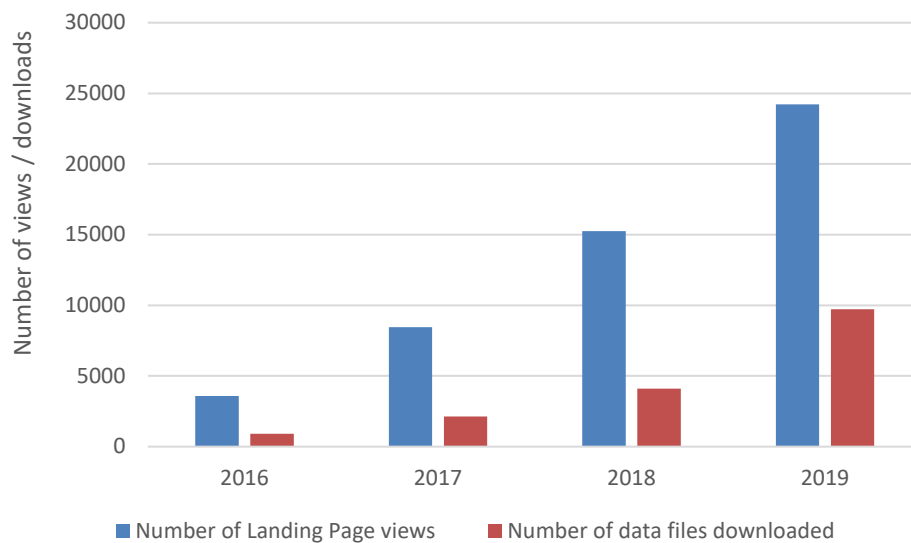


Figure 6: Page views and downloads from outside Ifremer (2016-2019)

The most requested repository was downloaded 1,002 times in 2019 ("Labeled SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode [TenGeoP-SARwv]"); the ten most downloaded datasets accounted for 46% of downloads in 2019. But despite this concentration, 75% of the repositories have been downloaded at least once.

Page views and downloads come from all continents (see Figure 7). The leading countries in terms of usage are China, France, the United States, India and Italy, followed by the United Kingdom and Germany.

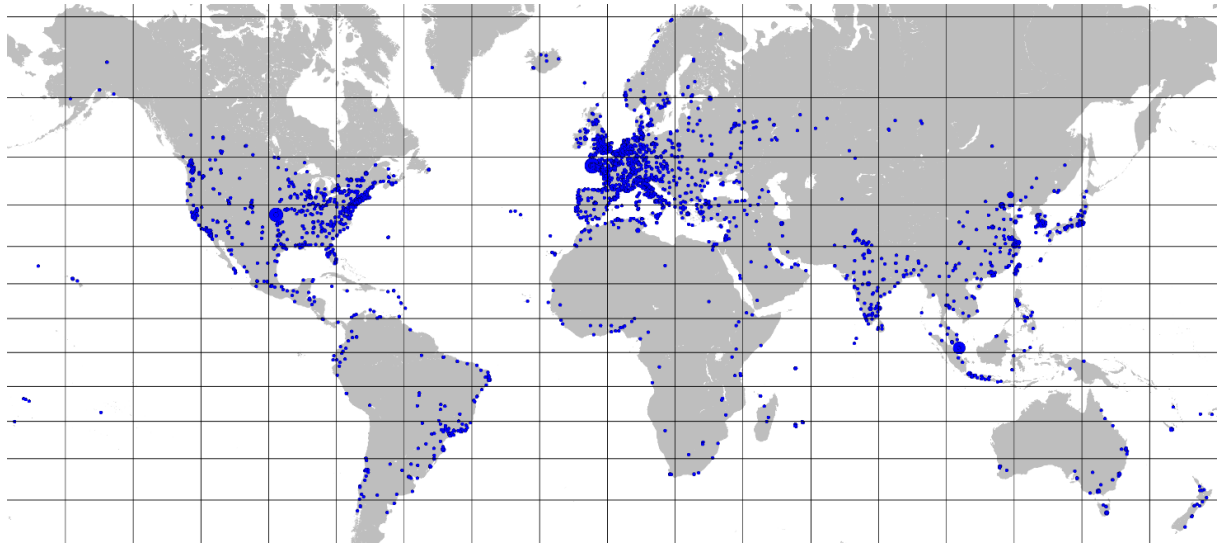


Figure 7: Geographical origin of the landing pages in 2019 (excluding Ifremer)

What do we know about the depositors? What is the feedback from users?

Between January and May 2020, 47 new datasets were published in *SEANOE*. Almost all of these datasets were published by scientists or engineers with a view to citing the dataset in a draft paper. 31 of these 47 datasets were published by French scientists. The other nationalities are varied (Australia, Brazil, Canada, China, Germany, Italy, Japan, Korea, Mexico, Qatar and the USA).

The feedback from scientists who publish their data is generally very positive. In fact, we regularly see the same scientists publishing new datasets. Among the positive points, the help provided during the deposit and the speed of the service come up most often.

What are the strong points of *SEANOE* for you?

*SEANOE* has benefited from the *Archimer* project's experience in publishing. For example, the mechanisms for SEO such as Google had never been implemented before in the systems managed by Sismar. The links with ORCID, the management of authors' affiliations, the links with article DOIs and the provision of download statistics to users are other examples of the contribution of the documentary world to this data publication project.

Conversely, the management of files of several tens of GB, advice on data formats, and links with the EMODnet Ingestion project are all contributions made by the Sismar team in terms of data management.

Among the strengths of *SEANOE*, we would mention:

- Availability and involvement of support during the deposit and after the publication of the dataset;

- The sustainability of the system based on a data center that has existed since 1984;
- Overview of article citations;
- The version control system using fragmentation within the same DOI;
- The SEO mechanisms in Google;
- The richness of the automatic cross-linking network;
- The readability and richness of DOI landing pages.

Does SAENOE consider the FAIR principles?

In general, yes, we strive to comply with the FAIR recommendations in *SEANOE*. Note that the Simer data center, which includes *SEANOE*, is *CoreTrustSeal*<sup>11</sup> certified. Simer has also been accredited as a national oceanographic data center by the IODE program<sup>12</sup>.

Here is a detailed description of *SEANOE*'s compliance with the various FAIR principles.<sup>13</sup>

### **Findability**

*F1. (Meta)data are assigned a globally unique and persistent identifier*

This is the case; each dataset published in *SEANOE* has a DOI.

*F2. Data are described with rich metadata (defined by R1 below)*

*SEANOE* imposes a list of mandatory metadata (title, authors, description, etc.) and proposes a set of optional metadata. We refuse to validate repositories that we consider insufficiently described (e.g.: title not explicit, etc.). But the detail of the description of a dataset differs according to the authors.

*F3. Metadata clearly and explicitly include the identifier of the data they describe*

The landing page of the dataset clearly presents the DOI in both the metadata and the suggested citation. In addition, some authors ask us to reserve a DOI for them before deposit so that they can add it to the data files. This is a practice we should encourage in the future.

*F4. (Meta) data are registered or indexed in a searchable resource*

Datasets are accessible from the *SEANOE* front-office. The metadata are also stored in a structured way in schema.org format in the landing pages which allows Google to index them in its Dataset Search tool. The metadata is also accessible through *SEANOE*'s OAI-PMH engine. The *SEANOE* REST API should be made available on the Internet in a future release (although we haven't had any requests for this yet).

### **Accessibility**

*A1. (Meta)data are retrievable by their identifier using a standardized communications protocol*

Metadata and data are freely accessible without authentication using HTTPS links (with the exception of embargoed data which is accessible on demand).

*A2. Metadata are accessible, even when the data are no longer available*

The *SEANOE* Terms of Use specify that it is forbidden to request the removal of a dataset already published. It is possible to update a dataset but not to remove it. In fact, we sometimes put a dataset

<sup>11</sup><https://www.coretrustseal.org/wp-content/uploads/2019/11/IFREMER-SISMER.pdf>

<sup>12</sup><https://archimer.ifremer.fr/doc/00389/50015/50604.pdf>

<sup>13</sup>See <https://www.go-fair.org/fair-principles/>

back under embargo at the request of an author while they release a correction when an error is detected in their dataset.

### **Interoperability**

*I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation*

The metadata are accessible in JSON LD format from the landing pages and in Dublin Core format from the OAI-PMH engine.

*I2. (Meta)data use vocabularies that follow FAIR principles*

This is something we need to improve. For example, by implementing the ability to list parameters using vocabulary lists defined in the SeaDataNet project.

*I3. (Meta)data include qualified references to other (meta)data*

It is possible to link the dataset to other resources, using:

- ORCIDs from the authors;
- NSCIs of geological samples;
- DOIs of related articles;
- DOIs of associated datasets.

All these *SEANOE* DOI / PID resource pairs are updated in the DataCite record.

On the other hand, we do not disclose the nature of the association with articles and datasets. We only use the `IsAssociatedTo` role. Specifying the nature of the link could be an evolution in a future version.

### **Reusability**

*R1.1. (Meta)data are released with a clear and accessible data usage license*

Datasets in *SEANOE* are published under a CC license.

*R1.2. (Meta)data are associated with detailed provenance*

This information is often available but scattered in the description and in the "Sensor metadata" field. But we do not propose a specific metadata field to record this information.

*R1.3. (Meta)data meet domain-relevant community standards*

The metadata collected by *SEANOE* is compatible with the DataCite and Dublin Core schema. The addition of fields (e.g., parameters) based on SeaDataNet vocabulary lists is a medium-term project.

Is *SEANOE* linked to EOSC or does it plan to be?

*SEANOE* is an EOSC data management service<sup>14</sup>; SeaDataNet uses *SEANOE* to assist researchers in publishing their datasets.

What are the current and future challenges?

In terms of tools, the back-office interface is aging and will need to be redesigned in the medium term. But it is above all the relative weakness of the number of deposits that could be problematic in the long term. The main challenge for *SEANOE* is to attract more deposits. Reaching a critical mass of

---

<sup>14</sup><https://marketplace.eosc-portal.eu/services/seadatanet-doi-minting-service>

repositories, which is difficult to estimate, would ensure that *SEANOE* had sufficient visibility at the international level, which in turn would lead to new repositories.

How will you develop *SEANOE*?

The filing forms in *SEANOE* are aging. They are based on a system that was developed in 2009. The redesign of the *SEANOE* back-office is envisaged from 2021. The first impression given by a data repository back office is important. Indeed, when an author has to publish the data associated with an article, he has at his disposal several hundreds of general or thematic repositories. We sometimes notice scientists passing by and quickly filling in a few fields before abandoning their deposit. We assume that they spend a few tens of seconds on a first selection of repositories before choosing one to publish their data. And when an author has experienced a repository system that suits him, he usually remains faithful to it for his next repositories. For *SEANOE*, the lack of immediate appeal of the current repository system is a drawback.

The development of an indexing system using a list of standardized vocabularies from the SeaDataNet project (e.g., list of measured parameters, country and contact point organisations, etc.) would allow the automation of data transfers between *SEANOE* and EMODnet Ingestion. The description of the data sets using standardized vocabularies is also a FAIR recommendation.

What will *SEANOE* look like in five or ten years?

At this stage, *SEANOE* is correctly fulfilling the role it has been assigned: to provide a simple and fast publishing interface to publish a dataset and obtain a DOI to cite it in an article. In ten years, this need should continue to exist and will probably be even stronger due to the societal demand for open access publication of research data.

In ten years, *SEANOE* should therefore be substantially the same as the current version. We hope that it will offer free access to a significant number of new data sets.

And in ten years, the datasets published by *SEANOE* should be part of a larger ecosystem of interconnected international resources.

Questions asked by Joachim Schöpfel