



# ICTS SOCIB

## Notes Before Completing the Application

*We have read and understood the notes concerning our application submission.*

True

*Reviewer Entry*

**Reviewer 1**

Comments:

**Reviewer 2**

Comments:

Yes

## CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

### Background & General Guidance

### Glossary of Terms

## BACKGROUND INFORMATION

### Context

*R0. Please provide context for your repository.*

*Repository Type. Select all relevant types from:*

Domain or subject-based repository, Research project repository

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:  
okay

### ***Brief Description of Repository***

The Sistema d'Observació i predicció Costaner de les Illes Balears (SOCIB) (<https://www.socib.es/>) hosts the SOCIB Data Repository (<https://socib.es/data>), which is a national marine research infrastructure included in the Spanish Large Scale Infrastructures (ICTS, Infraestructuras Científicas y Técnicas Singulares) Map, updated in November 2018.

The repository contains oceanographic and coastal data, which are collected from various observing platforms (e.g., high frequency radar, gliders, drifters, buoys, etc.) and forecasting services (numerical models). Ocean and coastal communities mainly determine the development and objectives of the repository, along with national or European marine directives.

In addition to SOCIB's observation programs and projects, the repository also includes data from research projects within the Spanish National Research Plan (MOCCA, MUSA, Pre-SWOT, etc.), and from European (Jerico-S3, EuroSea, Euro Argo RISE, etc.) and international research projects (eg. Calypso\_ONR). The metadata included in the files provides specific information about each dataset. The SOCIB repository is used in these projects as an archiving tool for current and future research projects requiring long-term legacy data input.

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:  
Good description

### ***Brief Description of the Repository's Designated Community.***

Ocean and coastal research communities continually share data and information. The SOCIB's repository contains all of its data assets from its observing and forecasting facilities.

Heslop et al. (2019) (<https://www.tandfonline.com/doi/full/10.1080/1755876X.2019.1582129>) identify 10 ocean and coastal sectors and associated communities that use the SOCIB Data Repository: marine and coastal research, maritime

safety, marine sports, beach and coastal communities, coastal protection and planning governance, sustainable marine ecosystems, ports and shipping, ocean management, sustainability of islands and climate change, and education. These users access the repository to query the data rather than modifying it, i.e., they are not depositors. As mentioned, SOCIB's observing and forecasting facilities are the principal depositors.

Thus, the Designated Community is composed as follows:

- Depositors: SOCIB's observing and forecasting facilities
- Consumers:
  - + SOCIB itself (data exploitation, research, etc.)
  - + Ocean and coastal research communities
  - + Other ocean and coastal sectors and associated communities

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:  
breadth of users and communities fully explained

### ***Level of Curation Performed. Select all relevant types from:***

C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation, D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:  
okay

### ***Comments***

The SOCIB data policy ( <https://socib.es/?seccion=dataCenter&facility=accessPolicy> ) explains that the repository constitutes “D. Data-level curation”. The generation of metadata follows interoperable and international standards and can thus facilitate data discovery, while adopting the European Directive INSPIRE (European Commission 2007, <https://inspire.ec.europa.eu/> ). All of the variables in the repository must comply with common policies for naming scientific parameters and data curation. Automatic curation is executed where possible. Aggregated data products are automatically built.

The levels of curation in the datasets are as follows:

- 95% of the data contain enhanced curation (C). All datasets included in the SOCIB repository are derived from raw data (typically but not exclusively ASCII or binary) transmitted in near-real time or after mission/campaign completion. All datasets are in NetCDF format and include rich metadata and links to related documentation (eg., quality user manuals).
- 5% of the data are curated (D) and include scientific post-processing, such as through gliders and research vessel instruments with corrections.

*Reviewer Entry*

**Reviewer 1**

Comments:  
Approved.

**Reviewer 2**

Comments:  
good description of levels

***Insource/Outsource Partners. If applicable, please list them.***

*Reviewer Entry*

**Reviewer 1**

Comments:

**Reviewer 2**

Comments:  
not applicable

***Summary of Significant Changes Since Last Application (if applicable).***

*Reviewer Entry*

**Reviewer 1**

Comments:

**Reviewer 2**

Comments:  
not applicable

***Other Relevant Information.***

SOCIB offers an essential Open Access Data Repository for oceanographic research. The repository contains a wide range of Essential Ocean Variables (EOVs, [https://www.gosocean.org/index.php?option=com\\_content&view=article&layout=edit&id=283&Itemid=441](https://www.gosocean.org/index.php?option=com_content&view=article&layout=edit&id=283&Itemid=441) ) from physical (sea water temperature, sea water salinity, currents, sea surface height, etc.) to biochemical (nutrients, oxygen, turbidity, etc.) to biological (chlorophyll mass concentration, etc.). Meteorological variables (atmospheric pressure, wind velocity, and direction, air temperature, humidity, etc.) are also obtained for some platforms.

These data are either reported by observing facilities (oceanographic buoys, drifters, profilers, gliders, high-frequency radars, etc.) or as outputs from hydrodynamical numerical (WMOP, [https://www.socib.es/?seccion=modelling&facility=forecast\\_system\\_description](https://www.socib.es/?seccion=modelling&facility=forecast_system_description) ), wave (SAPO, <https://www.socib.es/?seccion=modelling&facility=sapo> ), or meteorological (BRIFS, <https://www.socib.es/?seccion=modelling&facility=rissagaforecast> ) models, and facilitate the monitoring and forecasting of the ocean state and variability through projects and observation programmes that focus on the Western Mediterranean Sea.

Protocols and applications following the FAIR principles (Findable; Accessible; Interoperable; Re-usable) for scientific data management have been set up so users can access the SOCIB Data Repository. In addition, the data are also available through the DIGITAL.CSIC repository ( <https://digital.csic.es/handle/10261/228358> ).

These data, either reported by observing facilities (oceanographic buoys, drifters, profilers, gliders, high-frequency radars, etc.) or output by hydrodynamical numerical models (WMOP, [https://www.socib.es/?seccion=modelling&facility=forecast\\_system\\_description](https://www.socib.es/?seccion=modelling&facility=forecast_system_description) ), wave (SAPO, <https://www.socib.es/?seccion=modelling&facility=sapo> ), or meteorological (BRIFS, <https://www.socib.es/?seccion=modelling&facility=rissagaforecast> ), facilitates the monitoring and forecasting of the ocean state and variability in the frame of a number of projects and observation programmes with a particular focus on the Western Mediterranean Sea.

Following the FAIR principles (Findable; Accessible; Interoperable; Re-usable) for scientific data management, a number of protocols and applications have been set up for users to access the SOCIB Data Repository. In addition, the data are also exposed through the DIGITAL.CSIC repository ( <https://digital.csic.es/handle/10261/228358> ).

### *Reviewer Entry*

#### **Reviewer 1**

Comments:  
Accept

#### **Reviewer 2**

Comments:  
Not needed but a good opportunity to expand upon description of the repository if needed.

## **ORGANIZATIONAL INFRASTRUCTURE**

# 1. Mission/Scope

***R1. The repository has an explicit mission to provide access to and preserve data in its domain.***

## ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

### ***Reviewer Entry***

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository  
good description agreed with Level 4

## ***Response:***

The mission of SOCIB ( [https://socib.es/?seccion=historia&id\\_textotextes=historia](https://socib.es/?seccion=historia&id_textotextes=historia) ) is to operate a coastal ocean observing and forecasting system, a scientific and technological infrastructure that provides free, open, quality controlled and timely streams of oceanographic data, in addition to data stewardship and long-term preservation, in order to:

- Support research and innovation development on key internationally established topics such as the role of the ocean in the climate system at an inter-annual scale, the interaction between currents and eddies, vertical exchanges in physical and ecosystem variability, the variability in nearshore morpho-dynamics, and sea-level variability in response to climate change.
- Support (over the longer-term) the strategic needs of society in the context of global change: sustainable management, science-based mitigation, and adaptation strategies, along with policy development and operational tools for decision support.
- Consolidate operational oceanography in the Balearic Islands, in Spain, and Europe, thus contributing to the establishment of a well-structured centre of excellence in an international framework.

The main goal of the SOCIB Data Repository is to ensure that all data is (1) discoverable and accessible; (2) freely available; and (3) interoperable and standardised. The SOCIB Data Products Catalog ( <http://apps.socib.es/data-catalog/> ) is a browser-like interface built on top of the services rooted in the SOCIB API ( <http://api.socib.es/home/> ). It provides an easy and friendly gateway through which regular users can find, download and plot the data available at SOCIB. It is also

an example of how SOCIB API services can be interrogated and exploited, which developers can then reproduce and customise.

The SOCIB Data Management System (DMS) is the set of processes, services and documents aiming to effectively manage the SOCIB data assets lifecycle. The SOCIB Data Management Program is the framework in which the SOCIB DMS is managed, assessed and improved. The SOCIB Data Management Program is implemented based on the Data Management Maturity model ( DMM model, <https://www.cmmi.institute.com/dmm> ), and provides a best-practice roadmap and framework for improving the SOCIB data management function, including data strategy, governance, operations, quality and preservation.

The SOCIB Data Repository is framed in the SOCIB Data Management Program to ensure it complies with international quality assurance and interoperability standards. SOCIB manages its own Digital Object Identifiers (DOI) through a DataCite account ( <https://search.datacite.org/members/csic?data-center-id=csic.socib> ) for datasets and documents.

Long-term preservation is also a key goal for SOCIB, and so various initiatives have been implemented and documented as the following requirements, and particularly Requirement 10 (Preservation plan), indicate.

The management of the High Frequency Radar historical time series illustrates this. The SOCIB High Frequency Radar Facility ( <http://www.socib.eu/?seccion=observingFacilities&facility=radar> ) has led to guidelines on how to sync High Frequency Radar data with the European node (more information is provided here: [http://www.socib.eu/index.php?seccion=detalle\\_noticia&id\\_noticia=413](http://www.socib.eu/index.php?seccion=detalle_noticia&id_noticia=413) ).

The SOCIB mission statement has been endorsed by its funders (Ministry of Science and Balearic Government) through the Board of Trustees approved in 2014 and revised in 2021.

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Accept

##### **Reviewer 2**

Comments:

## **2. Licenses**

***R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.***

***Compliance Level:***

4 – The guideline has been fully implemented in the repository

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Extra supplied text demonstrates compliance level 4

### *Response:*

Most of the data contained in the SOCIB Data Repository are produced internally by SOCIB's Observing facilities and the Modelling and Forecasting facility, as detailed in Requirement 5. The various Agreements and Data Management Plans clearly state that the transfer of custody and any responsibility are handed over to the depositor and the repository from the various projects or observing networks. SOCIB has the right to preserve the data to ensure it remains accessible, usable and understandable over time by the designated community.

SOCIB data derived from permanent observing and forecasting SOCIB systems are “Free, open and quality controlled data streams, with data in adherence to scientific community standards”, following SOCIB founding principles and the original Implementation Plan of 2010 (page 13, [https://socib.es/?seccion=textes&id\\_textotextes=planEstrategico](https://socib.es/?seccion=textes&id_textotextes=planEstrategico) ).

The repository's Terms of Use can be found online ( <https://www.socib.es/data/data-terms-of-use> ) on the landing page. They include use licences, any statement of restriction to the open access policy, and the legal framework and regulations. SOCIB applies the Creative Commons Attribution License (CC BY 4.0) ( <https://creativecommons.org/licenses/by/4.0/> ) to its open data.

The Open Access policy is fully aligned with data policies from leading organisations such as IMOS ( <http://imos.org.au/facilities/aodn/imos-data-management/imos-data-policy/> ), SeaDataNet ( <http://www.seadatanet.org/Data-Access/Data-policy> ), EMODNet ( <https://www.emodnet.eu/> ), and CMEMS ( <https://marine.copernicus.eu/user-corner/service-commitments-and-licence> ), and is reinforced by Spanish and European regulations such as:

- The Water Framework Directive ( [https://ec.europa.eu/environment/water/water-framework/index\\_en.html](https://ec.europa.eu/environment/water/water-framework/index_en.html) ) that grants open and free access to fresh and coastal monitoring data to all citizens.

- The Inspire Directive ( <https://inspire.ec.europa.eu/inspire-directive/2> ), which encourages the creation of metadata (data description and discovery) and grants access to environmental data.

- The Marine Framework Strategy Directive



( [https://ec.europa.eu/environment/marine/eu-coast-and-marine-policy/marine-strategy-framework-directive/index\\_en.htm](https://ec.europa.eu/environment/marine/eu-coast-and-marine-policy/marine-strategy-framework-directive/index_en.htm) ) that covers access to marine environment monitoring data.

The European regulations and national directives also promote open and free access to all research data (“Open Research Data”).

However, in some cases the scientific data managed by the SOCIB repository may be subject to temporary or permanent restricted access. The aims of these restrictions are listed in the Terms of Use.

Access to this “restricted data” will be managed as stated in Requirement 16, either by guest users providing their credentials or through dedicated platforms with authentication protocols.

The current levels of data access can be summarised as follows:

- Moorings ( <https://thredds.socib.es/thredds/catalog/mooring/catalog.html> ): fully open for most platforms but only open by request for some wave recorders
- Drifters ( <https://thredds.socib.es/thredds/catalog/drifter/catalog.html> ): fully open
- Gliders ( <https://thredds.socib.es/thredds/catalog/auv/catalog.html> ): fully open except for the specific international projects (available through request)
- HF-Radar ( [https://thredds.socib.es/thredds/catalog/hf\\_radar/catalog.html](https://thredds.socib.es/thredds/catalog/hf_radar/catalog.html) ): fully open
- Research Vessel  
( [https://thredds.socib.es/thredds/catalog/research\\_vessel/catalog.html](https://thredds.socib.es/thredds/catalog/research_vessel/catalog.html) ): usually fully open except for some specific research projects (available through request)
- Operational models of three different types:
  - + hydrodynamical WMOP (surface and 3D)  
(  
[https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/hydrodynamics/wmop\\_surface/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/hydrodynamics/wmop_surface/catalog.html)  
and [https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/hydrodynamics/wmop\\_3d/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/hydrodynamics/wmop_3d/catalog.html)  
) : fully open
  - + waves SAPO ( Islas Baleares:  
[https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/wave/sapo\\_ib/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/wave/sapo_ib/catalog.html)  
Pitiuses:  
[https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/wave/sapo\\_pitiuses/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/wave/sapo_pitiuses/catalog.html)  
Mallorca: [https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/wave/sapo\\_mallorca/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/wave/sapo_mallorca/catalog.html)  
and Minorca:  
[https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/wave/sapo\\_menorca/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/wave/sapo_menorca/catalog.html) ) : fully open

Details about the preservation of the data included in the SOCIB repository, and therefore subject to free access, are

provided in Requirement 10.

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Accept

##### **Reviewer 2**

Comments:  
New pages off web page most useful in demonstrating compliance at level 4

### **3. Continuity of access**

*R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.*

#### ***Compliance Level:***

3 – The repository is in the implementation phase

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
3 – The repository is in the implementation phase  
Agreed since no formal written agreement for handover in case of service discontinuity.

##### **Reviewer 2**

Comments:  
3 – The repository is in the implementation phase  
Agreed, some planned activities underway currently and yet to be delivered but underway.

#### ***Response:***

SOCIB is the main and only depositor, and also the only custodian of the SOCIB Data Repository. SOCIB ensures the continuity of access and preservation in the data repository currently and in the future.

The data is published in the NetCDF format ( <https://www.unidata.ucar.edu/software/netcdf/> ), which is a well-established, de-facto standard in the scientific community and has long-term support. Metadata generation follows interoperable and international standards to facilitate data discovery, while adopting the European Directive INSPIRE ( European Commission 2007, <https://inspire.ec.europa.eu/> ). In addition, CF vocabularies (standard names, feature type) ( <http://cfconventions.org/> ) are part of the metadata offered by the repository. Some data sets have a Digital Object Identifier (DOI) managed by DataCite ( <https://datacite.org/> ).

The maintenance and evolution of the data in the repository is ensured by following these standards and best practices.

The Data Centre Facility and the Computing and IT Service team (more information in Requirement 5) constantly improve their services to meet the requirements of long-term access and data availability in the repository. The latest developments and technological advancements are closely monitored to ensure long-term preservation by staying informed of the best practices. The funding for the two departments in charge of the repository is sufficient to maintain core operations and activities such as data archiving and dissemination, and providing an information service.

In case of potential damage, a remote backup is stored in the secondary Data Centre (see Requirement 16) that can provide continuity of access to the data. The backup will then be restored to the primary Data Centre as soon as possible.

The SOCIB Strategic Plan for the period 2021-2024 (see R3-D1) separates the continuity business plan and the disaster recovery plan into two phases:

- Phase I (approximate start date Dec. 2021): Deploy a secondary backup environment over a cloud infrastructure, to ensure data availability.
- Phase II (approximate start date Nov. 2022): Deploy a fully replicated on-cloud datacenter infrastructure.

Once Phase II is complete, the SOCIB Data Centre will be able to resume service quickly in the cloud infrastructure, while recovering any data lost due to disruption in the main Data Centre infrastructure.

Several data collections managed by the SOCIB Data Repository have been ingested (replicated) by other international-level repositories, while their master copies remain in SOCIB. This will enable users to access the same data sets if SOCIB Data Repository discontinues its service, on any definite or indefinite timescale. These portals include:

- CMEMS ( <http://www.marineinsitu.eu/partners/socib/> ): In Situ data from SOCIB platforms
- EMODnet Physics ( <https://portal.emodnet-physics.eu/> ): In situ data for SOCIB platforms
- EMODnet Chemistry (SeaDataNet) ( <https://www.emodnet-chemistry.eu/> ): Analysis of laboratory sample data obtained from scientific cruises.
- EGO ( <http://www.red3m.eu/ego-network> ): Glider Near Real Time data
- Coriolis Argo GDAC ( <http://www.coriolis.eu.org/Observing-the-Ocean/ARGO> ): Profilers floats data in the Argo program ( <http://www.argo.ucsd.edu/> )
- ERDDAP NOAA ( <http://osmc.noaa.gov/erddap/index.html> ): Surface drifters data (SVP) through the Global Drifter Program ( <https://www.aoml.noaa.gov/phod/gdp/> )
- European HF-Radar node ( [http://150.145.136.27:8080/thredds/HF\\_RADAR/HFR\\_Ibiza/HFR\\_Ibiza\\_catalog.html](http://150.145.136.27:8080/thredds/HF_RADAR/HFR_Ibiza/HFR_Ibiza_catalog.html) ) : data from the Ibiza radar network.

As Requirement 6 indicates, SOCIB is a member of some of the institutions that manage these repositories.

The Spanish Council for Scientific Research (CSIC: <https://www.csic.es/> ) entered the SOCIB Consortium partnership in December 2020. This will be a major step forward that will guarantee the sustainability of both SOCIB and SOCIB data holdings.

In the unlikely case of a SOCIB closure, Spanish regulations deem that all information and scientific data will be transferred to the most appropriate partner institution, accompanied by any economic obligations (e.g., DIGITAL.CSIC, the institutional repository of the Spanish National Research Council, CSIC, <https://www.re3data.org/repository/r3d100011076> ).

As there is no formal agreement between the repository and the successor organisation, the compliance level is 3.

Linked documents:

R3-D1-SOCIB\_Strategic\_Plan\_2021\_2024 (

<https://repository.socib.es/repository/entry/show?entryid=328e6a1d-7630-4672-a086-0be6ebae5529> )

*Reviewer Entry*

**Reviewer 1**

Comments:

Approved.

**Reviewer 2**

Comments:

## 4. Confidentiality/Ethics

*R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.*

***Compliance Level:***

4 – The guideline has been fully implemented in the repository

*Reviewer Entry*

**Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

## **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Agreed

## ***Response:***

The Ethics and Confidentiality section of the Data Management web page ( <https://www.socib.es/data/data-management> ) explains the SOCIB policy in this regard.

All the data stored in the SOCIB repository is focused on oceanographic and coastal science and follows international Marine Research Infrastructure ( [https://ec.europa.eu/info/research-and-innovation/research-area/oceans-and-seas/marine-research-infrastructures\\_en](https://ec.europa.eu/info/research-and-innovation/research-area/oceans-and-seas/marine-research-infrastructures_en) ) norms and common practices (e.g., the profiling floats data included in the SOCIB repository are part of the Euro Argo ERIC of the European Research Infrastructure Consortium).

Since the beginning of 2020, the ethical dimension of all SOCIB activities has been reinforced through the new Responsible Research and Innovation Service, in which open access and ethics are among the six key dimensions (Public Engagement, Gender Equality, Science Education, Open Access, Ethics, Governance), and SOCIB also actively contributes to the Responsible Research and Innovation in Ocean Observing and Forecasting Factsheet from the EuroSea project ( [https://www.eurosea.eu/wp-content/uploads/2020/04/EuroSea\\_Factsheet\\_RRI.pdf](https://www.eurosea.eu/wp-content/uploads/2020/04/EuroSea_Factsheet_RRI.pdf) ).

The SOCIB Data Repository follows the ethical key points related to ocean observing, as Barbier et al., 2018 ( [https://www.researchgate.net/publication/329128255\\_Ethical\\_recommendations\\_for\\_ocean\\_observation](https://www.researchgate.net/publication/329128255_Ethical_recommendations_for_ocean_observation) ) and McMahon, Clive R. et al., 2021 ( <https://www.frontiersin.org/articles/10.3389/fmars.2021.751840/full#h8> ) note.

The datasets are open access and are disseminated to portals such as CMEMS ( <http://www.marineinsitu.eu/> ), EMODnet ( <https://www.emodnet.eu/en> ) and JCOMMOPS ( <https://www.ocean-ops.org/board> ). To optimise integration in these portals, the data in the repository are continuously adapted in terms of the criteria of interoperability, data quality, standardisation and best practices, where available.

As mentioned in Requirement 2, the SOCIB Data Repository follows an open data policy, but this can be limited in terms of confidentiality or special treatment of the data by the Principal Investigators. The SOCIB will then follow the Ethics in Research Criteria at CSIC ( <https://www.csic.es/en/csic/scientific-integrity-and-ethics-csic> ), which focuses on the moral perspective of research activity. This considers ethical aspects, characteristics and objectives (respect for the dignity of people, their autonomy of will in terms of data protection and privacy and confidentiality, the well-being of animals and environmental preservation).

Some datasets are not published or require authorization for access. The reasons for restricting the access to a given dataset are given in Requirement 2.

In terms of protected marine species, SOCIB has been tracking sea turtles through the use of tags since 2009. The registered trajectory in the repository is public but not the data itself. This project constitutes a consortium of research institutions, conservation organisations and public agencies (Oceanographic Turtles, <https://www.socib.es/?seccion=siasDivision&facility=oceanographic-turtles> ). Importantly, such programmes have clear biologically oriented objectives and consider welfare and ethical issues ( March et al., 2019, <https://doi.org/10.1111/gcb.14902> ).

There are plans to apply anonymization or aggregation to some datasets, but currently these datasets are not published (animal tracking) or not yet included in the repository (ship data from the Automatic Identification System (AIS) environmental repository). The AIS database contains vessels' names and unique vessel identifiers (e.g., MMSI or call sign number). This information has the potential risk of being linked to external databases with personal data (e.g., ship owners). In 2012 the European Data Protection Supervisor issued an opinion on the use of AIS and VMS data (European Data Protection Supervisor, Opinion of the European Data Protection Supervisor on the Commission Implementing Regulation (EU) No 404/2011, which provides detailed rules for the implementation of Council Regulation (EC) No 1224/2009. This establishes a Community control system for ensuring compliance with the rules of the common fisheries policy. 2012/C 37/01 2012.).

Coastal systems are considered environmentally sensitive areas. Some Beach Monitoring Facility ( <https://www.socib.eu/?seccion=observingFacilities&facility=beachMonitoring> ) products are public but a formal request through emailing ([info@socib.es](mailto:info@socib.es)) is required.

SOCIB Data Centre staff are trained in the management of data with disclosure risk, for example attending courses organised by the Ocean Teacher Global Academy.

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:

## **5. Organizational infrastructure**

***R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.***

## ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

### ***Reviewer Entry***

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

## ***Response:***

The repository is hosted by Sistema d'Observació i predicció Costaner de les Illes Balears (SOCIB), which since 2014 has been a national marine research infrastructure included in the Spanish Large Scale Infrastructures Map (ICTS, Infraestructuras Científicas y Técnicas Singulares). An international evaluation of SOCIB activities is conducted every four years, to maintain its inclusion in the ICTS map. It was renewed in 2018.

SOCIB is a public consortium with indefinite duration, and its structural funds can be expressed in terms of 35 FTE numbers. The total annual funding for SOCIB operations is up to two million Euros. The repository has therefore sufficient funding for staff resources, IT resources, and for attending meetings when necessary, over the next 3-5 years.

Additionally, SOCIB also has several partnership agreements

( [https://www.socib.es/?seccion=textes&id\\_textotextes=partners](https://www.socib.es/?seccion=textes&id_textotextes=partners) ) for international, national or regional research projects, and thus has additional project-based funding.

SOCIB organisation:

SOCIB is composed of three divisions plus three horizontal services

( [https://socib.es/?seccion=textes&id\\_textotextes=organizacion](https://socib.es/?seccion=textes&id_textotextes=organizacion) ) that provide central support to the divisions in achieving their objectives:

- The Systems Operation and Support Division (SOS Division) is responsible for establishing the ocean observing and forecasting network. The SOS Division is composed of three main subsystems (facilities):

- + an observing sub-system (Observing Facilities)
- + a forecasting and modelling sub-system (Forecasting and Modelling Facility)
- + a data management sub-system (Data Centre Facility)

The Engineering and Technology Development Division (ETD Division) provides the engineering and technical basis for

developing and operating the facilities of the SOS Division.

- The Strategic Issues and Applications for Society (SIAS Division) is designed to develop applications and operational tools for the science-based management of the coastal and marine environment. This division thus supports the development and transfer of strategic knowledge to meet the needs of society in the context of global change.

The three horizontal services are essential elements that support SOCIB activities:

- Management & Finance
- Computing & IT
- Outreach, Education, Training & Mobility (OETM)

Repository management:

As mentioned in Requirement 0 above, the data ingested by the SOCIB Data Repository comes from the SOS Division. The Observing Facilities ( <https://socib.es/?seccion=observingFacilities> ) and the Forecasting and Modelling Facility ( <https://socib.es/?seccion=modelling> ) produce raw data in many different file formats. The Data Centre Facility ( <https://www.socib.es/?seccion=dataCenter> ) then processes these data and transforms them into NetCDF files (in most cases) and applies quality control checks, resampling, and any other necessary operations. The resulting processed files are typically introduced into the SOCIB Data Repository automatically using appropriate processing software, or in some cases manually by Data Centre staff with the proper credentials.

The Data Centre Facility has developed a general data management system that is continually being improved to guarantee international standards, quality assurance and interoperability. The combination of sources and types of information requires appropriate methods for ingesting, cataloguing, displaying, distributing and preserving this information.

The Computing & IT Service team is responsible for providing the major computing infrastructure to support SOCIB Observing Facilities, Data Centre, Modelling and Forecasting, and general operations. This team also ensures that the SOCIB Data Repository, as a key asset, is protected and that SOCIB complies fully with the requirements of Data Protection and Data Storage regulations. All technical processes, such as security changes, hardware upgrades and the service desk, are managed and operated by Computing & IT Service employees.

Thus, the SOCIB Data Repository is mainly managed by two teams: the Data Centre Facility and the Computing & IT Service, which each have their own duties and staff. They jointly provide the IT infrastructure and computing services to acquire, maintain, access, distribute and preserve data of interest and develop visualisation tools. Both teams are ensured long-term sustainability and stability. The specific achievements of each team are described in the SOCIB Strategic Plan 2021-2024 (see R3-D1).

Staff involved in the Data Centre Facility are as follows:



- Permanent Staff:

+ Data Centre Facility Manager

+ 3 Software Engineers

+ 1 Web Engineer

+ Data Manager, Quality Control and Oceanographic Data

+ Data Steward and Oceanographic Data Specialist

- Non-permanent staff:

+ Web Engineer and Oceanographic Data Specialist

+ 1 Software Engineer

+ 1 Software Engineer/Data Analyst

Staff involved in the Computing & IT Service are:

- Permanent Staff: Computing & IT Service Manager

- Non-permanent and partial-time staff shared with the Data Centre: Computing & IT Technician

Staff training:

The SOCIB training program has two general objectives:

- To improve the performance of tasks in the daily work.

- To facilitate the professional development and career progression of SOCIB staff.

As a general rule, SOCIB is taking into account all the training proposals made by the staff. Particularly, SOCIB is providing training regarding languages, project management, hardware and software tools, occupational health and safety, and technical instrumentation.

The academic backgrounds of the Data Centre staff are deliberately diverse. They include oceanographers, mathematicians and software engineers, but all have mixed experience. The oceanographers have software development knowledge and the software engineers have acquired knowledge of oceanography. Together, they have the knowledge necessary to fulfil the Designated Community requirements.

The Data Centre staff also receive training ( [https://www.iode.org/index.php?option=com\\_content&view=article&id=546:1-1-september-2017-otga-qmf-iode-quality-management-framework-training-course-11-14-september-2017-oostende-belgium&catid=23&Itemid=115](https://www.iode.org/index.php?option=com_content&view=article&id=546:1-1-september-2017-otga-qmf-iode-quality-management-framework-training-course-11-14-september-2017-oostende-belgium&catid=23&Itemid=115) ) in how to implement a Quality Management System (QMS) for the SOCIB Data Management System, which is directly related to the repository management. This led to the creation of the SOCIB Data Management Program.

Attached documents:

R3-D1-SOCIB\_Strategic\_Plan\_2021\_2024 (

<https://repository.socib.es/repository/entry/show?entryid=328e6a1d-7630-4672-a086-0be6ebae5529> )

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:

Approved.

##### **Reviewer 2**

Comments:

## **6. Expert guidance**

*R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).*

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Agreed. Explanation provided on how expert guidance is fed back to the development, priorities and work plans.

### ***Response:***

The SOCIB Data Repository was developed by a team of data scientists and technicians with more than 20 years of international experience. The major principles and objectives of the repository are described in the 2010 Implementation Plan (see R6-D1), which adopts the European Directive Infrastructure for Spatial Information in the European Community (INSPIRE, European Commission 2007) and is aligned with international ocean data initiatives such as the Integrated Marine Observing System (IMOS, <http://imos.org.au/>) in Australia and the Integrated Ocean Observing System (IOOS, <https://ioos.noaa.gov/>) in the US. In Europe, SOCIB has evolved following major EU oceanographic data initiatives such as SeaDataNet (<https://www.seadatanet.org/>), Copernicus Marine In Situ TAC (Copernicus Marine Environment Monitoring Service In Situ Thematic Centre, <http://www.marineinsitu.eu/>) and EMODnet (European Marine Observation

and Data Network, <https://www.emodnet.eu/> ).

In November 2017, the SOCIB International Steering Committee ( [http://www.socib.eu/index.php?seccion=detalle\\_noticia&id\\_noticia=365](http://www.socib.eu/index.php?seccion=detalle_noticia&id_noticia=365) ) reviewed the advances made by SOCIB during the previous Strategic Plan in 2013, and the main scientific, technological and social achievements. Recommendations on future key actions and international priorities were given by the Committee. A further revision from the International Steering Committee was implemented in December 2020.

SOCIB has been an Associate Data Unit ( [https://www.iode.org/index.php?option=com\\_content&view=article&id=373&Itemid=100089](https://www.iode.org/index.php?option=com_content&view=article&id=373&Itemid=100089) ) of the IOC/IODE framework (Intergovernmental Oceanographic Commission/International Oceanographic Data and Information Exchange, [https://www.iode.org/index.php?option=com\\_content&view=article&id=3&Itemid=33](https://www.iode.org/index.php?option=com_content&view=article&id=3&Itemid=33) ) since 2018, which provides guidelines to the SOCIB Data Management Program that are implicit in the SOCIB repository. SOCIB is also an active member of major European Union data programs, such as the European component of the Global Ocean Observing System (EuroGOOS), EuroArgo (the European contribution to the Argo programme), CMEMS In Situ TAC and EMODnet. Together, these international partners play the role of an external advisory committee.

The repository communicates with experts through regular dedicated meetings that take place in the context of international research projects, such as the Ocean Data Interoperability Platform (ODIP2 , <http://www.odip.eu/> ), Jerico-S3 ( <https://www.jerico-ri.eu/> ) or EuroSea ( <https://www.eurosea.eu/> ), and international conferences and workshops such as the International Conference on Marine Data and Information Systems (IMDIS, <https://imdis.seadatanet.org/> ) and the European Geosciences Union (EGU, <https://www.egu.eu/> ).

Communication between the repository and its designated community (ocean data users) is maintained through periodical meetings with partners such as Puertos del Estado ( <http://www.puertos.es/es-es> ), Dirección General de Emergencias ( <https://www.caib.es/govern/organigrama/area.do?lang=es&coduo=2464> ), ICMAN.CSIC ( <http://www.icman.csic.es/es/> ), and news on the web and through social media channels (Twitter: [https://twitter.com/socib\\_icts](https://twitter.com/socib_icts) ).

The robust formal collaboration between SOCIB and the Spanish Maritime Safety and Rescue Agency (SASEMAR, <https://www.centrojoellanos.es/home> ) illustrates the communication and feedback that can occur with the Designated Community through regular meetings. These are convened to assess the results achieved and propose new objectives, through which SOCIB data and knowledge can help optimise SASEMAR operations. The meetings have agendas, minutes, summary key actions, teams and follow-up actions, such as the 3rd Steering Committee meeting between SOCIB-SASEMAR ( [http://www.socib.eu/index.php?seccion=detalle\\_noticia&id\\_noticia=432](http://www.socib.eu/index.php?seccion=detalle_noticia&id_noticia=432) ).

In terms of social media channels, new real-time data available in the repository is communicated to the community. For example, research vessel cruise and lagrangian platform data ( [https://twitter.com/socib\\_icts/status/1325834202584068096](https://twitter.com/socib_icts/status/1325834202584068096) ), or glider data ( [https://twitter.com/socib\\_icts/status/1295635490666876929](https://twitter.com/socib_icts/status/1295635490666876929) ) can be communicated, along with announcements of training programmes that may be helpful to SOCIB repository users, such as webinars concerning in-situ observations of the global ocean, or the CMEMS\_EU In Situ TAC ( [https://twitter.com/socib\\_icts/status/1308683448240963584](https://twitter.com/socib_icts/status/1308683448240963584) ).

The feedback received from these various groups, experts and users is considered by the heads of the SOCIB facilities and the management team (director and subdirector) when planning the work over the following years, and to develop a strategic plan every four years ( [https://socib.es/?seccion=textes&id\\_textotextes=planEstrategico](https://socib.es/?seccion=textes&id_textotextes=planEstrategico) ).

This feedback is a core element of SOCIB. Heslop et al., 2019 (see R6-D2) describes how the data in the repository responds to social requests in terms of data and visualisation. SOCIB applications for modern web browsers and mobile platforms ( <https://apps.socib.es/> ) include the Lifeguards app ( <http://seaboard.socib.es/lifeguard> ) and Jellyfish Observations ( <https://github.com/socib/grumers> ). To create these, SOCIB develops a minimum viable product (MVP, [https://en.wikipedia.org/wiki/Minimum\\_viable\\_product](https://en.wikipedia.org/wiki/Minimum_viable_product) ) and the product manager and relevant developers then organise a session with a cohort of potential users to test it. The team then collects feedback from the users manually and through forms (to extract KPIs). Finally, the information obtained is used in the next steps of the planning and development.

Linked documents:

R6-D1\_SOCIB\_Implementation\_Plan (

[https://repository.socib.es/repository/entry/get/plan\\_socib\\_web.pdf?entryid=4cec9eba-34c4-494a-8bde-4e8caf926060](https://repository.socib.es/repository/entry/get/plan_socib_web.pdf?entryid=4cec9eba-34c4-494a-8bde-4e8caf926060) )

R6-D2\_SOCIB\_integrated multi-platform ocean observing and forecasting

(<https://www.tandfonline.com/doi/full/10.1080/1755876X.2019.1582129> )

*Reviewer Entry*

**Reviewer 1**

Comments:

Approved.

**Reviewer 2**

Comments:

## DIGITAL OBJECT MANAGEMENT

### 7. Data integrity and authenticity

*R7. The repository guarantees the integrity and authenticity of the data.*

***Compliance Level:***

4 – The guideline has been fully implemented in the repository

*Reviewer Entry*

## **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

## **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Additional information provided to reviewers most help in making this assessment

## ***Response:***

NetCDF files (Network Common Data Form, <https://www.unidata.ucar.edu/software/netcdf/>) are the core digital objects within the SOCIB repository, and are stored in the unified SOCIB storage system. Several conventions are combined for this, such as the NetCDF Climate and Forecast convention (see R7-D1), OceanSITES 1.3 (R7-D2) and IMOS NetCDF (see R7-D3) data format conventions (more information is given in the SOCIB NetCDF user manuals and documentation: <https://repository.socib.es/repository/entry/show?entryid=e998fa9b-caf4-4d30-add4-d96b22214ded>).

The data managed in the SOCIB repository is derived from its own observation platforms (e.g., high frequency radar, gliders, drifters, moorings) and numerical models, e.g., the Western Mediterranean OPERational forecasting system, (WMOP [https://thredds.socib.es/thredds/catalog/operational\\_models/oceanographical/hydrodynamics/wmop/catalog.html](https://thredds.socib.es/thredds/catalog/operational_models/oceanographical/hydrodynamics/wmop/catalog.html)). The Data Centre also manages data from external providers.

The data lifecycle processes of the products in the repository are documented in the Data Management Plans (DMP), including metadata, data to be collected and managed, data format and standards, acquisition, assembly, processing, archival, preservation and dissemination, workflow, identity of the depositors and roles (see Requirement 9 for the Data Management Program). An example of the DMP of the high frequency radar can be found in R7-D10.

The web application Instrumentation (controlled by the Data Centre) is used to manage the platforms, instruments, sensors and metadata that link the processing configurations and the data products in the repository. More information can be found in R7-D4.

This information is held in a relational database, called “management”. When the raw data is processed to be transformed into a NetCDF file, the processing toolboxes query this database and add the metadata (global or variable attributes).

Only granted processes can perform updates within the files, in terms of machine users and groups. A permission policy document is available on request. For near real-time data, only the processing toolboxes update the data files automatically (typically every 10 minutes). Delayed time processing, delayed mode (i.e., scientific correction) and reprocessing operations are performed by Data Centre technicians in manual mode as per the Principal Investigator’s requests.

The NetCDF files included in the repository have two global attributes: “platform\_code” and “instrument\_code,” which are stored in an internal “management” database. The name of the file contains the name of the deployment in which the data were collected.

Changes in the data and metadata are noted at the NetCDF file level and the data product level:

- The global attribute “history” is added in the files, thus providing an audit trail for modifications to the original data, through a timestamp indicating the date and time when the program was executed. If an existing NetCDF file is generated again by a newer version of the processing application, the new version and timestamp will be appended to the current “history” value. This can occur either if it has been reprocessed or because new data has been added. In addition, corrections, data cleaning operations and delayed mode curation of the data also imply appending a new value to the corresponding “history” global attribute.
- At the data product level, each new version has an associated changelog, and as data products are collections of NetCDF files, this includes summarised information of all changes. This file is automatically generated based on the list of changes registered in the database.

All data objects, whether high level data products or individual NetCDF files, have a Persistent Identifier (PID) to identify them univocally. More information can be found in Requirement 13.

The repository does not maintain explicit URL links between datasets, but data are generated with processing levels: L0 (original data in NetCDF format), and L1 (L0 data plus quality control, derived data, homogeneous time intervals, resampling, etc.). The repository is organised hierarchically like a file system. L0 and L1 are at the same level (deployment folder), and thus there is an implicit link between the L0 and L1 dataset versions. Delayed mode changes and advanced data transformations are represented by L1\_corr and L2 processing levels, which are also implicitly linked to their source datasets.

The repository does not explicitly compare the essential properties of different versions of the same file, but the “history” attribute provides information about the versions of the processing tools used to generate the file. The essential properties can be easily determined by keeping track of the processing tools versions (Git repositories and related best practices in software development management) and tracking the database configuration (using a system that relates all changes to specific users). The nature of the data processing can thus be determined along with the properties generated.

The “data\_mode” global attribute indicates whether the data was processed in real time, delayed time or delayed mode. Further explanations of these data modes are given in the SOCIB data policy:  
<http://socib.es/?seccion=dataCenter&facility=accessPolicy>

As mentioned, the whole data lifecycle is automated, from the sensors sending their data to the final introduction into the repository made by the corresponding processing application. When a new instrument or sensor is added to the SOCIB sensor network, Data Centre technicians add the new configuration to the databases, set up the necessary QC automated tests, develop new parsing functions (if needed) and finally ensure that the data are correctly added to the repository.

Real-time QC tests are performed during the automatic process, which flag individual data bits with specific quality levels. These tests are discussed in the document R10-D5. They are based on several tests approved by the relevant

communities, and after checking the variables a quality flag is added, following the Argo Quality Flag scale (see R11-D1). The values applied to Quality Control mainly follow the framework of operational oceanography programs. For example, the recommendations issued by the EuroGOOS Data Management Exchange and Quality Working group (EuroGOOS Data MEQ, see R7-D6) are applied in addition to the quality control procedures of the Argo community: Argo CTD and trajectory data (see R7-D8), Bio-ARGO (see R7-D9) or EGO gliders NetCDF format (see R14-D2).

Additional information is given in Requirement 11, concerning the SOCIB Data Quality Strategy. SOCIB staff periodically revise the data manually. Any errors are then reported through an internal form and fixed as soon as possible.

All changes in the metadata and configuration databases, which are used by the processing tools, are tracked. The ingestion process is also automatically monitored, and the results are reviewed by technical and scientific staff to ensure the authenticity and integrity of all data collection. In addition, a software tool is run in the background to validate the metadata included in the resulting NetCDF files, and those used during the processing. This software sends a notification when any information appears to be missing or incorrect, and it registers the issue so it can be fixed in later revisions.

The main processing application has a notification system that sends an email to the Data Centre staff when a problem is detected. Any issues are stored in an internal database and remain open until they are fixed.

All changes in the configuration database are tracked. Thus, when any problem due to human error changes the configuration, both the change and who made it can be checked, thus reducing the time required to solve it.

#### Linked documents:

R7-D1\_NetCDF Climate and Forecast (CF) Metadata Conventions (

<http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.pdf>)

R7-D2\_OceanSITES Data Format Reference Manual (

[http://www.oceansites.org/docs/oceansites\\_data\\_format\\_reference\\_manual.pdf](http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf) )

R7-D3\_IMOS NetCDF User's manual(

[http://imos.org.au/fileadmin/user\\_upload/shared/emii/IMOS\\_netCDF\\_usermanual\\_v1.2.pdf](http://imos.org.au/fileadmin/user_upload/shared/emii/IMOS_netCDF_usermanual_v1.2.pdf) )

R7-D4\_PUM-DCF\_instrumentation-database-processing-configuration (

<https://repository.socib.es/repository/entry/show?entryid=97ea3a7f-a8a0-445f-860a-1352b78bd4a9> )

R7-D5\_ DataCite Metadata Schema Documentation for the Publication and Citation of Research Data(

[https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel\\_v4.0.pdf](https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf) )

R7-D6\_Recommendations for in-situ data Near Real Time Quality Control (

[http://eurogoos.eu/download/Recommendations-for-RTQC-procedures\\_V1\\_2.pdf](http://eurogoos.eu/download/Recommendations-for-RTQC-procedures_V1_2.pdf) )

R7-D8\_Argo Quality Control Manual for CTD and Trajectory Data ( <http://dx.doi.org/10.13155/33951> )

R7-D9\_Argo quality control manual for biogeochemical data ( <http://dx.doi.org/10.13155/40879> )

R7-D10\_DMP\_HFR\_SOCIB-Coastal-High-Frequency-Radar-Data-Management-Plan (

<https://repository.socib.es/repository/entry/show?entryid=181f7e65-081b-4973-ad46-379af372b8bb> )

R10-D5\_QUID\_DCF\_SOCIB-QC-procedures (

<https://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6> )

R11-D1\_Argo quality control manual (

<http://www.coriolis.eu.org/content/download/370/2828/file/argo-quality-control-manual.pdf> )

R14-D2\_EGO gliders NetCDF format reference manual( <https://archimer.ifremer.fr/doc/00239/34980/> )

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:

Approved.

##### **Reviewer 2**

Comments:

## **8. Appraisal**

*R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.*

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Extra guidance and reports helpful - some more generic information on preferred/acceptable formats (outside of DM plans) might be beneficial in the future.

### ***Response:***

The data selection for the repository is conducted by qualified technicians or scientists based on the Global Ocean Observing System (GOOS, <https://www.goosocean.org/> ). SOCIB adopts common standards and focuses on the Essential Ocean Variables ( [https://www.goosocean.org/index.php?option=com\\_content&view=article&id=14&Itemid=114](https://www.goosocean.org/index.php?option=com_content&view=article&id=14&Itemid=114) ).



As a general rule, the SOCIB Data Centre receives a minimum required level of metadata to interpret and use for the data provided (Principal Investigator, abstract, variables, project, units, instruments and sensors, etc.). This information is included in the management database (“Instrumentation” application).

If this information is not introduced properly, the processing tools will not produce the corresponding NetCDF files. Only essential file system users have write permissions for the repository. These few users are permitted to introduce or modify files following well-defined rules and attributions (a permission policy document is available to reviewers on request).

All data holdings are under the umbrella of Data Management Plan documents, in which all required metadata are specified. The level of curation of the datasets is increased in a later step (delayed mode), and at each step new metadata is added describing the method of the curation used.

For example, scientific calibration procedures on glider data collection are routinely produced. Information concerning the correction is recorded in the metadata file, along with the creation of the corresponding corrected variables and assignment of their associated errors. The information required in the metadata file must be precisely defined to guarantee the traceability of the processing (Allen et al., 2017, <https://repository.oceanbestpractices.org/handle/11329/884> ).

The preferred formats are listed in the Data Management Plans (DMPs). Each internal SOCIB program (fixed stations, lagrangian platforms, research vessel, beach monitoring, etc.) has a DMP, along with each European project (see example R7-D10).

Oceanographic data formats depend on the type of data (ocean profiles, time-series, satellite, geographic information systems, etc). The supplied version of a data file, which comes from the different instruments and sensors, is known within SOCIB as a “raw file”. This closely corresponds to the OAIS Submission Information Package (SIP), and the file formats are normally American Standard Code for Information Interchange (ASCII), comma delimited value (CSV), and Network Common Data Form (NetCDF) with Clime-Forecast (CF) conventions formats, although as mentioned, the exact file formats are described in each DMP.

Once the raw files are processed (ingested) through the corresponding applications, all elements of the deposited files are transformed into a valid preservation format, which is NetCDF as the de facto standard in the designated community (oceanography). This transformation is performed by in-house developed applications and toolboxes (a processing application, glider toolbox, radar toolbox). In some cases, they can be adapted to ingest new data formats (creating a new parsing function), but as this may not be trivial it must be considered during workflow planning. The maintenance of the processing applications and their backwards compatibility (reprocessing) plays a key role in ensuring SOCIB can ingest the data.

The information in the resulting OAIS Dissemination Information Package (DIP), or the NetCDF file, may come from a collection of SIPs. Thus, a one-to-one relationship between SIPs and DIPs may not always occur.

The version resulting from the ingestion process (NetCDF file) is also an Archival Information Package (AIP) in OAIS terms, as are the original raw files. All actions related to the preservation of the data are documented in the corresponding Data Management Plan (DMP).

More information about the pre-ingestion and the ingestion processes is provided in the Preservation Plan (see the Data Management section <https://www.socib.es/data/data-management> or R10-D1).

Linked documents:

R10-D1\_STRP\_SCB\_Data-Repository-Preservation-Plan (

<https://repository.socib.es/repository/entry/show?entryid=504dd999-0e5e-49c0-b2e1-790559055fd0> )

R7-D10\_DMP\_HFR\_SOCIB-Coastal-High-Frequency-Radar-Data-Management-Plan (

<https://repository.socib.es/repository/entry/show?entryid=181f7e65-081b-4973-ad46-379af372b8bb> )

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:

## **9. Documented storage procedures**

*R9. The repository applies documented processes and procedures in managing archival storage of the data.*

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:  
4 – The guideline has been fully implemented in the repository

### ***Response:***

As introduced in Requirements 1 and 6, the SOCIB Data Centre has been working since 2018 in a Data Management Program based on a Data Management Maturity model (CMMI Institute, <https://cmmiinstitute.com/resource-files/public/dmm-model-at-a-glance> ), to enable its data management capability to be developed, improved and measured. This Program is organised into five categories (Data Management Strategy, Data Governance, Data Quality, Data Operations, and Platforms & Architecture) and has a direct impact on the data repository. Each category contains different process areas that help with the organisation of the documentation and the management of the relevant processes and procedures.

The data lifecycle processes of the data products in the repository are documented in the Data Management Plans (DMP). These include metadata, data to be collected and managed, data format and standards, acquisition, assembly, processing, archival, preservation and dissemination, workflow, and the identity of the depositors and their roles.

As mentioned in Requirement 8, all of the SOCIB internal programs (fixed stations, lagrangian platforms, research vessel, etc.), their corresponding day-to-day operations, and projects (national and international) are documented in the corresponding DMP. An example of the DMP of the High Frequency radar can be found in R7-D10.

SOCIB uses the Data Management Program (DMP<sub>r</sub>) as a management techniques strategy, similar to OAIS ( <http://www.oais.info/> ). SOCIB's preservation plan (see R10-D1) is mainly based on the OAIS model, and is a good example of the relationship between them.

The Data Protection and Data Backup Plan is documented in several internal and private documents managed by the Computing & IT Service. The main elements of data protection and data integrity are:

1. Define an appropriate user profile strategy with different levels of access: domain users with read-only access and full-access facility superusers.
2. Maintain a filesystem access control policy including three access levels: read-only, full rights and root access.
3. Use suitable RAID levels on storage systems and servers, as a mechanism for protection against hardware failures and data corruption.
4. Establish and operate an IT infrastructure monitoring and alert system, for faster response and recovery.
5. Use an external monitoring service (checks every 30 minutes) to guarantee the availability of online data and services.
6. Keep IT redundancy, in terms of hardware components, servers and networks to ensure data and services availability.

The Computing & IT Service is responsible for data protection and backup policies, protecting data from compromise and ensuring that they can be restored quickly and safely after any corruption or loss. The data backup strategy for each data stream is designed in accordance with its nature, volume, criticality, reproducibility and high availability requirements. The strategy is divided into five main categories: data file systems, application and log file systems, virtualization services, databases and users' data. (More details are available on request).

Backups are kept in two locations (SOCIB and IMEDEA offices) separated by more than 10 kilometres (for more information, see Requirement 16).

The consistency and integrity of backups is ensured by a backup quality control system, which establishes a combination of manual and automatic methods for guaranteeing the quality of copies. The system consists of three components:

1. A monitoring and alert system to ensure the integrity of the backup procedure (both for local and remote procedures).
2. IT Staff carries out a weekly manual overview to ensure that all the backups are properly completed (weekly summary control).
3. Historical backups are checked every three months.

The Backup Quality Control System is a work-in-progress, aimed at improving the quality and availability of the backup copies.

All storage equipment (see Requirement 15 for a complete description) is fully maintained and monitored, to ensure its level of performance and high reliability. Currently, the Computing & IT Service maintains hardware support agreements and service contracts for all main storage systems, to guarantee the operability of the SOCIB Data Centre. In addition, all storage systems are permanently monitored through a combination of built-in functions, custom procedures and different alert and notification services, to detect and react to any problem.

Linked documents:

R10-D1\_STRP\_SCB\_Data-Repository-Preservation-Plan (

<https://repository.socib.es/repository/entry/show?entryid=504dd999-0e5e-49c0-b2e1-790559055fd0> )

R7-D10\_DMP\_HFR\_SOCIB-Coastal-High-Frequency-Radar-Data-Management-Plan (

<https://repository.socib.es/repository/entry/show?entryid=181f7e65-081b-4973-ad46-379af372b8bb> )

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:

## **10. Preservation plan**

***R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.***

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

## *Reviewer Entry*

### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Extra information is most useful - publishing openly yearly preservation plans in the future would be beneficial in demonstrating repository work in this area.

## *Response:*

The SOCIB Data Repository's Preservation Plan is published in the Data Management section of the repository web page: <https://www.socib.es/data/data-management> (also see R10-D1).

Following the licences policy stated in Requirement 2, SOCIB has the rights to preserve the data in the repository to ensure it remains accessible, usable and understandable over time by the designated community. SOCIB will take any necessary action, such as updating metadata, forward migrating formats or re-processing data sets. The SOCIB staff are aware of the state-of-the-art of the related formats, conventions and software, and regularly participate in conferences and workshops.

Changes in the technology, methodologies and standards used by the Designated Community are considered in the SOCIB Data Repository. Through several partnerships, SOCIB is co-author/collaborator on some of the formats and standards, which are updated regularly.

SOCIB is already working on several plans related to future migrations (2021-2024), as Requirement 14 shows.

Specific actions regarding digital preservation are:

- Annual security backups of the data (Requirement 9)
- The documents repository is maintained and preserved through the implementation of the SOCIB Data Management Program
- Some SOCIB data products have a DOI (see Requirement 13 for details), and all the metadata describing them is preserved through the global DOI system (i.e., through Digital.CSIC and DataCite)
- The metadata of a NetCDF file is contained in the file itself (autocontained) in addition to the data, so the relationship between the metadata and its corresponding data is ensured.
- Each published NetCDF file has a unique identifier, which is a global metadata attribute.
- Continuity of the NetCDF format ( <https://www.unidata.ucar.edu/software/netcdf/> ), CF conventions ( <http://cfconventions.org/> ), DataCite metadata schema ( <https://schema.datacite.org/> ) and DOI system (for metadata).
- Continuity and improvement of the SOCIB Data Catalog ( <http://apps.socib.es/data-catalog/> ), management database and SOCIB web services in general. Ensuring preservation in the long-term, may thus be more difficult, but the DOI's strategy for data products and their snapshots help to solve this problem (see R10-D2).

In terms of managed documentation and preservation levels, SOCIB is continuously improving documents such as the NetCDF user manual and associated documents (see R10-D4), the Quality Control procedures (see R10-D5), and the Product User Manual data catalog (see R10-D6), to support usability in the designated community. These are publicly available through the Data Management section of the repository web page: <https://www.socib.es/data/data-management>.

SOCIB has a public Data API ( <http://api.socib.es/home/> ) to access the repository (meta)data independently of the actual file format. In addition, the SOCIB Data Centre developed a web application called Data Catalog ( <http://apps.socib.es/data-catalog/> ), on top of the API and other SOCIB web services, to provide a more user-friendly interface and to support the SOCIB DOI system landing pages. Thus, the end user need not be concerned with format changes.

The SOCIB Data Centre will make the necessary changes to the API and web services if a file format changes, to ensure continued access to the data.

The management of the data lifecycle is a main objective of the SOCIB Data Centre. This can achieve the effective management of the SOCIB data holdings included in the repository. Data management includes data acquisition processes, metadata and formats, processing and assembly, and archival and dissemination. Data Management Plans are tools used to identify and describe all steps in the data lifecycle and the actors involved. Further documentation describing specific processes, workflows and related information are linked to each Data Management Plan document. The Data Management Plans will be published in the Data Management section ( <https://www.socib.es/data/data-management> ) of the SOCIB Data Repository website.

In situ data and metadata are stored in a unified storage platform, following the backup criteria explained in Requirement 9. Automatic quality control is conducted, making the data available in a few hours (Near Real Time service is required for part of the data). These data are disseminated to other data portals like CMEMS ( <https://marine.copernicus.eu/> ) or EMODnet ( <https://www.emodnet.eu/> ). For more information, see the Data Dissemination Report draft document (R10-D3).

Data processing involves managing different processes such as standardisation, data conversion, and data validation. Processes include data ingestion, quality control, the generation of new products and data archiving. The generation of metadata follows interoperable and international standards, to facilitate data discovery, through adopting the European Directive INSPIRE ( <https://inspire.ec.europa.eu/> ) (European Commission 2007).

Linked documents:

R10-D1\_STRP\_SCB\_Data-Repository-Preservation-Plan ( <https://repository.socib.es/repository/entry/show?entryid=504dd999-0e5e-49c0-b2e1-790559055fd0> )

R10-D2\_Doi\_Versioning\_System (\*private document, shared with the reviewers)

R10-D3\_IREP\_DCF\_data-dissemination-report\_DMPPr (\*private document, shared with the reviewers)

R10-D4\_Metadata\_Management (

<https://repository.socib.es/repository/entry/show?entryid=e998fa9b-caf4-4d30-add4-d96b22214ded> )

R10-D5\_QUID\_DCF\_SOCIB-QC-procedures (

<https://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6> )

R10-D6\_PUM\_DCF\_data-products-catalog-user-manual (

<https://repository.socib.es/repository/entry/show?entryid=927c8211-610f-45fd-b63f-d844f135655c> )

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

Accept

#### **Reviewer 2**

Comments:

## 11. Data quality

*R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.*

### *Compliance Level:*

4 – The guideline has been fully implemented in the repository

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Data strategy addresses this criteria.

### *Response:*

The SOCIB Data Quality Strategy (see R11\_D3) is publicly available through the repository website on the “SOCIB Data Management” section ( <https://www.socib.es/data/data-management/> ). This is reviewed on a yearly basis. It is currently being re-addressed to meet the new requirements from the coastal ocean community and to ensure the IT systems that

power SOCIB are up to date. This new Data Quality Strategy is driven by the improvement in both (1) data assets FAIRness and (2) data quality processes and procedures. The SOCIB Data Management Program is being updated accordingly to guarantee that SOCIB benefits from the new Data Quality Strategy. Although not fully implemented, the upgraded Data Management Program is currently offering new tools and processes that improve the quality of the SOCIB data assets. Updates of the Data Management Program are conducted following two main courses of action. The first includes the three processes of Quality Assurance, Quality Control and Quality Assessment, and the second requires the IT architecture to be upgraded to ensure data quality.

As explained in Requirements 7 and 12, the real-time quality control checks applied to the data are specified in document R10-D5 . These checks are based on several tests accepted by the relevant community and after checking a variable containing a quality flag value is added, following the Argo Quality Flag scale (see R11-D1).

In Requirement 7, the documentation of the completeness of the data and metadata is discussed. The NetCDF files provided by the SOCIB repository follow the requirements of the Climate and Forecast conventions ( <http://cfconventions.org/> ). Our metadata is also partially based on the OceanSites Data Format Description (see R7-D2), SOCIB's definitions include sufficient metadata to describe the variables and their physical units and dimensions. The quality of data and metadata is assessed by considering the requirements defined in international programs.

The processing applications that create the NetCDF files ensure that all the metadata fields are filled. To check if the metadata requirements are adhered to for the relevant schema, a systematic evaluation of the metadata (socib\_Data\_checker software) is periodically run to determine whether the data quality meets the requirements of the different projects. When an error is found, it is registered and noted. After the evaluation, the data asset is reprocessed to maintain the predefined quality requirements.

SOCIB is directly involved in several international projects, which provide the opportunity to share comments and actions for rating the data and metadata.

The SOCIB Repository is interoperable, which means that Google Dataset Search ( <https://datasetsearch.research.google.com/search?query=SOCIB> ) is able to harvest the data product metadata. This is a work-in-progress that is being tested with some of our data products and serves to ensure the quality of our metadata, by considering Google's requirements.

According to the CMEMS In Situ TAC System Requirement Document (see R11-D2), the regional production units that collect data from nearby providers are designed to provide feedback when any original data Quality Control flag is revisited and changed due to a data reassessment. This is conducted through an ASCII comma-separated list (with one line per modified station) called the "Quality control feedback list". As a large proportion of SOCIB data is being disseminated to CMEMS, it is subject to comments by leading European parties involved in CMEMS data integration.

In addition, SOCIB operational oceanography and coastal in-situ data are continuously replicated in other regional or global data centres, according to the data management plan of international programs such as Copernicus Marine



Environment Monitoring Service (see R11-D2 ), EMODNET, EGO, Coriolis Argo GDAC, and ERDDAP NOAA (see draft document: R10-D3). A master copy of the data collections is kept at SOCIB.

Metadata catalogues (such as Cruise Summary reports, CSR, <https://www.seadatanet.org/Metadata/CSR-Cruises> ) are included at the European level within the SeaDataNet European infrastructure.

For example, a comparison is made between glider data and the lowered CTD data from the research vessel and corrections are made where necessary This is cited in the variable salinity\_corr (e.g., background\_data\_used\_for\_correction: Background comparison Cruises used: 1) dep0012\_socib-rv\_scb-sbe9001\_L1\_corr\_2019-03-05).

Linked documents:

R10-D5\_QUID\_DCF\_SOCIB-QC-procedures (

<https://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6> )

R11-D1\_Argo quality control manual (

<http://www.coriolis.eu.org/content/download/370/2828/file/argo-quality-control-manual.pdf> )

R7-D2\_OceanSITES Data Format Reference Manual (

[http://www.oceansites.org/docs/oceansites\\_data\\_format\\_reference\\_manual.pdf](http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf) )

R11-D2\_Copernicus Marine in situ TAC - System Requirements Document ( <https://archimer.ifremer.fr/doc/00297/40846/> )

R11-D3\_ADPL\_SCB\_data-quality-strategy (

<https://repository.socib.es/repository/entry/show?entryid=e4e6d093-63ec-4fd4-9bf3-6da66ae92c4d> )

R10-D3\_IREP\_DCF\_data-dissemination-report\_DMPr (\*private document, shared with the reviewers)

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

Approved. It is nice to see this new initiative to leverage the Google dataset search as well.

#### **Reviewer 2**

Comments:

## **12. Workflows**

***R12. Archiving takes place according to defined workflows from ingest to dissemination.***

***Compliance Level:***

4 – The guideline has been fully implemented in the repository

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

### *Response:*

SOCIB is working on several Data Management Plans (DMPs) (see example R7-D10) to establish which type of data and metadata are collected and stored, documented, made accessible, and preserved for future use. The DMPs state whether the data should be maintained under embargo, provide the necessary steps to either anonymize files or to provide secure access and to establish the security processes. The DMPs also define roles and responsibilities regarding data collection, processing, custody, access and preservation. They are live documents in which information regarding the data can be revised and will mature based on the progress of the project, and the observing and/or modelling program.

Data users can contact SOCIB through email ([info@socib.es](mailto:info@socib.es)) for information required to handle the data in specific datasets.

SOCIB Data Centre is working on several Business Process Model and Notation (BPMN, <http://www.bpmn.org/>) documents. These are graphic notations that logically describe how the work flows inside a project or observing system. Only generic roles are included (e.g., researcher, technician, data steward, etc.). Every BPMN is associated with a specific task in Asana (a software-as-a-service designed to improve team collaboration and work management) that considers the privacy of subjects/users (as the actors in the workflows).

Several approaches provide useful information concerning qualitative and quantitative checks of the outputs in the repository:

- Real time and historical data and metadata are checked manually. Errors found are reported to the Data Centre and the SOCIB facilities so appropriate action can be taken.
- Real Time Quality Control (see R10-D5) is applied to the data through checks based on several tests accepted by the relevant community. After checking a variable a quality flag is added (following the Argo Quality Flag scale, R11-D1).
- Software is executed in the background (`socib_Data_checker`). This validates the metadata included in the NetCDF file and the metadata used during the processing (deployments, processing, instrumentation installation, etc.). If some information appears to be missing or incorrect, the software will note this and register it for a later revision.
- An ongoing implementation of automatic checks is conducted within the frame of the SOCIB Key Performance Indicators strategy, which ensures operations and overall system activity information is available for community engagement and in-house performance monitoring. This implementation will rely on the Management Services, which is a set of services

returning in dataset-creation-time (indexing time) specific metrics regarding observation quality (percentage of observations qualified as good, bad etc), for example.

Before disseminating data publicly, the Principal Investigator of the observing program/project provides the Data Centre with a list of variables, units, conversion and metadata that can be made public (through the corresponding DMP). The data that do not fall within the mission are stored in the corresponding raw input data folder. This is the main folder that contains all of the non-processed data (raw format from the point of view of SOCIB Data Centre processing tools) from the various platforms, instruments, etc. SOCIB Data Centre processing tools read the files in the /socib/raw directory and select the data related to the mission/collection profile, and ignore the rest. For details about historical preservation, see Requirement 10.

The type of data managed are ocean observations with a focus on Essential Ocean Variables ( [https://www.gooscean.org/index.php?option=com\\_content&view=article&id=14&Itemid=114](https://www.gooscean.org/index.php?option=com_content&view=article&id=14&Itemid=114) ). The workflows are in continuous iteration. If ad-hoc decisions are needed to deal with special cases, they are made by the Principal Investigator and the Data Centre and documented in the metadata.

The change management of workflows is conducted through the Data Management Plan, as it is important to know who is in charge (person of contact) of every step throughout the data lifecycle (acquisition, assembly and processing, dissemination, etc).

Linked documents:

R10-D5\_QUID\_DCF\_SOCIB-QC-procedures (

<https://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6> )

R11-D1\_Argo quality control manual (

<http://www.coriolis.eu.org/content/download/370/2828/file/argo-quality-control-manual.pdf> )

R7-D10\_DMP\_HFR\_SOCIB-Coastal-High-Frequency-Radar-Data-Management-Plan (

<https://repository.socib.es/repository/entry/show?entryid=181f7e65-081b-4973-ad46-379af372b8bb> )

### *Reviewer Entry*

#### **Reviewer 1**

Comments:  
Approved.

#### **Reviewer 2**

Comments:

## **13. Data discovery and identification**

***R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.***

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### ***Reviewer Entry***

##### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Agreed - Good range of portal access to data

### ***Response:***

The search and access interfaces related to the repository are listed below:

- Observational Multi-Platform Ocean data & Ocean forecasting model data:

+ Access provided by a THREDDS Data Server: <https://thredds.socib.es>.

+ Catalog and discoverability provided by a web site: <https://apps.socib.es/data-catalog> (at present for Ocean and coastal Observing data).

- Documentation: catalogued and discoverable repository of documentation provided by a RAMADDA server:

<https://repository.socib.es>

- Data REST API: <https://api.socib.es/home/> provides search and discovery through the three dataset levels that are managed by the repository. The SOCIB API (Fernández J.G, et al., 2018, see R13-D1) data-entries endpoint allows quick access to the files containing the metocean observations.

- Some of the SOCIB datasets (

<https://datasetsearch.research.google.com/search?query=socib&docid=2QVBICJQq8pDqQxwAAAAAA%3D%3D> ) can be accessed through Google Dataset Search, which helps researchers locate online data that is freely available for use.

- Several web applications have been implemented as a response to interest from a wide range of users:

+ SACOSTA: a web-based map viewer for cartographic data such as environmental sensitivity of the coastline ( <https://gis.socib.es/sacosta> ).

+ LW4NC2: a web application for multidimensional data from netCDF files, usually from numerical models (

<https://thredds.socib.es/lw4nc2> ).

+ BEACH DATA VIEWER: a web-based map viewer that displays historical and beach survey data (

<https://gis.socib.es/viewer> , currently being rebuilt)

+ Dapp: a web application to display information related to trajectories from mobile platforms (e.g., gliders, drifter buoys, ARGO profilers; <https://apps.socib.es/dapp> )

+ SOCIB app: for real-time data from fixed stations (oceanographic buoys, sea level stations and coastal weather stations, etc.), glider trajectories and numerical models (hydrodynamics and waves). The app is available for iOS (

<http://itunes.apple.com/us/app/socib/id482542716?mt=8> ) and Android (

<https://play.google.com/store/apps/details?id=com.socib&hl=en> ).

The datasets can currently be found through name, date, status (active, completed), variable, platform type, instrument type, feature type (time series, grid, trajectory), or using a specific bounding box. SOCIB is working on different strategies to define a better search following the Designated Community criteria.

The generation of metadata follows interoperable and international standards to facilitate data discovery, whilst also adopting the European Directive INSPIRE (European Commission 2007). In addition, CF vocabularies (standard names, feature type) are part of the metadata offered by the repository and are also searchable through the various discovery interfaces. In addition, Google Dataset Search harvests SOCIB data.

The SOCIB Data REST API enables machine-to-machine data and metadata harvesting regarding multiplatform observations and data-products. This system is recommended for third party aggregators, enabling them to obtain up to date reprocessed datasets.

Digital Object Identifier (DOI) system:

A Digital Object Identifier (DOI) managed by DataCite ( <https://datacite.org/> ) is assigned to a data product, after SOCIB's scientific and technical staff consider that the level of curation and quality are adequate. The metadata scheme offered by DataCite (see R7-D5) is followed to manage the data product metadata.

Data products with a DOI have a strict semantic versioning policy ( <https://semver.org/> ) and an archiving system to guarantee reproducibility any time the data product is requested. The data product is served in a compressed file that contains the corresponding datasets (netcdf files). The compressed file that represents a data product is periodically updated, replacing the previous one. In addition, the versioning policy indicates if a copy of the previous version must be archived to maintain reproducibility. The different versions of a data product include the time range in which such a version is valid. In addition, each version includes a change-log file with appropriate information about the changes that were applied compared to the previous version. Every time a new version is generated, regardless of its importance or relevance, the metadata in DataCite is updated. The DOI assigned to a data product never changes.

Only the latest version of the product is directly available. Older versions are available on request. Further information can

be requested from the SOCIB Data Centre concerning the DOI Versioning system.

Data objects identification:

SOCIB data objects with Persistent Identifiers (PID) can be divided in three categories:

- First, high level data products, which are identified through the DOI system.
- Second, data collections belonging to high-level data products. These collections are identified by a unique local hash ID within the SOCIB data category (e.g., space name), and it is managed by an automatic system and a relational database.
- Third, the NetCDF files that comprise the data collections. Here, the PID is the unique ID attribute in the global metadata. Additionally, the file name is unique within SOCIB. These file names are managed by an automatic system and follow a documented naming convention (see R13-D2).

Ultimately, all of the data in the SOCIB Repository are contained in NetCDF files (the third category above) that are easily accessible through the Thredds server individually, regardless of the high-level data product they belong to.

Acknowledgment and citation:

The Terms of Use web page ( <https://www.socib.es/data/data-terms-of-use> ) explains how to properly cite datasets and acknowledge the source of the data.

Linked documents:

R13-D1\_IMDIS 2018 ( <https://oceanrep.geomar.de/45129/1/IMDIS2018.pdf> )

R13-D2\_SPEC\_DCF\_SOCIB\_netcdf-file-naming-convention (

<http://repository.socib.es/repository/entry/show?entryid=37944bfd-1245-48cf-8a0a-7e3f74b6d458> )

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

Approved. Numerous pathways and parameters are available to facilitate discovery, and DataCite DOIs are applied once data are curated.

#### **Reviewer 2**

Comments:

## **14. Data reuse**

***R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.***

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### ***Reviewer Entry***

##### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

Agreed - within the context of the NETCDF framework this criteria is met - wider usage outside source domain in 'use cases' or 'impact stories' might be useful in the future.

### ***Response:***

The SOCIB Data Repository has experience in preserving and disseminating data for re-use since 2009. The NetCDF classic format is used as UNIDATA ( <http://www.unidata.ucar.edu/> ) encourages the use of this format to distribute data for maximum portability. SOCIB produces extensive metadata records using the NetCDF Climate and Forecast ( <http://cfconventions.org/> ), OceanSITES 1.3 (see R7-D2), and IMOS NetCDF (see R14-D1) data format conventions. However, SOCIB's conventions may diverge from these, either with internal conventions or through using other standard conventions (e.g., EGO, see R14-D2). Following the requirements of the CF conventions, SOCIB's definitions include sufficient metadata to describe the variables and their physical units and dimensions.

The metadata corresponding to the data products are mapped to the DataCite metadata schema ( <https://schema.datacite.org/> ) and every effort is made to be as complete as possible.

Changes in technology, methodologies and normalities employed by the Designated Community are accounted for by the SOCIB repository. Through the various partnerships, SOCIB is co-author/collaborator of some of the formats and standards and is updated regularly.

SOCIB is working on several plans related to future migrations (2021-2024):

- New metadata schemas and formats will be adopted for specific platform types (e.g., Glider and HF-Radar) following ongoing European/international working groups: AZTI ( <https://www.azti.es/en/> ), CNR-ISMAR (

[http://www.ismar.cnr.it/index\\_html-1?set\\_language=en&cl=en](http://www.ismar.cnr.it/index_html-1?set_language=en&cl=en) ), and EGO.

- Integration with the SeaDataNet European data infrastructure by implementing their metadata schema and metadata completeness and accuracy are under revision.

- Sensor data management: Sensor Model Language (ML).

- Sensor Web Enablement (SWE) paradigm: through implementing and adopting the Eurofleets+ (

<https://www.eurofleets.eu/> ) research vessel processing system and data management.

To make the data understandable, SOCIB repository has several reference manuals available to the community. The most important are the NetCDF user manual and associated documents (see R10-D4), the Quality Control procedures (see R10-D5) and the Product User Manual data catalog (see R10-D6) to support usability by the Designated Community. They are publicly available through the Data Management section of the repository web page:

<https://www.socib.es/data/data-management>.

The SOCIB API home page ( <https://api.socib.es/home/> ) contains a documentation section that supports the understanding of the data through a collection of API examples (see [https://github.com/socib/API\\_examples](https://github.com/socib/API_examples) ). In addition, several web applications that are accessible through the SOCIB corporate web page ( <https://www.socib.es> ) have been implemented, responding to interests from a wide range of users that provide information on the data itself (as mentioned in Requirement 13 in the Search Interfaces section).

Linked documents:

R7-D2\_OceanSITES Data Format Reference Manual (

[http://www.oceansites.org/docs/oceansites\\_data\\_format\\_reference\\_manual.pdf](http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf) )

R10-D4\_Metadata\_Management (

<https://repository.socib.es/repository/entry/show?entryid=e998fa9b-caf4-4d30-add4-d96b22214ded> )

R10-D5\_QUID\_DCF\_SOCIB-QC-procedures (

<https://repository.socib.es/repository/entry/show?entryid=a85d659d-b469-4340-ae88-c361333c68b6> )

R10-D6\_PUM\_DCF\_data-products-catalog-user-manual (

<https://repository.socib.es/repository/entry/show?entryid=927c8211-610f-45fd-b63f-d844f135655c> )

R14-D1\_IMOS\_NetCDF\_file\_naming\_convention (

[http://imos.org.au/fileadmin/user\\_upload/shared/emii/IMOS\\_netCDF\\_filenameing\\_convention\\_v1.3.pdf](http://imos.org.au/fileadmin/user_upload/shared/emii/IMOS_netCDF_filenameing_convention_v1.3.pdf) )

R14-D2\_EGO gliders NetCDF format reference manual( <https://archimer.ifremer.fr/doc/00239/34980/> )

### *Reviewer Entry*

#### **Reviewer 1**

Comments:

Approved.

#### **Reviewer 2**

Comments:

## TECHNOLOGY



## 15. Technical infrastructure

***R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.***

### ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

#### ***Reviewer Entry***

##### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

##### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

### ***Response:***

The total storage space capacity is around 270 TB for main storage and 350 TB for backup purposes, geographically distributed between the two data centres (SOCIB and IMEDEA) but globally conforming to a unified storage system. Independent file systems provide a hierarchical storage system, organising data into the four main components of data (both observational data and numerical models), users and areas (Windows, MacOS and Linux users, and storage related to facilities and services), general services (virtual machines, applications, logs and databases) and HPC (both serial and parallel computing). Data is stored in separated shared file systems, enabling its concurrent and distributed access. The storage systems use various technologies and architectures, aimed at providing the optimum storage capability and performance for all users and groups.

Briefly, Flash storage memory and SSDs drives support high-performance storage tiers, whereas NL-SAS and SAS drives provide high-capacity tiers. Backup systems mainly use NL-SAS and SATA disks, prioritising capacity over performance.

NFS and CIFS provide the ability to share centralised file systems between different users and servers, while the iSCSI service stores servers' OS and configurations for the virtualization environment. The backup system consists of various external storage systems, for backup up of all scientific and staff data, applying appropriate archiving policies. Both storage and backup systems implement different levels of RAID technology (RAID 5 or RAID 6) to offer a high availability fault tolerance storage solution. Storage and backup policies ensure storage data are compliant with current legislation

and provide a suitable recovery system.

All storage equipment is fully maintained and monitored, to ensure a high level of performance and reliability. The Computing & IT Service maintains hardware support agreements and service contracts for main storage systems and service equipment, to guarantee the operability of SOCIB Data Centre. The Investment Plan 2021-2024 (see R3-D1) incorporates necessary upgrades, new acquisitions and any required renewal in terms of IT infrastructure, which guarantee the continuity of operations and data availability.

Software tools are provided and maintained according to the versions of standardised formats and the community guidelines. This makes them easier to generate and readable in the long term. Most software produced by SOCIB is stored in a private Git repository (document available on request). A catalog of software resources used in the infrastructure is available. A SPARQL database was recently created and the population with this information is being implemented, to make it easier to maintain and operate useful queries (work in progress).

Information about Operating Systems, software packages, firmware versions or any relevant patches or upgrading installations are fully documented, managed and kept up-to-date by IT staff.

The Data Centre system's main components are:

- An instrumentation application ("Instrumentation") developed at SOCIB to manage all platforms, metadata and processing configurations, centralised in a database through a web interface. This is developed in Java, and makes extensive use of standard, open source, community-based libraries.
- The "management" database contains the configurations of all platforms as long as the metadata or any other data necessary to perform the processing are implemented with PostgreSQL. The software used is pgAdmin.
- A processing application has been developed at SOCIB to address all collected platform data and to perform data calibration, derivation, quality control and standardisation. This is developed in Java and makes extensive use of standard, open source, community-based libraries. It uses the NetCDF Java library from Unidata. The tools to develop and manage this project include Apache Maven (build system), JUnit 5 and the Mockito framework for unit testing.
- SOCIB Data Catalog API: a REST API that enables the discovery of the SOCIB data catalog and retrieves its data directly in JSON format and through interoperable web services (OPeNDAP, WMS, WCS). The technologies used are Python, Django REST framework, Panda and Swagger (OpenAPI) among others.
- SOCIB Data Products Catalog: a user-friendly website to access and discover SOCIB's data repository. SOCIB DOI system: technologies used include Backbone.js and Django.
- A layer of RESTful web services has been developed at SOCIB to ease the development of both internal and external applications, such as web or mobile applications.

- A set of tools for data visualisation and real time monitoring developed at SOCIB.

A new architecture based on DevOps principles is under development, to introduce Continuous Integration and Continuous Deployment (CI/CD) practices in all its software projects.

Some software produced by SOCIB, which is used internally, is distributed as open-source software through SOCIB's Github repository: <https://github.com/socib>. These applications or libraries are used extensively by the community. Examples: Glider toolbox ( [https://github.com/socib/glider\\_toolbox](https://github.com/socib/glider_toolbox) ), Leaflet.TimeDimension ( <https://github.com/socib/Leaflet.TimeDimension> ).

Connectivity for real-time to near real-time data streams is ensured and has sufficient bandwidth to meet the global responsibilities of the repository. Whenever possible, remote stations are provided with reliable communications based on fiber-optic networks. Complex stations, characterised by difficult access conditions, use wireless mobile telecommunication networks of at least 4G or 3G. The SOCIB Research Vessel is equipped with a double-network infrastructure, which combines a mobile communication system (for near-shore communications) and a global mobile satellite communication system (for offshore communications). Communication from the main office is accomplished through a 1Gbps monomode optic fiber connection with the University of the Balearic Islands (UIB). In addition, another monomode optical fiber link between the secondary office (IMEDEA) and UIB is in place. UIB is the local node of the RedIRIS Spanish National Research and Education Network (NREN), which connects most Spanish universities and research centres.

Linked documents:

R3-D1-SOCIB\_Strategic\_Plan\_2021\_2024 ( <https://repository.socib.es/repository/entry/show?entryid=328e6a1d-7630-4672-a086-0be6ebae5529> )

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:  
Good description of supporting infrastructure and software technologies used.

## **16. Security**

***R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.***

## ***Compliance Level:***

4 – The guideline has been fully implemented in the repository

### ***Reviewer Entry***

#### **Reviewer 1**

Comments:

4 – The guideline has been fully implemented in the repository

#### **Reviewer 2**

Comments:

4 – The guideline has been fully implemented in the repository

## ***Response:***

The head of the Computing & IT Service is the nominated security officer. This person is qualified and has a Master's degree in Computer Science from the University of the Balearic Islands. The Computing & IT Service team performs daily security oversight and incident response.

Data holdings are stored in two Data Centres separated by more than 10 kilometres; the main Centre is located at the ParcBit IT Campus ( [https://www.socib.es/?seccion=textes&id\\_textotextes=localizacion](https://www.socib.es/?seccion=textes&id_textotextes=localizacion) ) (Palma, Mallorca) and the secondary Centre at IMEDEA ( <https://imedea.uib-csic.es/donde.php?lang=en> ) (Esporles, Mallorca). Access to both campuses is controlled by gatekeepers and security services.

These two Data Centres are secured areas with access limited to authorised personnel (Computing & IT Service staff). Access is only possible using keys that are kept in a secure and safe place and only accessible to IT staff. Occasional access for external personnel is permitted, but only under supervision of one of the members of the Computing & IT Service.

Both facilities are equipped with security and environmental monitoring systems over the network. A camera system allows video surveillance and monitoring to record human activity, while integrated sensors monitor variables such as temperature, humidity or air flow. These active network monitoring systems protect data holdings from security or environmental threats, providing an automatic alert notification system to ensure immediate detection and enable rapid response. In addition, the SOCIB Data Centre is equipped with an automatic fire detection and suppression system, specifically designed to operate within data centre/computing facilities. Access to the internal network (LAN) is only granted to SOCIB personnel or to authorised guest users under the supervision of SOCIB staff. Access to IT resources is granted through a user-password authorization policy, mainly based on an LDAP directory service.

Both facilities provide secured wireless networks (WPA2), protected by password and isolated from the wired local network. SOCIB personnel and guest users access segregated networks (with different identifier SSIDs), to control their access to network resources.

When members leave the organisation their user accounts are immediately disabled. Any former member maintains access to their personal data for a month before being erased. The work-related content is kept in an archive location in the organisation's storage resources.

External network access is controlled at the first level by the University of the Balearic Islands (UIB) IT and Network Department, and at a second level by SOCIB C&IT Service. Most of the data are open and accessible to the general public, but in special cases some network access control policies are applied to ensure private access to protected resources. Access control is provided by different mechanisms such as hardware firewalls (under UIB supervision), software firewalls or intrusion prevention software (prevent DDoS attacks). All network access is monitored and access is logged and monitored; any suspicious requests are black-listed and access is prohibited.

Access to some of the services and applications are restricted only to authenticated users. Typically, the authorization relies on the LDAP service described above. In some circumstances, external users with granted access to specific restricted resources are managed by means of a hidden (access protected within the organisation) configuration file or database record containing their credentials. This method will be overridden when the new SOCIB corporate website is deployed (early next year), as this will contain a new

Authentication and Authorization system. These restriction methods are used to manage the private access to the data sets that are not fully accessible (publicly), as mentioned in R2.

All IT equipment connected to the SOCIB network is fully protected with complete, updated and trustable software (antivirus, firewall, anti-malware software, etc.). The E-mail service is protected with spam filters, antivirus software, phishing protection and management of authorised devices.

SOCIB personnel must agree to an Information Technology contract clause to ensure good practices for the users of the IT infrastructure.

#### *Reviewer Entry*

##### **Reviewer 1**

Comments:  
Approved.

##### **Reviewer 2**

Comments:

## **APPLICANT FEEDBACK**

### **Comments/feedback**

*These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.*

## ***Response:***

### ***Reviewer Entry***

#### **Reviewer 1**

##### Comments:

Overall, this was a comprehensive and well-written application. A demonstrated major strength of this repository is its attention given to discipline-specific metadata, formats, curation and data quality approaches, along-side active community engagement to continually improve these practices.

#### **Reviewer 2**

##### Comments:

A strong application for a project repository, compliance demonstrated at levels 3 and mostly 4.

Clearly laid out within the CTS document provided, the web sites links and additional documents made available to the reviewers.