The GHRSST Central Catalogue
User Guide

Document Management			
Reference:	UserGuideCatalogue1-0.doc		
Version:	UserGuide Catalogue 1.1	**Document Revision:**	1
Date of issue of this version	25 May 2023		
Document type:	Microsoft Word [Compatibility Mode]		
Book Captain:	Jean-François Piollé (IFREMER)		
Author:	The Group for High Resolution Sea Surface Temperature Science Team		
Location:	**Approved online version:** https://www.ghrsst.org/ghrsst-data-services/for-sst-data-producers/guidelines-for-data-producers/		

The GHRSST Catalogue User Guide

History of revisions

Catalogue 1.0	22 October 2022	Compiled by IFREMER and the Task Team Evolution of the Regional/Global Task Sharing (R/G TS) Reviewed by International GHRSST Project Office Published by the International GHRSST Project Office in October 2022
Catalogue 1.1	16 May 2023	Update on the checker, PO.DACC checker replaces with IOOS checker.

Table of Contents

Introduction

The GHRSST Data Discovery and Cataloguing (DDC) service is designed to provide users with a single entry point to search for all GHRSST datasets. It comprises a centralised catalogue describing all GHRSST datasets and a file (granule) search service to retrieve any GHRSST data file of a given dataset within some time and space criteria.

The GHRSST Catalogue is accessible on the GHRSST website at: https://www.ghrsst.org/ghrsst-data-services/ghrsst-catalogue/. It is centralized, meaning it stores the description (metadata) of the different GHRSST datasets in a single repository. The metadata of these datasets needs to be edited and updated remotely directly from the catalogue's main interface. This task is performed by GHRSST Data Producers (GDP) and/or GHRSST Data Assembly Centres (DAC), each one editing and curing the metadata of the datasets it is responsible for.

This document presents how to use the GHRSST Central Dataset Catalogue from the producer and DAC perspective.

The metadata catalogue uses the Ifremer Sextant solution. Based on Geonetwork it implements ISO 19115-x metadata standards to manage spatially referenced resources.

This document describes:

- the editing/publishing methodology for a new dataset and for editing an existing one,
- the different roles and associated responsibilities (producers and reviewers), and
- provides step-by-step explanations on how to complete the catalogue forms.

1. Publishing a GHRSST dataset

Publishing a GHRSST-compliant dataset requires the producer and the distributing centre(s) to comply with the requirements: The dataset's file format must comply with the latest GDS version (GDS v2.1 at the time of writing),

- access to the data through HTTPS, FTP, OpenDAP, THREDDS, etc...
- The metadata describing the dataset, which is defined through a new entry in GHRSST Central Catalogue,
- The searchability of the dataset files through an OpenSearch service.

These requirements are defined in the GHRSST R/G TS Architecture Document, which can be downloaded from the GHRSST Zenodo community: https://doi.org/10.5281/zenodo.4926440

We provide below a short overview of the different steps that need to be fulfilled to ensure a GHRSST dataset is published.

- STEP 1: Verification of the format.
- STEP 2: Description of the dataset and metadata.
- STEP 3: Provision of data access and search points.

STEP 1: Format verification

A producer shall verify that the format of a dataset's files is fully compliant with the latest version of the GDS.

- This is verified by uploading a test file to the online GHRSST metadata compliance checker available at https://compliance.ioos.us/index.html
- If the file passes successfully the compliance checker, the result (an attached pdf file) of this check shall be sent by email to the GHRSST Catalogue **coordination team** (Email: catalogue@ghrsst.org) with the following subject: *"New GHRSST dataset <the identifier of your dataset>".* Please remember to add in cc to your email Chiara Bearzotti chb@dmi.dk

See the picture below.

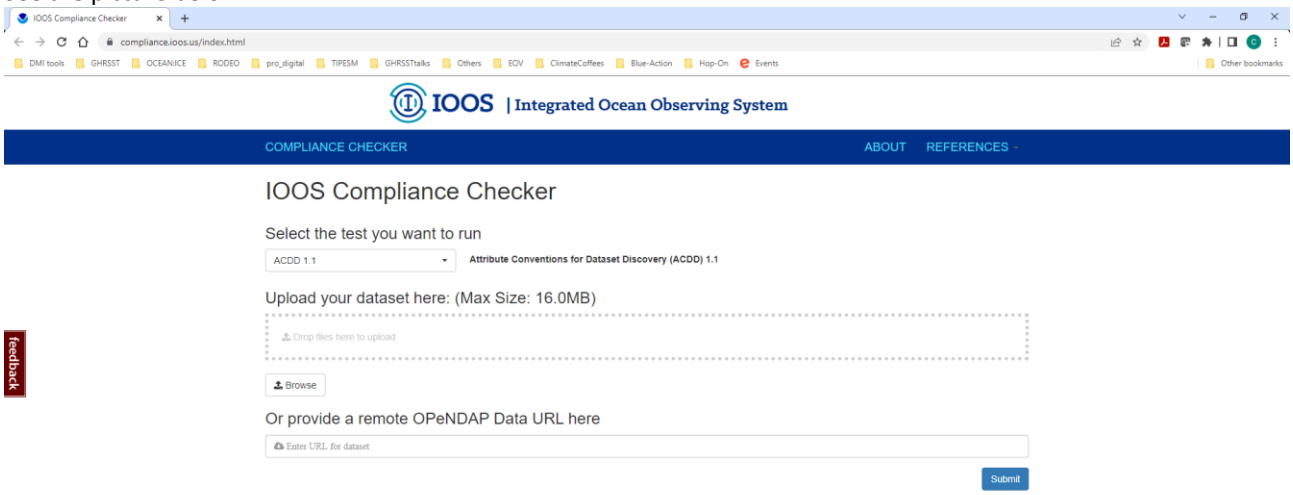

Figure 1: GHRSST metadata compliance checker.

How to attribute the correct identifier to your dataset?

The identifier of a GHRSST dataset is defined by naming conventions detailed the section 7.9 ("GHRSST Unique Text Strings and Numeric Codes") of the GHRSST format specification GDS 2.1. https://zenodo.org/record/6984989

STEP 2: Dataset description and metadata

When sending the email above, the data producer enters the description of the submitted dataset into the GHRSST Central Catalogue. This description also must be kept up to date throughout the whole life of the dataset. Please check section 2 of this manual on how to create a new catalogue entry.

STEP 3: Providing the data access and search points

The dataset producer shall also make sure that all the data access points for its dataset are described in the submitted entry to the GHRSST Central Catalogue. This information may be filled in by the producer itself or by the manager of the Data Assembly Centres (re)distributing the dataset. Please check section 3 of this document on how to edit a new catalogue entry or update an existing one.

The GHRSST Central Catalogue

The GHRSST Catalogue is designed to provide users with a single entry point to search for all GHRSST datasets and is accessible on the GHRSST website at: https://www.ghrsst.org/ghrsst-data-services/ghrsst-catalogue/

The GHRSST Catalogue is centralized, meaning it stores the description (metadata) of the different GHRSST datasets in a single repository. The metadata of these datasets needs to be edited and updated online directly from the catalogue web interface. This task is performed by GHRSST Data Producers (GDP) and/or GHRSST Data Assembly Centres (DAC), each one editing and curing the metadata of the datasets it is responsible for.

This section presents how to use the GHRSST Central Dataset Catalogue from the editor's perspective.

The metadata catalogue uses the Ifremer Sextant solution. Based on Geonetwork it implements ISO 19115-3 metadata standards to manage spatially referenced resources.

The following sections describe the editing/publishing methodology, the different roles and associated responsibilities, and gives step-by-step explanations.

Who does what? Editors and reviewers

Two main roles are set up for the editing/publishing process: the ***editor*** and the ***reviewer***.

- ***Editors*** are **GHRSST Data Producers (GDP) and Data Assembly Centres (DACs):** they respectively fill in the description of their dataset and the data access points and services (FTP/HTTPS accesses, cloud accesses, THREDDS, etc...) for these datasets. It can be one single person since many GHRSST producers distribute their data themselves and therefore act as DACs too.
- ***Reviewers*** are the members of the **GHRSST Catalogue coordination team**, a group of R/G TS task team ([1]) and GHRSST Project Office representatives ([2]), whose members are listed on the page https://www.ghrsst.org/about-ghrsst/contact-us/. The team members verify the eligibility, completeness and accuracy of the submitted information: by providing this double check they help the producer to ensure their dataset is properly accessible and documented for external users and to spot any possible issue or obvious mistake.

[1] https://www.ghrsst.org/about-ghrsst/task-teams/task-team-on-evolution-of-the-regional-global-task-sharing-r-g-ts-tt/

[2] https://www.ghrsst.org/about-ghrsst/contact-us/

While ***editors*** can only create a new metadata entry in the catalogue and/or edit an existing one, only the ***reviewers*** can make this metadata visible to the public.

Reviewers can also create and edit metadata.

	Create/modify datasets	Publish datasets
Editors (i.e. GDP, DAC)	YES	**No**
Reviewers (i.e. the GHRSST Catalogue coordination team)	YES	Yes

Table 1: Roles and permissions

2. For editors: How to create a new entry into the catalogue?

Request an account

First, you need an account to access the GHRSST Central Catalogue as an editor.

Please request an account by sending an email to: catalogue@ghrsst.org with the following subject: **"*GHRSST Catalogue Access Account*".**

In the body of the e-mail provide:

- Your full name,
- The name of your organization.

An email will be sent back to you with your credentials, usually within one or two working days.

These accounts will be created for all identified GDP, DAC, R/G TS and GPO representatives.

Please try to create only one account per GDP or DAC, to void the multiplication of accounts.

Log in to the catalogue

After getting an account, you can access the GHRSST catalogue at this link: https://www.ghrsst.org/ghrsst-data-services/ghrsst-catalogue/

To sign in, click on the "sign in" button at the top right of the catalogue (see Figure 2) and enter your credentials (see Figure 3).

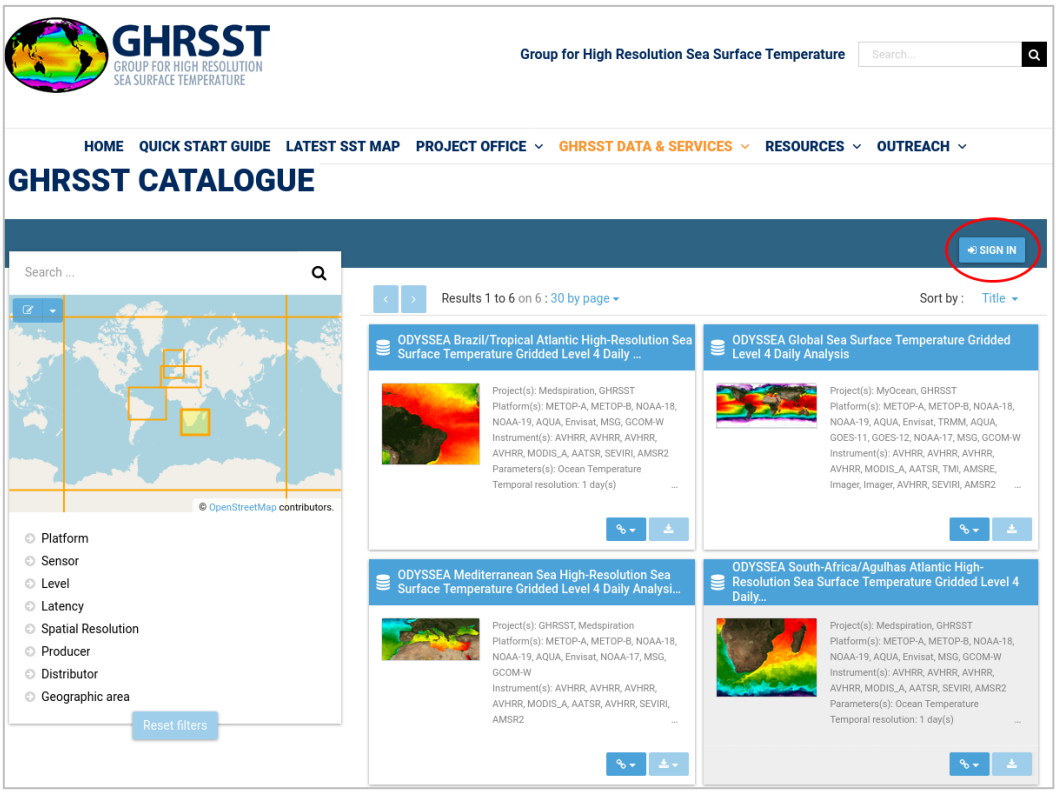

Figure 2: Main catalogue view.

Figure 3: Entering credentials. “Enter your ID (=identifiant) and your password (mot de passe)” and click on “Se connecter”.

How to enter a new catalogue entry (metadata)

Opening the editing form

After logging in, a new *Administration* menu

is available.

- Open the menu,
- Click on *New Metadata,*
- Select *Template GHRSST* and
- click on *Create*.

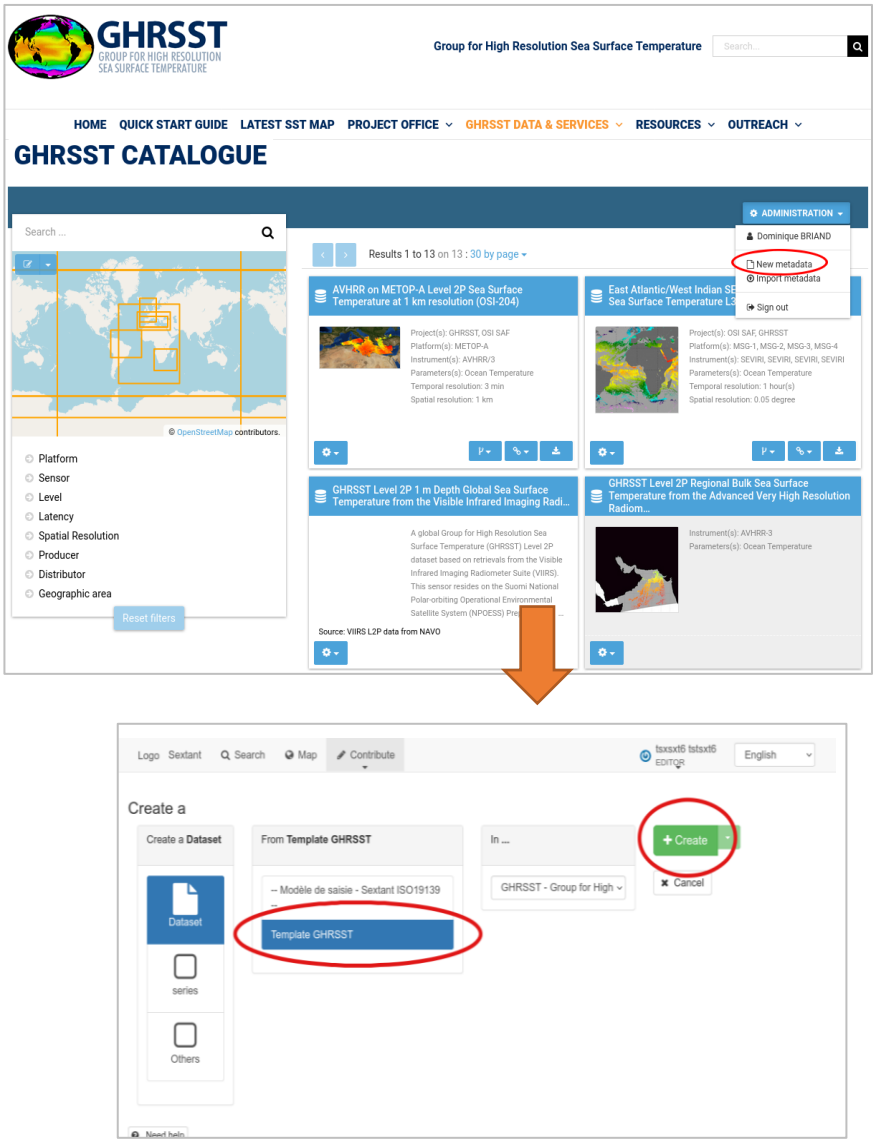

Figure 4: Create a new entry/ metadata

Additional fields for you to complete are now displayed. See the picture below (Figure 5).

- On the left side of the panel: you provide here the general description of the dataset provided in the main panel, by adding information to several tabs, i.e. **Description, Contact, Access and Usage, History**.
- On the right-side panel: you can edit the online **associated resources**, such as data access points, links to documents, etc...

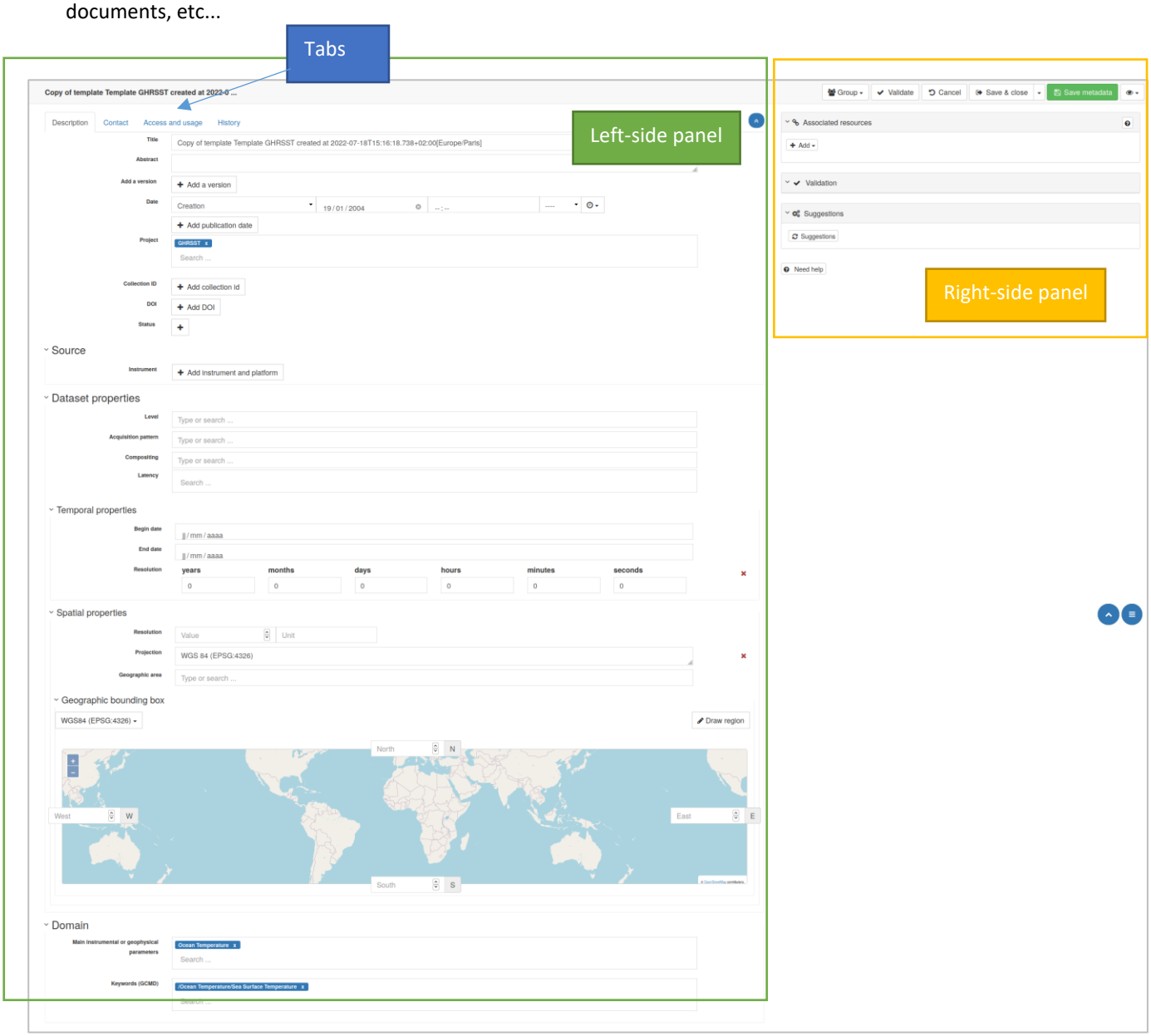

Figure 5: Editing the form with general dataset description.

Editing the form (left side of the panel)

Tab: Description

In the main panel displayed, you can now enter the following general information:

- **Title**: the label by which the dataset will be displayed in the catalogue's list of products. Be short, for instance: "GHRSST Level 4 OSTIA Global Foundation Sea Surface Temperature Analysis".
- **Abstract**: the description of the dataset. This field is the most important as it describes the dataset and should give a comprehensive view of the dataset content and processing. You can use multiple sentences and paragraphs. It is displayed in the detailed view of the dataset.
- **Version**: the current processing version of the dataset, e.g. "v1.0".
- **Creation date**: the date at which the dataset started to be produced. This information is required if you need a DOI to be created for this dataset by GHRSST Central Catalogue's management team.
- **Publication date**: the date at which the dataset started to be distributed. This informal catalogue's required if you need a DOI to be created for this dataset by GHRSST Central Catalogue's management team.
- **Project**: the project(s) in the context of which the dataset is produced and/or distributed. You should use at least "GHRSST" in addition to the other projects attached to this dataset.
- **Collection Id**: the GHRSST identifier of your dataset (the one used in the NetCDF attribute called "id"), e.g. "OSTIA-UKMO-L4-GLOB-v2.0".
- **DOI**: the parent DOI defined for your dataset (whether it was created by your organization or the GHRSST Central Catalogue's management team).
- **Status**: the current status of your dataset. Among the different values available in this selection list, the most common application cases include:
    - "*completed*": a dataset which is not updated anymore (because the corresponding mission is terminated or in the case of a completed reprocessing)
    - "*ongoing*": a dataset which continues to be updated regularly updated (such as operation datasets)
    - "*retired*": an obsolete dataset, meaning it is no more distributed, replaced by a newer dataset.

Source

The satellite mission(s) and instrument(s) are processed within the dataset. The provided lists of available missions and instruments are thesauruses. You should have only one mission and instrument for L2P and L3C datasets, whereas multiple pairs apply to L3S or L4 datasets. If some missions or instruments are missing in these thesauruses, please report the missing ones to catalogue@ghrsst.org and they will be added to the form.

Dataset properties

- **Level**: the dataset processing level (L2P, L3C, etc...). Use one of the provided values provided in the selection list.
- **Acquisition pattern**: the type of granularity corresponding to the data of the dataset:
    - Swath L2P data will be usually broken down into full *orbit*, subdividing *into*, or granules of several minutes of consecutive acquisition (*3-min granule, 5-min granule, ...).*
    - Gridded L3C or L4 data correspond usually correspond to *snapshots* (e.g. for a geostationary single acquisition) or *composite* (averaging of multiple swaths or snapshots).
    - If some granularities are missing in these thesauruses, please report the missing ones to catalogue@ghrsst.org and they will be added to the form.
- **Compositing**: in case of fusion of multiple swaths or snapshots, this field describes how the data were resampled into the dataset grid. Possible values include: "*collated*" (L3C, L3S), "*uncollated*" (L3U), or "*interpolated*" (L4). For L2P use "*None*".
- **Latency**: the timeliness of the dataset files (number of days or hours after the acquisition the files are made available to users). For completed datasets (reprocessings, etc.) use *historical* (since the data are not used anymore), otherwise use one of the typical timeliness values available in the selection list. If some timelinesses are missing in these thesauruses, please report the missing ones to catalogue@ghrsst.org and they will be added to the form.

Temporal properties

- **Begin date**: the start date of the dataset (acquisition date of the first file available in the time series).

- **End date**: the end date of the dataset (acquisition date of the last file available in the time series). Don't fill this field for datasets that keep on being updated (operational datasets).
- **Resolution**: the temporal extent of the files in the dataset. This corresponds to the time difference between two consecutive files in the dataset time series.

Spatial properties

- **Resolution**: the mean spatial resolution of the dataset, as a numerical value and a unit (km or degree), e.g. 1.1 and km respectively for a 1.1 km resolution.
- **Projection**: the used projection, as textual information (EPSG code may be added). Some recommendations:
  - *Satellite native projection* for swath products (L2P)
  - *Equirectangular* for isolat isolon cylindrical projection (most L4 datasets in GHRSST).
- **Geographic area**: a label for the covered area. Predefined choices are available ("*Global"*, "*Arctic*", "*Indian*", etc.). If some common area names are missing in these thesauruses, please report the missing ones to catalogue@ghrsst.org and they will be added to the form. It is also possible to enter this information as a free-form text.

Geographic Boundindataseter here the west/east/north/south limits of the area covered by the dataset. Use the minus sign for negative longitudes (expressed between -180 and 180).

Domain metadata

The main parameters correspond to the dataset. The values are prefilled here for GHRSST context and you do not need to edit this field.

Tab: Contact

The people and organizations attached to the dataset can be filled in by clicking on the **Contact** tab in the edition form.

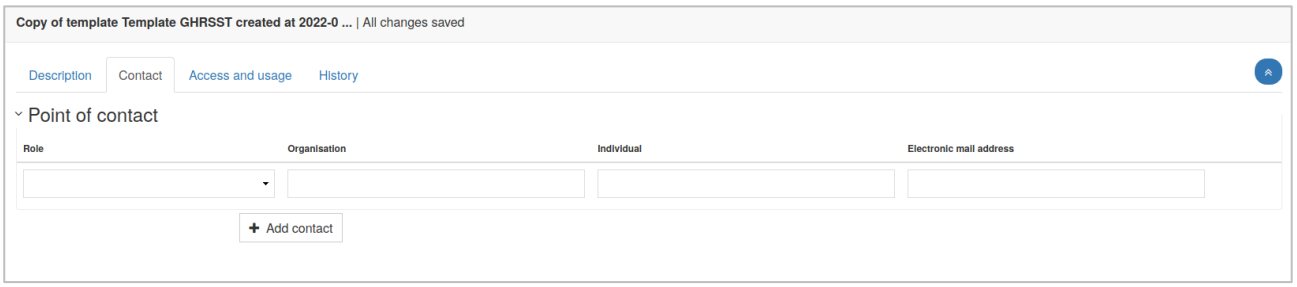

Figure 6: Editing form, adding information to the "Point of contact".

Each contact can be a physical person (by filling in the **Individual** field) or a more al entity (just filling in the **Organisation** field but not the **Individual** field).

This is where you add the following information:

- The authors of the dataset, by selecting "*Principal Investigator",* which will be displayed in the main dataset description right after the dataset abstract,
- The official point(s) of contact (which can be a help desk) for access or product issues, by selecting "*Point of Contact",*
- The producer agency or person ("*Originator*"),
- The distributing DAC or person ("*Distributor*"),
- The funding agency or person ("*Funder*").

Note that several persons or organisations can be defined for the same role e.g., a list of authors and co-authors, by defining multiple "*Principal InveInvestigators"* they will appear in the order they have been added to the system.

Tab: Access and usage

The information related to data policy and conventions can be edited by selecting the **Access and Usage** tab, which displays the following form:

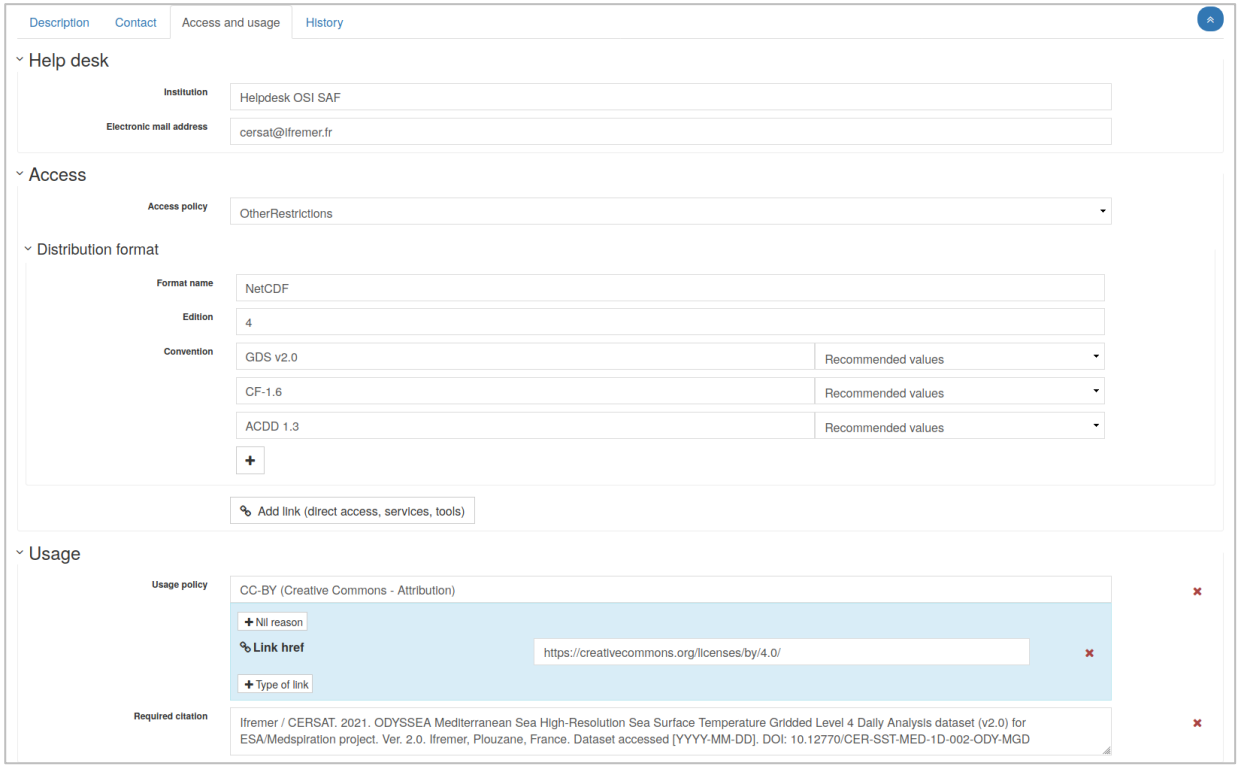

Figure 7: Access and usage.

The following fields should be filled in: help desk, access, distribution format and usage.

Help Desk

The main **Help desk** with its reference label in **Institution** and the contact email in **Electronic mail address**: this is the primary contact to which questions on the dataset should be asked, usually the producing agency. This contact point shall be able to answer or relay the question to any of the attached producer(s) or DAC(s) for the dataset.

Access

The recommended value is "*Unrestricted*" as GHRSST **data are free and open.**

Distribution format

The data format is edited in the **Distribution format** section:

- **Format name**: currently in GHRSST, it should be "*NetCDF*"
- **Edition**: the version number of the format. For GDS 2 .x datasets, it should be "*4*" as all data are in NetCDF4 format
- **Convention**: the format convention used for NetCDF; the minimum information here should be the GDS version followed by the dataset (one of *"GDS 1.0", "GDS 2.0", and "GDS 2.1"*). Other used conventions used can be added such as "CF 1.6" and "ACDD 1.3" which are used by GDS.
- **Add link (direct access, services, tools)**: this is to define the different access points, online resources (documents, etc.) for the dataset. This is equivalent to editing the online resources on the right-side panel (see section 0).

Usage

- **Usage policy**: the type of usage license associated with the data. It is recommended here to use a known license. A pretty common and recommended license for open data is "*CC-BY (Creative Commons - Attribution)*".
- **Required citation**: the citation by which the dataset should be referenced within a scientific paper or any application. The following example can be used as a template: Producer, Dataset title and version, Access date, DOI.

An example is "Ifremer / CERSAT. 2021. ODYSSEA Mediterranean Sea High-Resolution Sea Surface Temperature Gridded Level 4 Daily Analysis dataset (v2.0) for ESA/Medspiration project. Ver. 2.0. Ifremer, Plouzane, France. Dataset accessed [YYYY-MM-DD]. DOI: 10.12770/CER-SST-MED-1D-002-ODY-MGD"

Tab: History

This tab will be updated in the next future, as it refers to a foreseen evolution.

Editing the form (right side of the panel)

In this section of the panel, you need to complete the following.

Online resources

The online resources must be added on the right side of the panel.

Please click on the "+ Add" button and then **Link an online resource**.

Online resources include the description of the data access points, available documents and an image quick look illustrating the dataset.

Figure 8: Adding online resources under "Associated resources".

Image quick look image

To add an image quick look illustrating the dataset.

- Go to "**Link a"** and select "*Overview for the full resource or a subset*" from the drop-down menu.
- Add an image from your local drive by clicking on "*Choose or drop resource here*", and selecting your image file (.png, .jpeg, etc.).
- Then click on the added line in the **Metadata file storage** box. This will automatically fill in the **URL**, **Overview** and **Resource name** on the left side.
- Finally, click on the **Add an online resource** button to validate and exit.

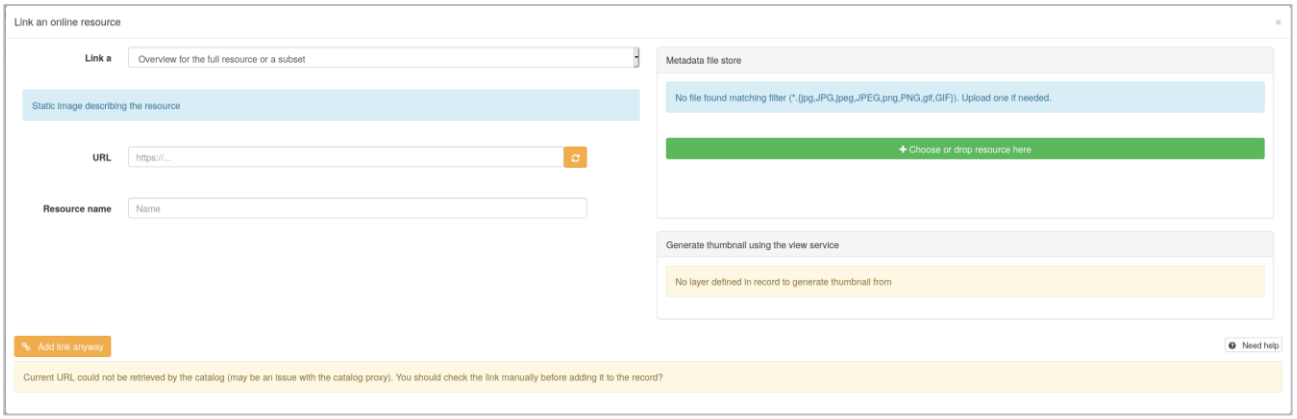

Figure 9: Adding a quick look image (1).

See the example here below.

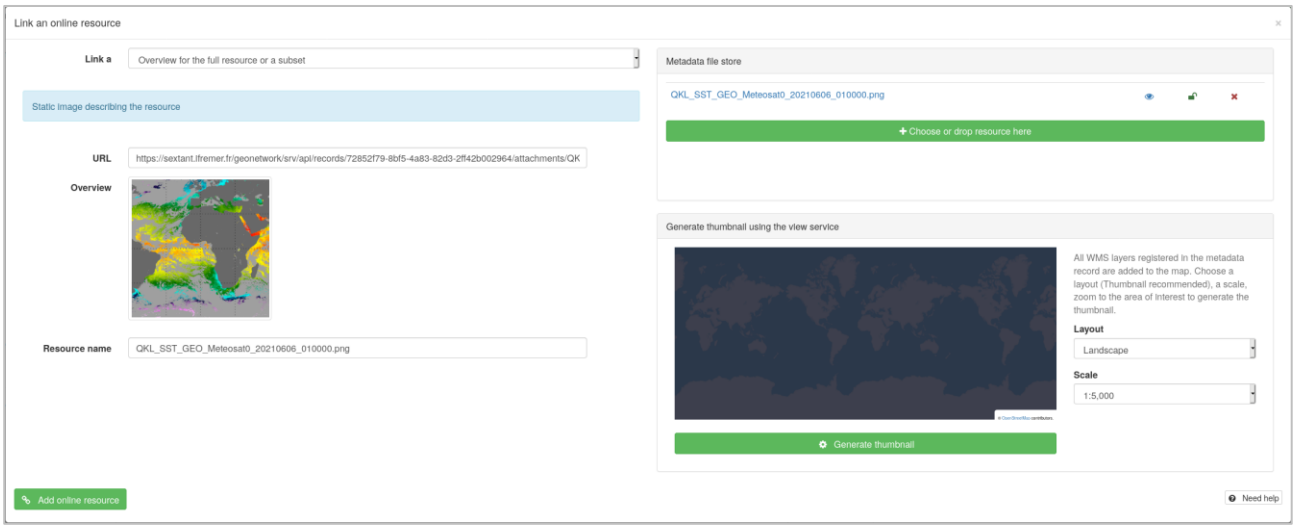

Figure 10: Example of a quick look image added to the catalogue.

Data access points and services

To add a data access point or service attached to a dataset, follow the instructions below:

- Go to "**Link a"** and select one of the items under "*Direct download*" or "*Services*" from the drop-down menu.
- This opens a new modal box that may slightly differ, depending on the download or service type.

> If some download or service types are missing in the **Link** drop-down menu, please report the missing ones to catalogue@ghrsst.org and they will be added to the form.

For instance, for an FTP access point, please provide the **URL** and **Resource name**: i.e. the label associated with the access point, e.g. "PO.DAAC FTP server".

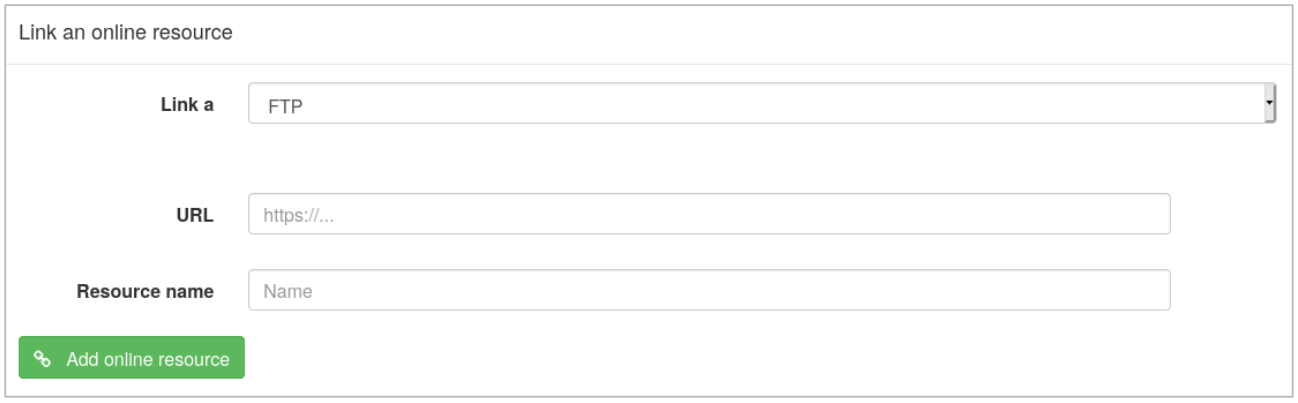

Figure 11: Adding a data access point.

Add a document to the dataset

To add a document attached to the dataset, select one of the items under "*Document*":

- *"User Guide"*
- *"Product and Validation"*
- *"Other"*

According to the category it falls into.

If your document is already available online, just copy the URL into the **URL** field and enter the document description (title, reference, etc.) into the **Description** field.

Alternatively, the document can be uploaded to the GHRSST catalogue by clicking on "*Choose or drop resource here*".

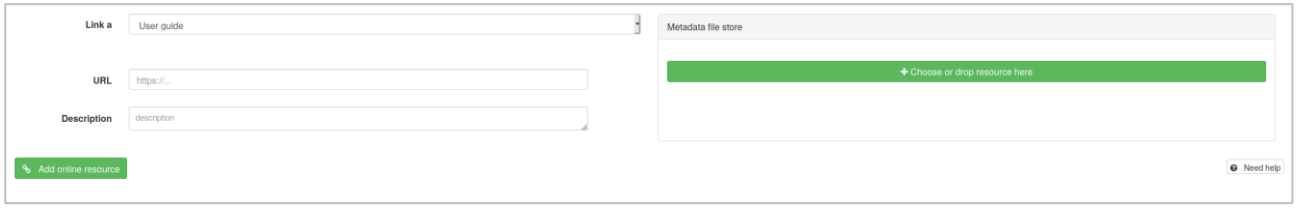

Figure 12: Adding a document.

Publishing the dataset

Once the dataset form is completed, click on **save & close**.

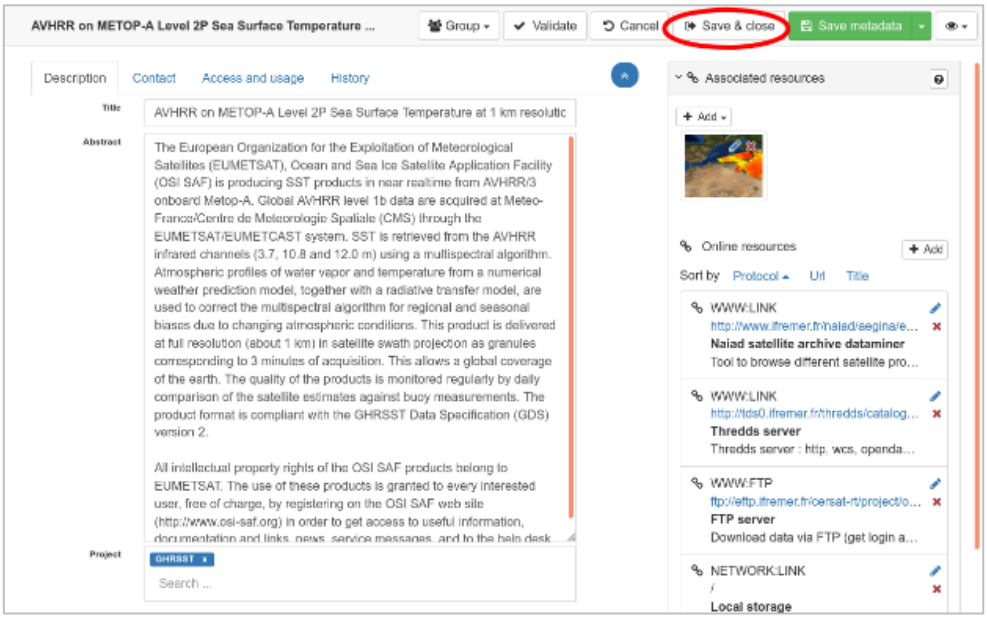

Figure 1: Saving a dataset and exiting the edition mode

At this stage, the dataset description is not yet visible publicly in the catalogue due. To publish the dataset and make the dataset entry publicly visible in the catalogue, a **reviewer** has to validate the dataset, its description and format, and then change its status to “public”.

The **dataset editor** must therefore initiate the review (i.e. the *publication workflow*) once she/he has completed the dataset description.

To activate the publication workflow, the dataset editor must:

- go back to the main catalogue view at: https://www.ghrsst.org/ghrsst-data-services/ghrsst-catalogue/
- click on the **Submit for review** menu item of the dataset entry.

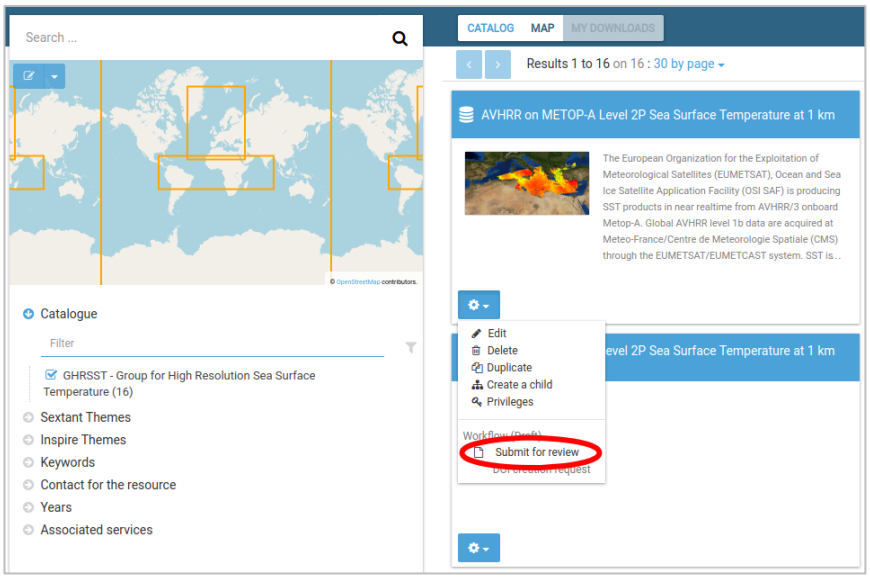

Figure 2: Submitting a dataset to reviewers

The publication workflow status for this dataset is visible in the same menu. The submission can be cancelled by the editor at any time. When the dataset is submitted for review, all reviewers receive a notification email with a link to the dataset entry.

Creation request for a Digital Object Identifier

It is recommended to provide a Digital Object Identifier (DOI) for each dataset, to make easier future reference and citation.

A DOI can be obtained in various ways, GHRSST does not recommend a preferred way to create or obtain this DOI and producers or distributors (DAC) are free to obtain this DOI in any way they want.

The GHRSST catalogue provides a way for producers to ask for a DOI creation for a dataset. To ask for DOI, go into the main catalogue view and click on ''**DOI creation request**" in the tool menu.

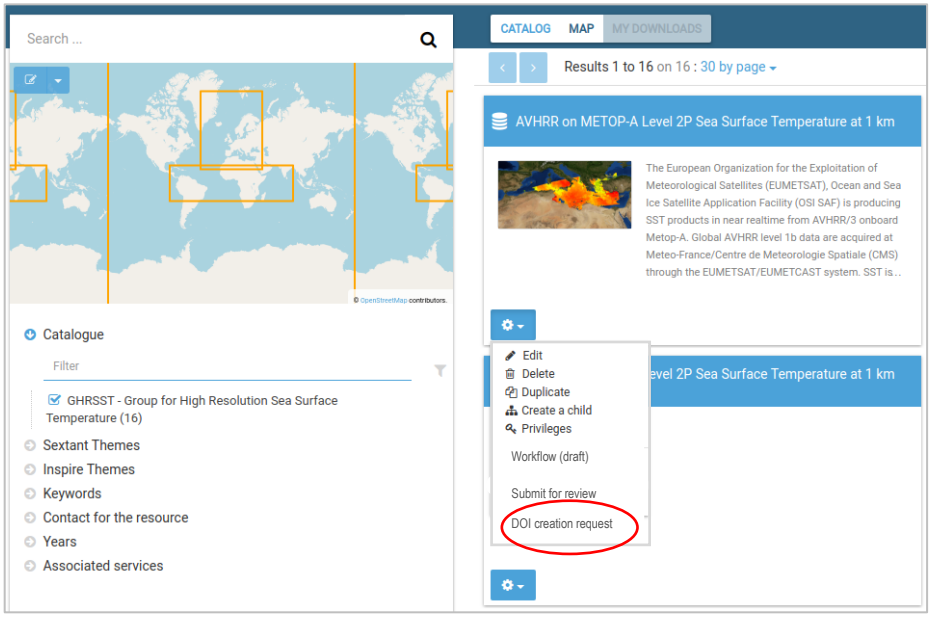

Figure 13: DOI creation request.

3. For editors: How to edit an existing catalogue entry?

Editors are requested to keep the entries up to date!

The description in the catalogue also must be kept up to date **throughout the whole life of the dataset**.

If the editor wishes to modify metadata, the editor must click on “Edit” in the metadata preview:

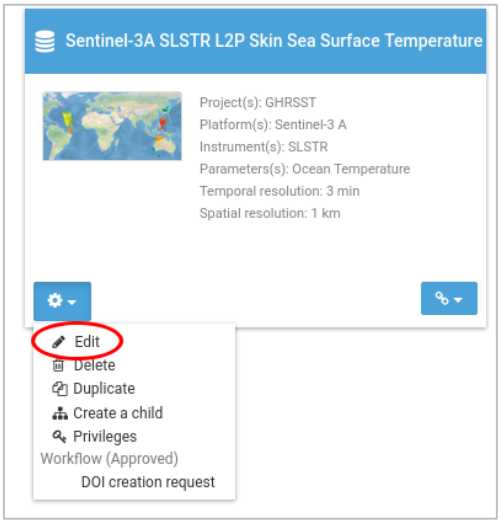

Figure 14: Editing an existing dataset.

The status of the metadata (workflow and visibility) is kept as it was before editing.

4. For reviewers: How to review a dataset?

This section is for the reviewers of a newly submitted dataset, whose role is to verify the submitted information and file format and content are comprehensive and compliant with GHRSST requirements, and help the producer or DAC to fix it if needed.

Review of the dataset description and validity of the resources

To visualise the dataset description waiting for publication, reviewers can click on the link received by email. In the top right tool menu of the record, a **workflow** section is available **to approve or reject** the dataset.

The review is a double check to support editors in publishing accurate and complete information on their dataset. In particular, the reviewers shall verify:

- **The completeness of the dataset description;**
- **The validity of the links and data access points listed with the dataset.**

Any issue (obvious error, typo, missing information) shall be reported by the reviewer to the editor of the dataset description so that the editor can fix the open issues before the dataset is made ready for publication.

How to approve or reject a dataset?

Please follow the steps indicated in the figures below.

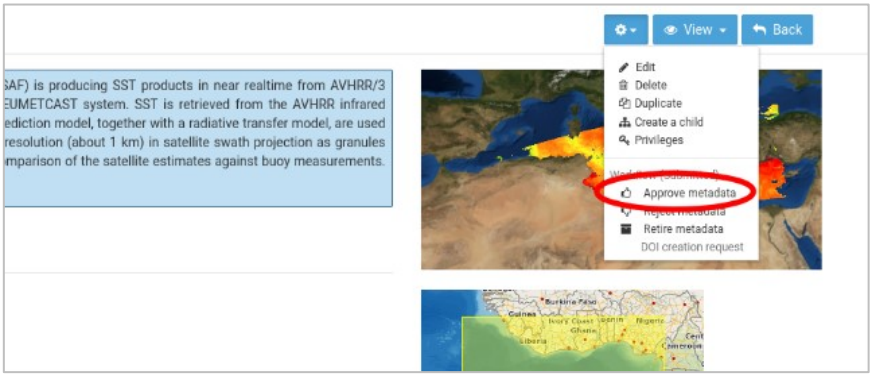

Figure 15: Approving a dataset.

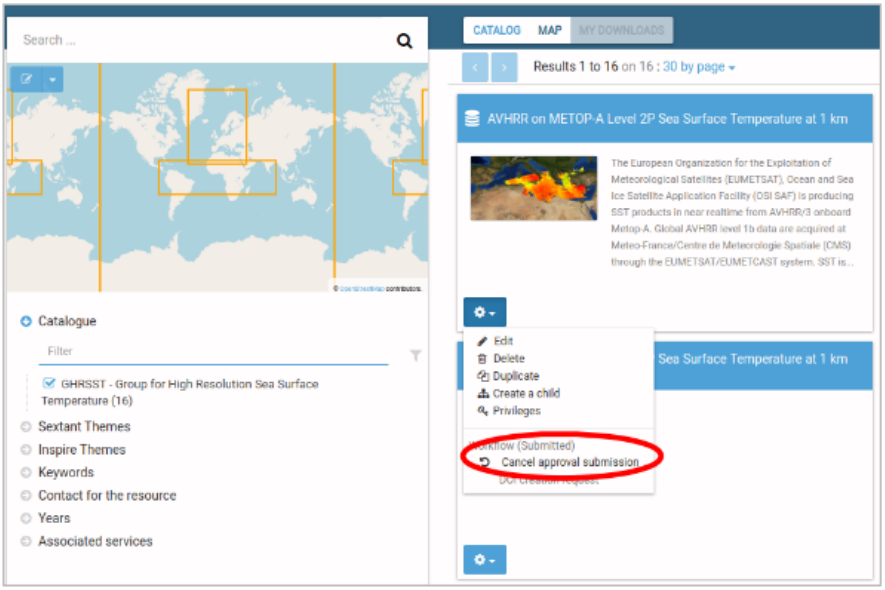

Figure 16: Rejecting a dataset, in case of missing information or some errors.

Authorisation of the dataset publication

Once the dataset is approved, the reviewer can authorize the publication of the dataset.

To publish a dataset and thus make it publicly visible the reviewer needs to **update its privileges.**

"Privileges" are accessible for each dataset from the tool menu, as in the following figure:

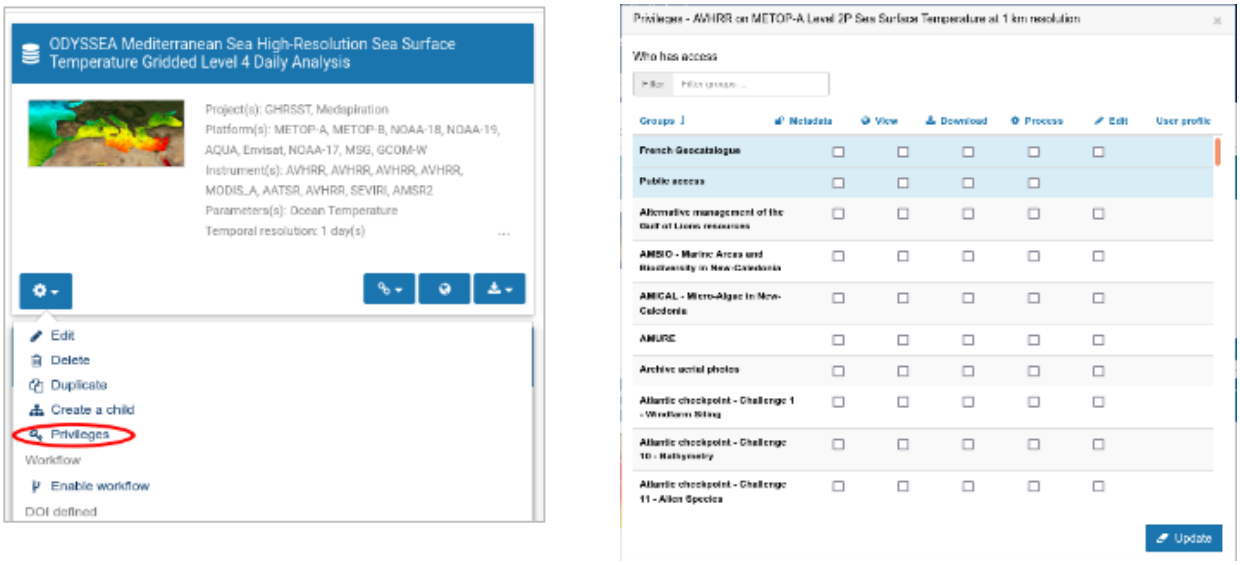

Figure 17: Opening (left side) the dataset publication status (right side).

To turn the metadata public, a reviewer needs to tick the "**Public access"** check box for the **Metadata** column as in the figure below:

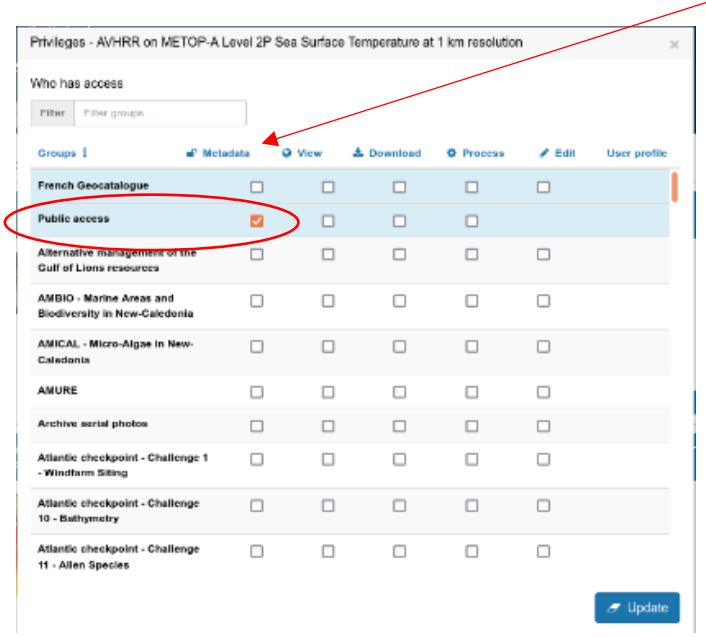

Figure 18: Checking the dataset (Metadata) for public visibility.

The metadata is now public.

5. Using the GHRSST federated granule search service

The official URL for the GHRSST federated Open search granule search service is: https://opensearch-ghrsst.ifremer.fr

The site provides examples of requests for every dataset searchable in GHRSST and a full description of the API of the OpenSearchsearch query.

A tutorial notebook was made available on the GHRSST Github account at: https://github.com/GHRSST/ghrsst-opensearch/ The tutorial provides Python functions to perform various types of search and decode the results into Python objects, and a complete step-by-step example on how to use them.