Geoscience Data Journal

WILEY

SPECIAL ISSUE Open-Source Software in the Geosciences

Geoscience Data Journal RMetS

DATA SERVICES ARTICLE OPEN ACCESS

TGCAT—A Tool to Analyse the Content of Sea Level Data Portals

Laurent Testut[1] | Adrien Laval[1] | Clémence Chupin[2] | Mikaël Guichard[1] | Begoña Pérez Gómez[3]

[1]LIENSs, CNRS – La Rochelle University, La Rochelle, France | [2]Lab-STICC, UMR 6285, ENSTA Bretagne, Brest, France | [3]Puertos del Estado, Madrid, Spain

Correspondence: Laurent Testut (laurent.testut@univ-lr.fr)

Received: 8 October 2024 | **Revised:** 10 November 2025 | **Accepted:** 5 December 2025

Keywords: data portals | EuroGOOS | GLOSS | sea level | tide gauge

ABSTRACT

The volume of tide gauge data available to the sea level community has grown substantially, with information distributed across numerous global, national, and institutional data centres. As a result, the main challenge is no longer accessing data, but identifying the most relevant dataset for a given application. Currently, more than 15 global data centres provide sea level information, each tailored to different users and use cases (e.g., real-time monitoring, delayed-mode analysis, monthly means). For users unfamiliar with tide gauge data, selecting the appropriate source can be difficult. Tide Gauge CATalog (TGCAT) is a software tool developed to address this challenge. It helps users discover where specific tide gauge data are available and assists data providers and centres in identifying inconsistencies, such as misreferenced stations or discrepancies in metadata. TGCAT collects metadata from global and national sea level data centres to produce intercomparable catalogues. It also allows visualisation of data availability timelines across multiple sources. Written entirely in Python and linked to an online dashboard (www.sonel.org/tgcat), TGCAT is designed as an open, community-based platform. Its goal is to improve data discoverability, support better referencing practices, and help users navigate the complex landscape of tide gauge data portals.

1 | Introduction

Most of the knowledge about the recent coastal sea level variations comes from the analysis of tide gauge observations which sample the world's coastline in an inhomogeneous manner. Historically, sea level measurements were taken by individual organisms or countries using various methods, technologies and levels of standardisation. Early in the twentieth century, the scientific community recognised the importance of having a consistent dataset of sea level observation to better understand the sea level variability. This led to the creation in 1933 of the Permanent Service for Mean Sea Level (PSMSL) in Liverpool, which is now hosted by the UK National Oceanography Center (NOC). This service is in charge of the collection, dissemination and analysis of mean sea level values (monthly and annual) from tide gauges worldwide. The PSMSL played a crucial role in the development of standardised methods for measuring and analysing sea level. Today it maintains a comprehensive database of more than 1500 stations worldwide and is the primary source of information for studying the long-term changes in sea level. For several decades, the PSMSL was the only one of its kind of what we will refer to as a "data portal".

The effort of standardisation and creation of global sea level data centres gained new momentum with the creation in 1985 of the Global Sea Level Observing System (GLOSS) under the auspices of the UNESCO Intergovernmental Oceanographic Commission (IOC). One of the main goals of this program was to establish a core network of about 300 permanent high-quality sea level stations homogeneously distributed around the global ocean (IOC 1990, 1997, 2012). The GLOSS program has also been instrumental in setting standards in terms of measurement metrology, data exchange formats, tide gauge control, and coordination of international data distribution efforts through the establishment of global sea level Data Assembly Centers (DACs). The GLOSS program currently relies on six DACs responsible for broadcasting different types of sea level information (real-time, fast mode, delayed mode, hourly data, mean sea level, …), targeting different applications. The six GLOSS DACs are:

- PSMSL: responsible for the dissemination and analysis of monthly and annual mean sea level;
- BODC (British Oceanography Data Centre), which acts as a GLOSS Delayed Mode Centre for assembling and controlling the hourly sea level observations;
- UHSLC (University of Hawaii Sea Level), which acts as the GLOSS Fast Delivery Centre for hourly sea level observations;
- SLSMF (Sea-Level Station Monitoring Facility) operates by the Flanders Marine Institute (VLIZ, Belgium) which acts as a GLOSS real-time centre. VLIZ also maintains the Sealevel Station Catalog (SSC) which is currently the most comprehensive metadata catalogue for tide gauges;
- SONEL (Système d'Observation du Niveau des Eaux Littorales), which is the dedicated GNSS at the tide gauge data centre for providing vertical land movement information.

In the 1990–2010 decades, the number of data integrators has grown significantly, providing sea level data with different latency levels and temporal sampling. The early 2000s marked an explosive growth in informatics and the internet. Alongside the need for integrating sea level data into tsunami warning systems, profoundly transformed the digital landscape and how users access sea level data. The earthquake of December 26th 2004 and the devastating tsunami it generated left a lasting impression and made the community of tide gauge data providers aware of the lack of stations in the Indian ocean and the need for real-time data. This event led to the creation of the SLSMF GLOSS real-time centre. In 2009 the European Marine Observation and Data Network (EMODnet) has been created to provide in situ information for the marine environment. In 2014 the European Union launched the ambitious Copernicus program of which Copernicus Marine Service (CMEMS) is a key component specifically focused on providing regular and systematic information on the state of the oceans. CMEMS recently released a reprocessed version of coastal sea level observations (Lin-Ye et al. 2023). In the meantime, other datasets or portals have been implemented which display tide gauge data from GLOSS DACs or from national institutions such as the Global Extreme Sea Level Analysis dataset (GESLA) (Woodworth et al. 2016) or the Minute Sea-Level Analysis (MISELA) (Zemunik et al. 2021) high-frequency sea-level analysis global dataset.

Therefore, sea level data users have now access to a huge number of data portals to find data, including those managed at national or provider level to such an extent that the main problem is no longer where to find data, but where to find the most relevant data for our needs (see Table A1 in Appendix A for a list of data portals). Figure 1 shows the number of tide gauges that are stored in the main global data and metadata portals.

This observation is at the root of a study to identify gaps and redundancies in the current data portals. This task was carried out as part of one of the Task Team of the European components of the Global Ocean Observing System (EuroGOOS). The Global Ocean Observing System (GOOS) was established in 1994 to provide a framework for worldwide ocean observation. The EuroGOOS Tide Gauge Task Team (hereafter EuroGOOS-TGTT) brings together the tide gauge communities in the European and adjacent seas. Its role is to compile information on existing sea level networks, advise operators and scientists, and support national and regional sea level initiatives. In the frame of the EuroGOOS-TGTT and partly funded by the H2020 EuroSea project (Pérez Gomez et al. 2021), an online live metadata catalogue for European and adjacent areas Tide Gauge Network Inventory (EUTGN) has been developed and is now operated by the Irish Marine Institute (EUTGN, ref., https://eutgn.marine.ie/). In the same framework, the Tide Gauge CATalog (TGCAT) has been developed by the French Système d'Observation du Niveaux des Eaux Littorales (SONEL) in order to identify gaps and redundancies in the main global data portals and to guide users on tide gauge data discovery. The tool has recently been extended to an online dashboard. We will present in the next section the general principle of the TGCAT tool (Section 2), the methodology (Section 3), an overview of the dashboard (Section 4), and a short description of the architecture of the code (Section 5). We will address the conclusion and perspective in Section 6.

2 | General Principle of the Tool

TGCAT was originally designed as a command line tool developed in Python to help the EuroGOSS-TGTT in analysing gaps and duplicates of the main data portals. The key concept behind the command line tool is based on the "set theory", the branch of mathematical logic that studies sets, which are collections of elements. In our case each element is a tide gauge, also called a station or a site hereafter. Each station is represented by a Python class called a **Site**. The Site object is composed of attributes (lat, lon, name, provider, link, id, data, h3) that describe the main metadata associated with a tide gauge station and necessary to identify it. These attributes are generally collected from data portals during the harvesting process (more details in Section 3). Some simple methods are also associated with the Site object to find a station by its name, id, or to compare two sites. The h3 index attributes allows each station to be uniquely assigned to an hexagonal cell at a given resolution and is a way to compare with other stations. This h3 index is computed using the open-source geospatial indexing system which considers a hierarchical hexagonal grid to decompose the world surface into hexagonal cells (Uher et al. 2019). From the longitude and latitude of the tide gauge a unique h3 index is computed at a given resolution and represented as a 15-character (or 16-character) hexadecimal string (see Figure 2). The average hexagon areas range from more than $4 \times 10e6\,\mathrm{km}^2$ (at coarse resolution: hex0) to less than a square meter (at fine resolution: hex15) (see

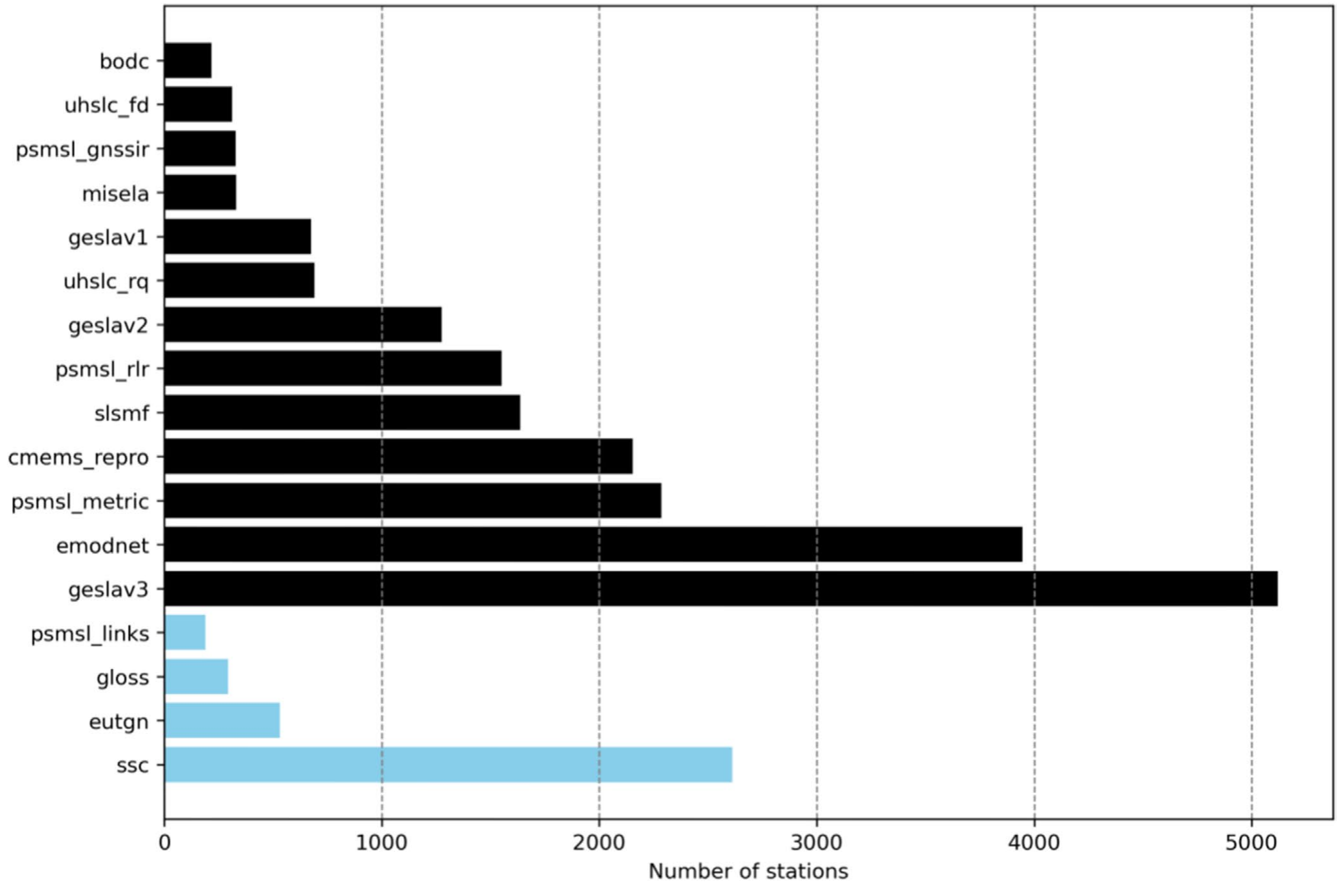

FIGURE 1 | Number of stations referenced in the main global tide gauge catalogues. The light blue bars are for metadata catalogues that don't provide any data. A detailed table of the different catalogues is shown in Table A1.

https://h3geo.org/docs/core-library/restable for more details on the resolution).

One of the main drawbacks of the existing tide gauge data management and flow is that the name and geographical coordinates of a tide gauge may substantially differ from one portal to another. Then, the coordinates and the name of a tide gauge are not good identifiers for a station. This is why each Site is assigned an h3-index.

Each Site object is concatenated into a collection of sites that compose a Catalogue object (Figure 2). The Catalogue object contains a list of Site objects and some export methods to other classical Python Objects (DataFrame, GeoDataFrame).

3 | Methodology

This section describes the main steps of the current workflow that aims to collect data and metadata from sources, in order to store them into structured Python objects (Site and Catalogue) that will be called and displayed in the online web tool (Figure 3).

The workflow in Figure 3 can be decomposed into 3 important steps.

3.1 | Metadata Harvesting

The collection of information from a source is the primary step. In our case, a source can be (i) a data portal that distributes sea level data observations (PSMSL, CMEMS, ...) (ii) a metadata portal that only gives metadata information on tide gauges (GLOSS, EUTGN, SSC, ...) or (iii) sea level datasets that can be downloaded but are not often regularly updated (GESLA, MISELA, CMEMS-REPRO). A Python function is defined for each source in order to fill the Catalogue object with metadata from this source. This stage involves using web scraping techniques or API requests to retrieve metadata information describing a tide gauge (latitude, longitude, station name, ...), that will feed the Site object for each tide gauge. Then a Catalogue containing all sites is created and named after the source. Not all metadata information may be available depending on the source and thus some Site attributes can stay unfilled. It has to be noted that for the metadata catalogue (GLOSS, EUTGN, ...) the process stops at the creation of the Catalogue.

3.2 | Data Download

The second step for a data Catalogue is to download the associated sea level time series. The download methods involve scraping web pages, accessing APIs, or using FTP protocols. The software is designed to be flexible and able to adapt to

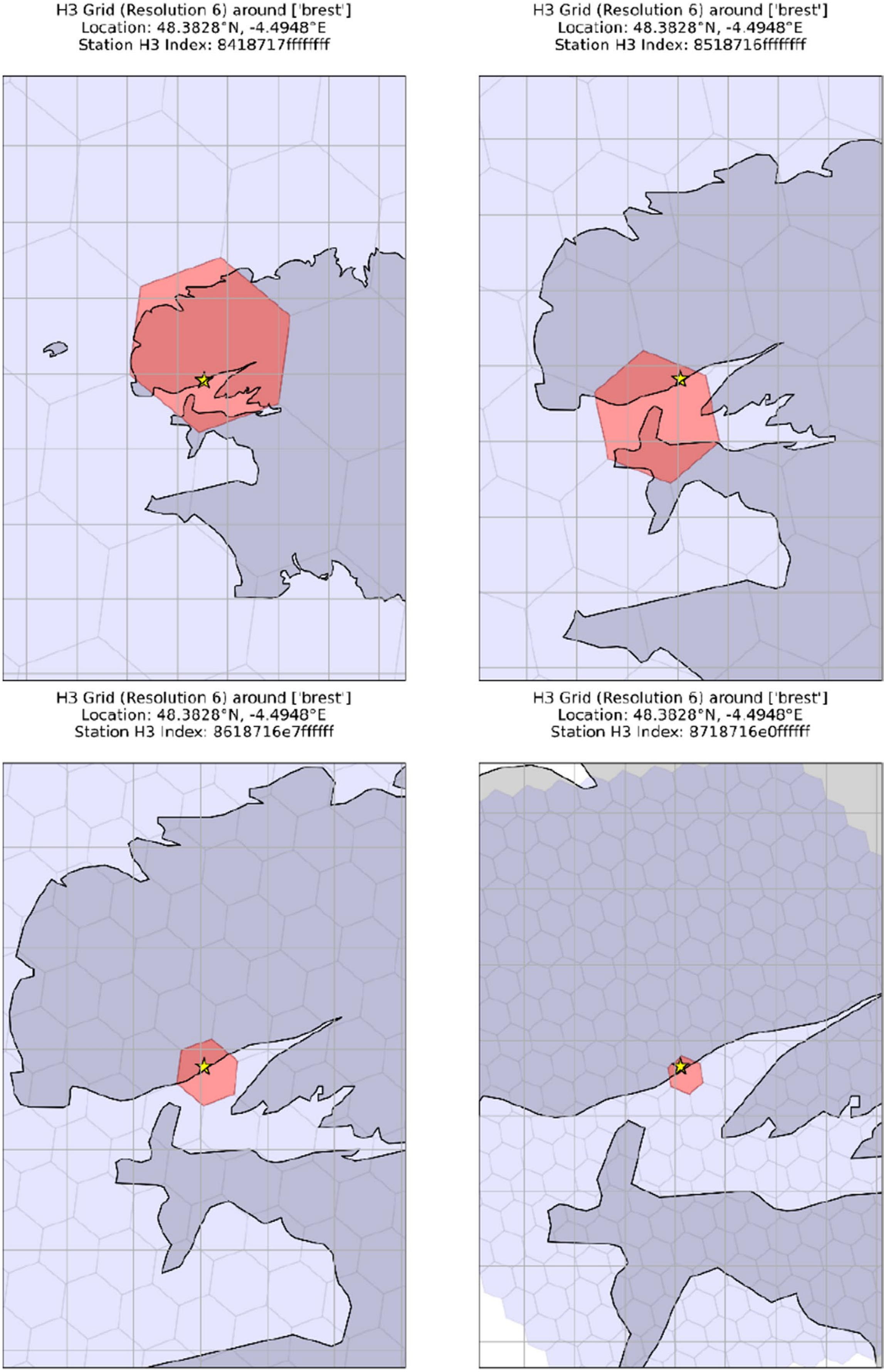

FIGURE 2 | Example of the hexagonal grid around the Brest tide gauge in France at different h3-index resolutions.

different data sources. The goal is not to make a copy of the sea level data, but rather to compute from them the chronograms. Once the chronograms have been computed, the data can be deleted.

3.3 | Chronogram Computing

Once the raw data has been downloaded, the chronograms are computed. A chronogram is a schematic representation of the sea level observations which shows the availability of data on the portal for specific periods and stations. This stage involves data processing techniques to clean and normalise the data. As the data can have several formats (.csv, .nc, .txt, .json, ...), the first step is to convert each data into comma separated value file (.csv) and homogenise the data (adding NaN where it has been omitted or annotated) to create a pandas single column DataFrame with the time and the sea level value. A custom pandas accessor has been added to the DataFrame class. The accessor takes the time of observation, the name of the station, and a timedelta value. The timedelta is the time duration that is considered to be a gap in the data. This timedelta is a threshold value and has been defined for each Catalogue. For example, all catalogues with hourly sampling data have a timedelta of a bit more than one hour. It means that all missing values in the observation greater than 1 h will be reported with a gap in the chronogram (Figure 4). For monthly sampling, the timedelta has been set to 31 days and 366 days for annual values.

When all catalogues have been downloaded and chronograms have been computed, the connectivity table is created for each Catalogue. This table associates for each h3 cell, at each

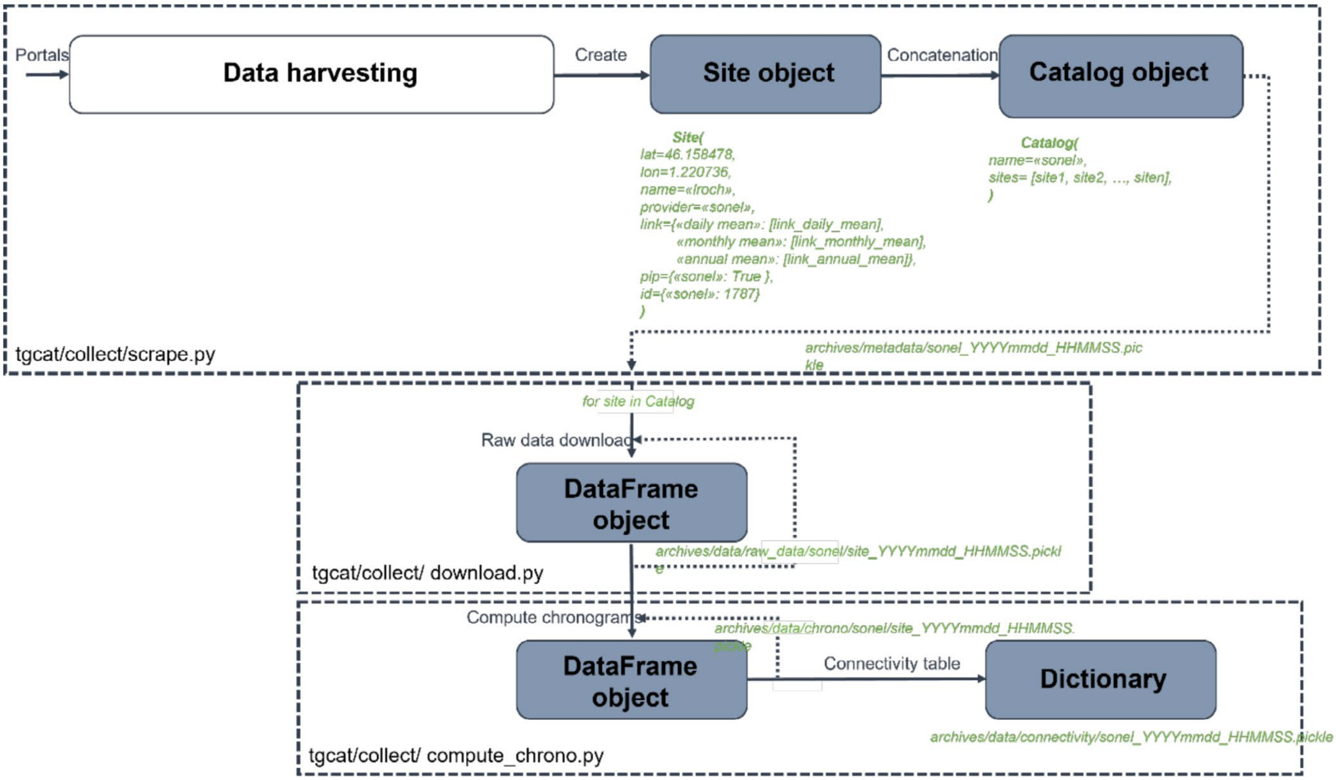

FIGURE 3 | Current workflow of the TGCAT processing chain. The first step is to collect information from a source using different data harvesting methods (ftp, api, web scraping, ...) in order to fill a Site object that is concatenated into a Catalogue object named after the source. Then if the Catalogue is a data catalogue, the sea level observations are downloaded and a chronogram is computed. The green inside the dashed box indicates the module in the gitlab repository that is concerned by this step.

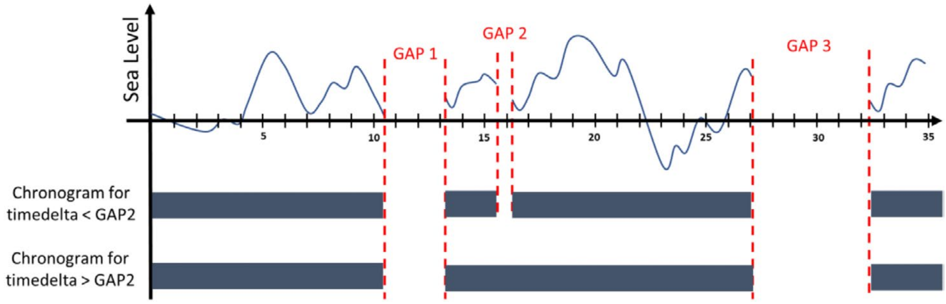

FIGURE 4 | The impact of the choice of the timedelta threshold on the representation of the chronogram. For the same timeseries, the choice of a timedelta lesser or greater than GAP2 in data will include or not a gap in the associated chronogram.

resolution, the name of all the stations it contains and the link to the associated chronogram. This connectivity table is used in the online dashboard.

4 | Dashboard

Based on the data generated by the processing chain, an online dashboard has been developed and is accessible at www.sonel.org/tgcat. Currently the dashboard is hosted in the SONEL GLOSS DAC. A job scheduler has been set up on the server to download the raw data on a weekly, monthly or annual basis depending on the data source. The online tool has been developed using the low-code framework for building data apps in Python called Dash (https://dash.plotly.com/). The entire tool (processing chain and online tool) has been developed fully in Python and should be easy to maintained, even by non-professional programmers.

Figure 5 shows a snapshot of the tool. At the opening, the homepage displays a blank map. The first action is to select the catalogues to be visualised by clicking on the popup 'hamburger' menu on the top left. It opens the catalogue selection panel, with two dropdown cells to select either metadata or data catalogues. Each catalogue can be selected by its name, which is generally the name of the source (PSMSL, GELSA, ...). Following the name of the catalogue, the 3-letter abbreviation in parentheses indicates if it is a global (GLO) or a national catalogue. In the case of a national catalogue, the 3-letter abbreviation is its country code. For example, selecting SONEL (FRA) will display the French tide gauges stored in the SONEL database. Multiple selection is possible between catalogue and metadata catalogues. Users should be aware that the order of the selection matters as the latter selected catalogue will appear on top of the others on the map and can possibly mask underlying catalogues. Once the selection is done, clicking on the "Load selected catalog(s)" button will display them on the map. By default, each catalogue has a predefined colour, but this colour could be changed using the "colour selection" menu of the catalogue selection popup panel.

When catalogues are loaded and displayed on the map, a "statistics panel" gives the users information on the total number of stations found, a table summarising the stations loaded, and a barplot with the number of stations per catalogue (Figure 6). The "statistics panel" is updated when a selection is done on the loaded catalogue. The "clear selection" button cancels all the filters.

To select sites on the dashboard, 3 filters are available (Figure 5, top of the dashboard):

1. By region or project. Predefined polygons allow quick access to some region of interest. When a region is selected on the dropdown menu, all the sites inside the predefined polygon are selected and the statistics are updated. One way to

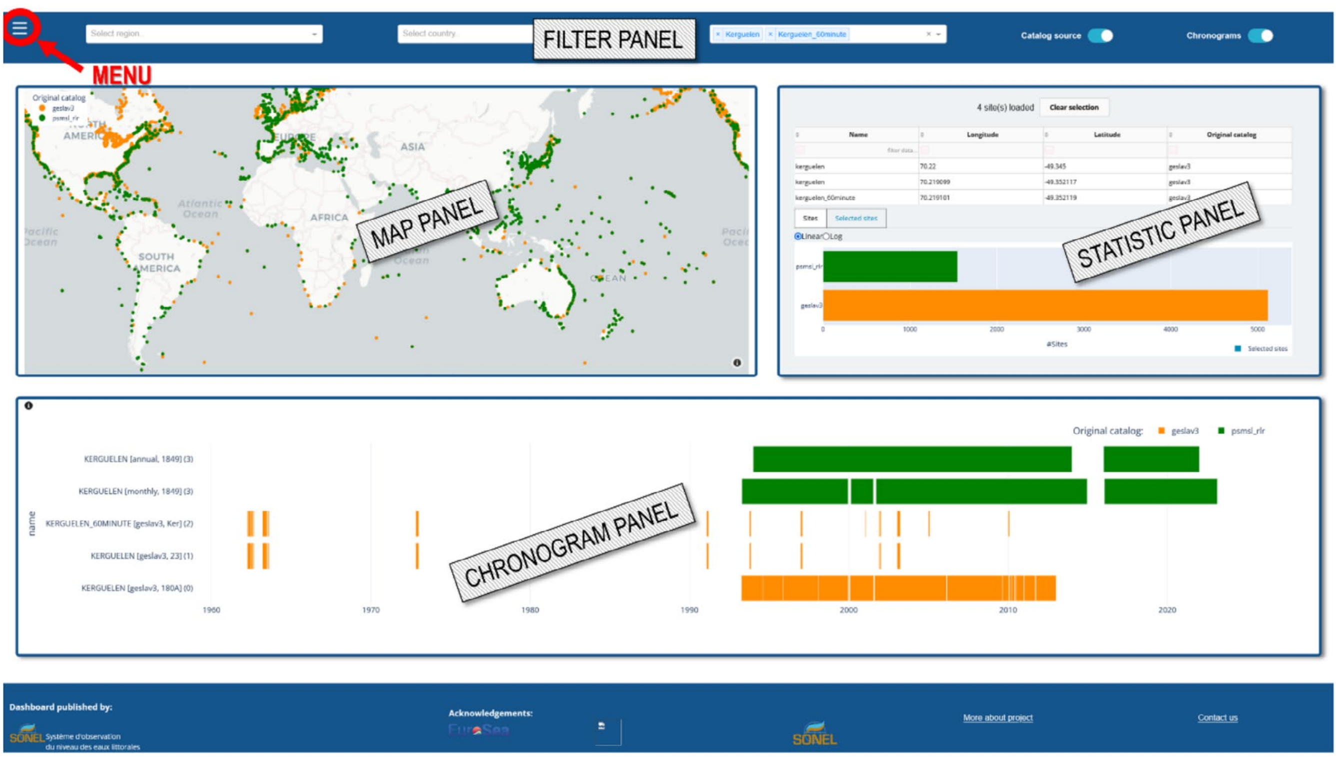

FIGURE 5 | Online version of TGCAT with 2 catalogues open (psmsl, gesla3) and 1 station selected (Kerguelen Island in the Southern Indian Ocean). The catalogue loading, export and colour selection are in the popup top left 'hamburger' menu (not open in the present Figure but showed in red). There are 3 possible filters (region, country and tide gauge) on the top of the dashboard. The two switchers on the right of the filters are to switch on/off the colour of the catalogues and the bottom chronograms. The centre of the windows is composed of a map on the left and a statistic panel on the right. The bottom chronogram panel only appears when the chronogram switcher (top right) is on.

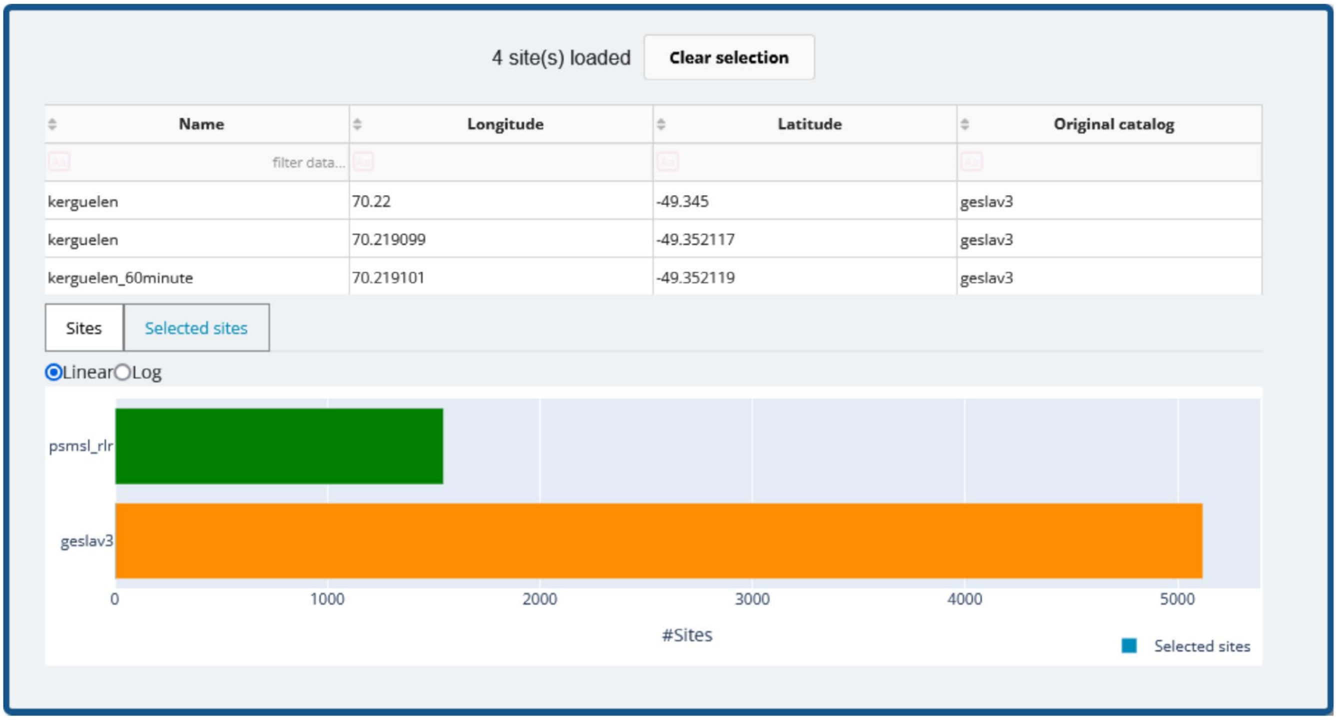

FIGURE 6 | Zoom on the statistic panel of the dashboard.

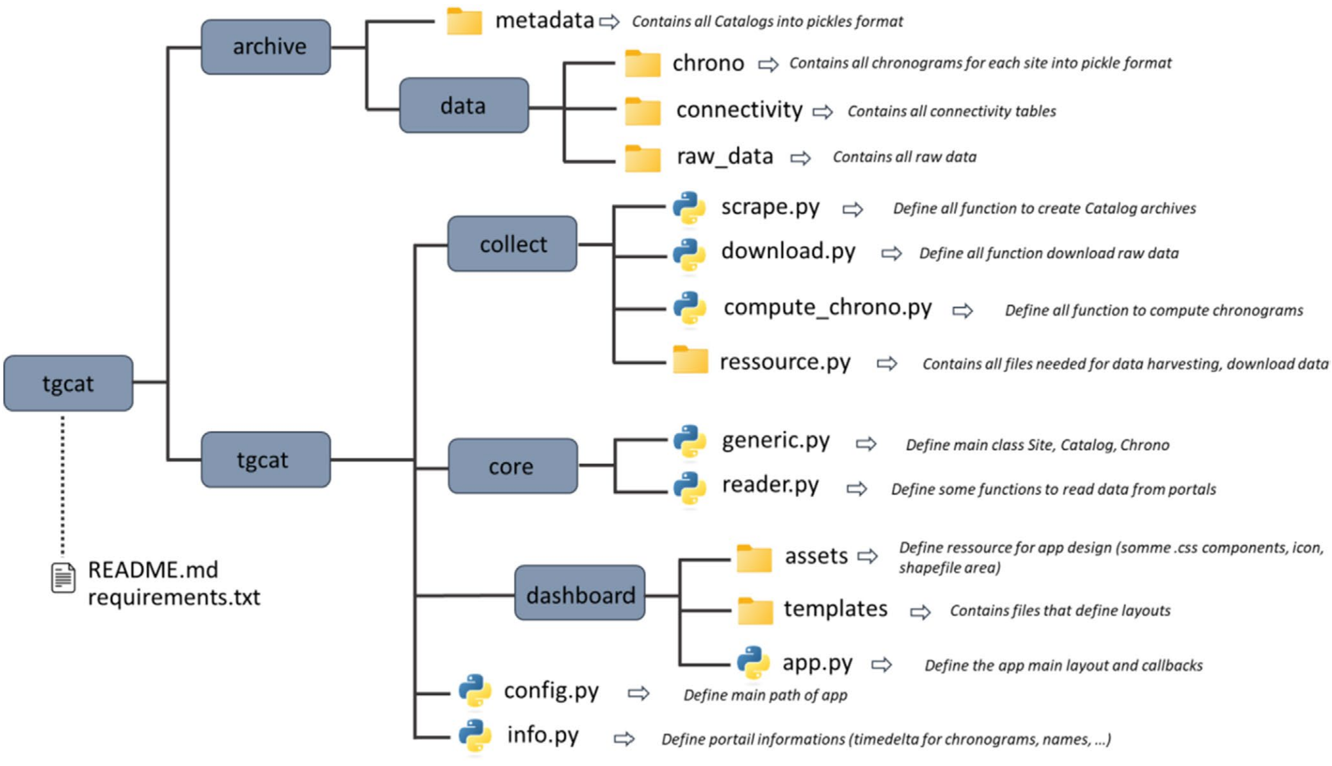

FIGURE 7 | Schematic representation of the architecture of the distributed code.

better view the selected sites is to switch off the catalogue source "toggle switch" on the top right of the page. In that case, all sites on the map turn to grey except the selected ones that now appear in blue. It's important to note that users can easily implement their own region of interest.

2. By country. This selection is based on the union of world country boundaries and Exclusive Economic Zone EEZs (IMIS, VLIZ, coarse Flanders Marine Institute 2020).
3. By name. During the catalogues loading, all site names are stored and can be accessed through this filter. Users should

be aware that one site can have different naming depending on the catalogue, so multiple selection is available with this filter.

5 | Code Architecture

The code is publicly available on the gitlab of the La Rochelle University (https://gitlab.univ-lr.fr/sonel-public/tgcat). The architecture of the code is shown in Figure 7.

6 | Conclusion

TGCAT is a free, open-source tool designed to facilitate the discovery, comparison, and referencing of tide gauge data across multiple global and national portals. By harmonising metadata and visualising data availability through a geospatial indexing approach (H3), TGCAT supports both end-users and data providers in navigating the increasingly complex ecosystem of sea level observations. At the moment 15 data catalogues are available on the online version as well as 6 metadata catalogues. This amount for more than 19,000 sites displayable on the web site. The chronograms available spans several centuries. In the case of the Brest station from 1711 to 2025. Because the raw data are not stored on the server side, but only the chronograms and the connectivity information the amount of data on the server is around 250 Mb.

For researchers and data users, TGCAT offers an efficient way to identify which stations and datasets are available across different repositories, helping to locate the most appropriate data for specific applications—whether for long-term climate studies, operational oceanography, or coastal risk assessments.

For data providers, TGCAT serves as a quality control tool. It allows them to verify whether their stations are properly referenced, to detect inconsistencies in location or naming across platforms, and to assess integration into key global data centres. TGCAT can also be a platform to host historical tide gauge information that has not yet been digitised or shared—opening the door for the community to build a more complete and coherent inventory of sea level observations.

Looking ahead, we plan to expand TGCAT by incorporating more national datasets and welcoming contributions from providers who wish to include their catalogues.

Acknowledgements

This project has received funding from the EuroSea European Union's Horizon 2020 research and innovation program under grant agreement No. 862626. We thank Philip Woodworth for providing the first version of the GESLA dataset (named GESLA1 in the dashboard).

Funding

This work was supported by Horizon 2020 Framework Programme (862626).

Data Availability Statement

There is no dataset associated with the paper because it is a code and a dashboard. The source code is publicly available at https://gitlab.univ-lr.fr/sonel-public/tgcat and the online dashboard can be accessed at www.sonel.org/tgcat.

References

Flanders Marine Institute. 2020. "Union of the ESRI Country Shapefile and the Exclusive Economic Zones (Version 3)." https://www.marineregions.org/. https://doi.org/10.14284/403.

IOC. 1990. "Global Sea Level Observing System (GLOSS): Implementation Plan 1990." Implementation Plan 35. IOC. Technical Series. Intergovernmental Oceanographic Commission. https://unesdoc.unesco.org/ark:/48223/pf0000087839.

IOC. 1997. "Global Sea Level Observing System (GLOSS) Implementation Plan, 1997." https://unesdoc.unesco.org/ark:/48223/pf0000112650.

IOC. 2012. "Global Sea Level Observing System (GLOSS): Implementation Plan 2012." JCOMM Technical Report Series N°066 194. GOOS Report/IOC Technical Series 100. Intergovernmental Oceanographic Commission. https://unesdoc.unesco.org/ark:/48223/pf0000217832.

Lin-Ye, J., B. Pérez Gómez, A. Gallardo, et al. 2023. "Delayed-Mode Reprocessing of In Situ Sea Level Data for the Copernicus Marine Service." *Ocean Science* 19, no. 6: 1743–1751. https://doi.org/10.5194/os-19-1743-2023.

Pérez Gomez, B., L. Testut, A. Hibbert, et al. 2021. "New Tide Gauge Data Flow Strategy." Report. EuroSea. https://doi.org/10.3289/eurosea_d3.3.

Uher, V., P. Gajdoš, V. Snášel, Y.-C. Lai, and M. Radecký. 2019. "Hierarchical Hexagonal Clustering and Indexing." *Symmetry* 11, no. 6: 731. https://doi.org/10.3390/sym11060731.

Woodworth, P. L., J. R. Hunter, M. Marcos, P. Caldwell, M. Menéndez, and I. Haigh. 2016. "Towards a Global Higher-Frequency Sea Level Dataset." *Geoscience Data Journal* 3, no. 2: 50–59. https://doi.org/10.1002/gdj3.42.

Zemunik, P., J. Šepić, H. Pellikka, L. Ćatipović, and I. Vilibić. 2021. "Minute Sea-Level Analysis (MISELA): A High-Frequency Sea-Level Analysis Global Dataset." *Earth System Science Data* 13, no. 8: 4121–4132. https://doi.org/10.5194/essd-13-4121-2021.

Appendix A

TABLE A1 | List of data portals implemented in the TGCAT software.

Catalogue name	Description	#stations	Type	Region	Time (s)
psmsl_rlr	Permanent Service for Mean Sea Level	1551	Data	GLO	0.72
psmsl_metric	Permanent Service for Mean Sea Level	2286	Data	GLO	0.89
psmsl_gnssir	GNSS Interferometric Reflectometry from PSMSL	328	Data	GLO	0.37
uhslc_fd	University of Hawaii Sea Level Center—Fast Delivery data	313	Data	GLO	2.71
uhslc_rq	University of Hawaii Sea Level Center—Research Quality data	691	Data	GLO	2.50
slsmf	Sea Level Station Monitoring Facilities	1637	Data	GLO	4.81
bodc	British Oceanography Data Centre	216	Data	GLO	1.44
cmems_repro	Copernicus Marine In Situ (Reprocessed product)	2155	Data	GLO	31.61
cmems	Copernicus Marine In Situ TAC Near RealTime	0	Data	GLO	0.00
emodnet	European Marine Observation and Data Network	3945	Data	GLO	5.72
misela	Minute Sea-Level Analysis	331	Data	GLO	2.46
geslav1	Global Extreme Sea Level Analysis	675	Data	GLO	79.18
geslav2	Global Extreme Sea Level Analysis	1276	Data	GLO	1.21
geslapv2	Global Extreme Sea Level Analysis	77	Data	GLO	0.10
geslav3	Global Extreme Sea Level Analysis	5119	Data	GLO	0.82
sonel_db	SONEL Tide Gauge entries in database	1249	Meta	FRA	0.63
gloss	Global Sea Level Observing System	294	Meta	GLO	0.55
ssc	IOC/Unesco Sea Level Station Catalog	2607	Meta	GLO	13.42
psmsl_links	Tide gauge linked to a geocentric ellipsoid from PSMSL	188	Meta	GLO	0.59
eutgn	European and adjacent areas Tide Gauge Network Inventory	531	Meta	GLO	0.22
ptwc	Pacific Tsunami Warning Center	0	Meta	GLO	0.00
indonesian	Indonesian Tide Gauge Inventory	277	Meta	IND	1.86
bandaid	Band AID Belmont Forum Project	0	National	BGD	0.04
redmar	Spanish Tide Gauge Network (Puertos del Estado)	42	National	ESP	0.01
refmar	Reference Network of French Tide Gauges	162	National	FRA	62.38
shom_rescue	Inventory of historical data at Shom archive	228	National	FRA	5.92
sonel	Observing System of the Coastal Sea Level	91	National	FRA	3.80
norwtgn	Norwegian Hydrographic Service Tide Gauge Network	30	National	NOR	0.67
noaa	US National Oceanic and Atmospheric Administration	295	National	USA	1.97

Note: The #stations represent the number of stations scrapped for this catalogue. The time column indicates the wall time in seconds taken by the function to collect the metadata information from the website. The web link used to collect the information for each catalogue is shown in the web application (www.sonel.org/tgcat) in the menu for the selection of the metadata and data catalogue.