Project Information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	**EuroSea**
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D3.3
Deliverable title	**New Tide Gauge Data Flow Strategy**
Description	New Tide Gauge Data Flow Strategy
Work Package number	WP3
Work Package title	Network Integration and Improvements
Lead beneficiary	Puertos del Estado
Lead authors	Begoña Pérez Gómez, Laurent Testut, Angela Hibbert, Andrew Matthews
Contributors	Elizabeth Bradshaw, Guy Westbrook, Guy Wöppelmann, Francisco Hernández, Sylvie Pouliquen, Antonio Novellino, Marta de Alfonso, Stephane Tarot, Fernando Manzano, Marta Marcos, Bart Vanhoorne, Mathieu Belbeoch, Long Jiang, Silvia Chacón Barrantes, Bernardo Aliaga, Claire Fraboul
Due date	31/12/2020
Submission date	26/03/2021
Comments	Delayed due to COVID-19 impact on topic related first workshop

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626.

https://doi.org/10.3289/eurosea_d3.3

Table of contents

1. Introduction to EuroSea

Although the Ocean is a fundamental part of the global system providing a wealth of resources, there are fundamental gaps in ocean observing and forecasting systems, limiting our capacity in Europe to sustainably manage the ocean and its resources. Ocean observing is “big science” and cannot be solved by individual nations; it is necessary to ensure high-level integration for coordinated observations of the ocean that can be sustained in the long term. For Europe, EuroSea will point the way for the current and future cooperation between science and industry, politics and the public with the common goal of a sustainable blue economy and the responsible handling of the sensitive marine ecosystems. The project will make a significant contribution to not only generating, processing and linking information about our oceans, but also to make long-term and extensive use of this and the resulting knowledge in a wide variety of areas. As a link between sectors and disciplines, EuroSea faces a very big challenge.

2. Rationale for improving existing tide gauge data flow

Coastal sea level data from tide gauges provide information on oceanographic processes at different spatial and temporal scales, allowing its use for a wide range of applications: research (climate change, ocean circulation, extremes), sea level related hazards (storm surges, meteotsunami and tsunami warning systems), tidal predictions, port operations or datum definitions for national or state boundaries.

The first tide gauges were installed at the main harbours of the world in the XIXth century. Since then, most of the well-developed countries have created and maintained national or local networks that have led to the hundreds of stations available nowadays. Their use for scientific studies on climate change was enhanced thanks to the implementation in 1985 of the Global Sea Level Observing System (GLOSS, https://www.gloss-sealevel.org/). GLOSS has promoted the establishment of a global tide gauge network and have published several manuals on best practices through the last decades, about type of sensor, accuracy and maintenance requirements, or quality control and data processing (see section 4.1). Several GLOSS data portals compile tide gauge data and provide final products to end-users since then. In recent years, GLOSS recommendations include the multi-purpose approach and use of real or near-real time data for storm surge and tsunami warning systems.

Only in Europe and adjacent seas, there are today more than 600 tide gauges operated by different types of national, regional and local institutions: ports, hydrographic offices, geodetic and oceanographic institutions, meteorological services, etc. As in other parts of the world, most of these stations are today transmitting data in near-real time, what has become important for the strong operational oceanography community in the region. These data are today aggregated and distributed by new international programs such as Copernicus or the European Marine Observation and Data network (EMODnet) - Physics thematic lot -, along with many other types of variables, from in-situ stations, remote sensing techniques and circulation numerical models.

The EuroGOOS Tide Gauge Task Team was established in 2015 in order to bring together the European and adjacent seas tide gauge community. The initiative aimed to foster collaboration by enhancing countries capacity and the development of the European Ocean Observing System (EOOS), under the umbrella of GLOSS, the Regional Operational Oceanographic Systems (ROOS’s) in EuroGOOS and the new international programs of data exchange such as CMEMS, EMODnet or SeaDataNet. This task team launched a questionnaire in 2016, that revealed than more than half of the institutions and near 30% of the tide gauge stations in the region would be facing problems of funding. Fulfilling adequate end-users' requirements is important for the sustainability of the system. Therefore, since 2018 the group has focused on assessing the access to tide gauge data and derived products through existing data portals. A first analysis has revealed the existence of gaps and duplications that confirm the importance of increasing the coordination between mentioned data aggregators and data providers. The objective of this document is to propose several recommendations and actions that could help on that direction.

3. Background

3.1. WP3 Initiatives: EuroSea Data Management Plan

The EuroSea Data Management Plan (DMP) has been produced by WP3 task 3.10 as part of the Open Research Data Pilot that EuroSea complies with, setting the framework for the handling of data produced in EuroSea from acquisition over curation to dissemination (Deliverable 3.1). Built up on existing observing networks and already established best practices, this DMP deals with harmonizing and improving data management procedures and implementing FAIR principles to serve Copernicus Monitoring Environment Marine Service (CMEMS), EMODnet, SeaDataNet and historical National Oceanographic Data Centres at a later stage. The primary data providers in EuroSea are existing observing Networks as described in WP3 that have been established before the start of EuroSea mostly under EuroGOOS Task Team activities.

Based on EuroGOOS and OCG (Observations Coordination Group) data recommendations, this system of systems harmonizes work-flows, data processing, quality assurance procedures, and distribution across in-situ observing network systems in EuroSea WP3, WP5, WP6 and WP7 and integrates in-situ observations into existing European and international data infrastructures, in EuroSea termed "data Integrators".

The driving principles within EuroSEA is to foster the implementation of recommendations of the OceanObs19 white paper [https://doi.org/10.3389/fmars.2019.00440.]:

1. **Findable:** Each dataset should be identified by a unique persistent identifier and described by rich, standardized metadata that clearly include the persistent identifier. The metadata record should be indexed in a catalogue and carried with the data.

2. **Accessible:** The dataset and its metadata record should be retrievable by using the persistent identifier and a standardized communications protocol. In turn, that protocol should allow for authentication and authorization, where necessary. All metadata records should remain accessible even when the datasets they describe are not easily accessible.

3. **Interoperable:** Both metadata and datasets use formal, accessible, shared, and broadly applicable vocabularies and/or ontologies to describe themselves. They should also use vocabularies that follow FAIR principles and provide qualified references to other relevant metadata and data. Importantly, the data and metadata should be machine accessible and parsable.

4. **Reusable:** data must already be findable, accessible, and interoperable. Additionally, data and metadata should be sufficiently richly described that it can be readily integrated with other data sources. Published data objects should contain enough information on their provenance to enable them to be properly cited and should meet domain-relevant community standards.

The New Tide Gauge Data Flow Strategy presented in this report is therefore the contribution to EuroSea DMP of the European Tide Gauge Network and the EuroGOOS Tide Gauge Task Team, collecting the voice of European Data Integrators Infrastructure (EMODnet Physics, CMEMS INSTAC, SeaDataNet), and following GLOSS existing requirements and needs in the region.

3.2. AtlantOS Project and EuroGOOS DATAMEQ initiatives

Well-founded data management systems are of vital importance for the Ocean Observing system as they ensure that essential data are not only collected but also retained and made accessible for analysis and application for current and future users. Data management systems that facilitate access, use and interpretation of data and products should be included as essential elements of the Observing System. Effective data management is based on collaboration across activities including observations, metadata and data assembly, quality assurance and control (QA/QC), and data publication that enables local and interoperable discovery and access, and secure archiving that guarantees long-term preservation.

The existence of a multitude of disparate data management infrastructures currently imposes problems for the observing system that include delayed and duplicate data receipts, versioning issues, missing data and metadata, and non-documented data processing procedures. Therefore, modern data management infrastructures are needed so that all activities along the data flow pipeline are tracked. Progressively the systems should be advanced towards interoperability; this serves both the routine data exchanges within and between the observation networks, as well as user-friendly tools for data/products discovery viewing and access. Community standards for metadata, data formats, communication protocols, and data server software infrastructure are the foundation for interoperability.

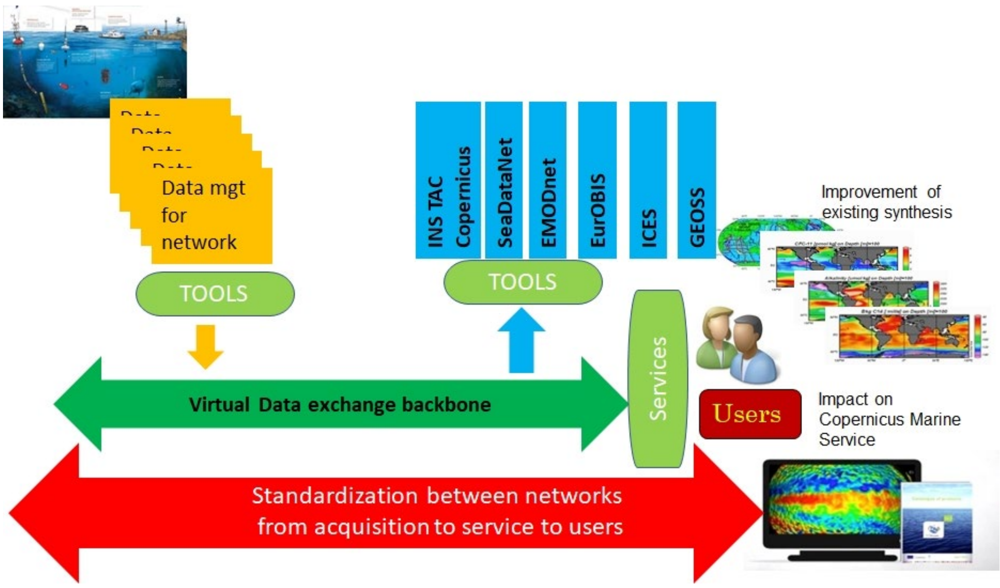

Figure 1: The integrated European data system: in yellow the observing network systems, in blue the integrators and in green and red the harmonization and integration elements.

An important driver within AtlantOS has been to ensure that data from different and diverse in-situ observing networks are readily accessible and usable to the wider community, including international ocean science communities and other stakeholders in this field. The AtlantOS strategy proposes to move towards an integrated data system (Figure 1) that harmonizes work flows, data processing and distribution across the in-situ observing network systems in the Atlantic Ocean to existing European and international data infrastructures and Integrators portals (e.g. Copernicus In Situ Thematic Assembly Centre (INS TAC), SeaDataNet network of national oceanographic data centres (NODCs), EMODnet, ICES, EurOBIS, GEOSS).

The actors for such an integrated system are overall mature systems with long-term experience and established procedures for data collection and management often agreed at international level to collaborate. Consequently, trying to implement a sovereign and rigid set of rules for all the actors to comply with, would be highly challenging and not in the best interest of Atlantic Ocean observing community and product users. In agreement with the EuroGOOS "DATA Management Exchange and Quality" working group recommendations (DATAMEQ), the AtlantOS community recommended to:

1. Rely on existing European and international standards and protocols, first **focussing on metadata** by implementing a set of mandatory information and using agreed vocabularies at all level of the processing chain (Common Vocabulary for parameters, Common Unique ID for Platform and code for Institution, etc.) that allow easier traceability of the observations within the processing and distribution data flow.
2. Encourage **open and free** data policy.
3. **Focus on data quality** by implementing a set of common near real time QC procedures for 7 Essential Variables (Temperature, Salinity, Current, Sea-Level, Oxygen, Chlorophyll, Nitrate and Carbon) acquired in near real time (distributed within a few hours to days).
4. **Enhance access to network data by setting up a unique entry point to discover and download** existing data, either by integrating the data in existing international network Global Data Centres (EGO for gliders, OceanSites for fix point platforms and transport array, ICOS-Marine for some VOS and GO-SHIP data), or by setting up new ones like the new GDAC for drifters endorsed by DBCP/GOOS
5. **Connect to existing integrators**.
6. **Enhance monitoring facilities** offered by OceanOPS.

Relying on existing infrastructure that will last after the end of the AtlantOS project, AtlantOS has moved forward on the implementation of the FAIR principles (Findable Accessible Interoperable Re-Usable) for Atlantic observations, through a system of data management systems where all ocean observations are made available to users on a free and unrestricted basis, ensuring full and open exchange of data, metadata and products at minimum time delay.

Such principles have been published in partnership with international partners at the OceanOBS19 conference in the paper Tanhua & al "Ocean FAIR Data Services" (Front. Mar. Sci. 6:440.,doi: https://doi.org/10.3389/fmars.2019.00440) and if implemented, even partially, would significantly enhance access to existing tide gages data and derived products.

4. International programs and data portals providing tide gauge data

A description of the main international programs and data portals with tide gauge data and/or products is presented in this section. National data portals are not included in this report, although they exist in many countries and serve a large number of usually national and local users. The different data portals presented below deal with different temporal sampling and latencies, as well as quality control and processing, depending on their main application.

It is important to emphasize that since the tsunami of December 2004, most of the tide gauges in the world have been progressively upgraded to lower temporal sampling (< 1min), and real time data transmission has become an important requirement. This has been a challenge for existing tide gauge data flow and processing, but today these new data are a key new source of information for extreme sea levels research.

Real time raw data from tide gauges are used by the National Tsunami Services, and by the Intergovernmental Coordination Groups and regional warning systems established by UNESCO/IOC for the Tsunami and Other Coastal Hazards Warning in the Pacific (ICG/PTWS), Indian Ocean (ICG/IOTWMS), Caribbean Sea (ICG/CARIBE-EWS) and North-East Atlantic and Mediterranean region (ICG/NEAMTWS). A list of Tsunami Service Providers currently operating under the IOC-UNESCO framework, can be found at: http://www.ioc-tsunami.org/index.php?option=com_content&view=article&id=426&Itemid=368&lang=en. In this report, we will consider Tsunami Services as users of the tide gauge network that report sea level readings as TSUNAMI OBSERVATIONS, usually based on Decision Support Systems (DSS) that ingest directly data from different sources.

4.1. The Global Sea Level Observing System (GLOSS)

The Global Sea Level Observing System (GLOSS, https://www.gloss-sealevel.org/) is an international program established by the UNESCO Intergovernmental Oceanographic Commission (IOC) in 1985. Its aim is to implement a well-designed, high-quality sea level observing network to support scientists and operational applications, as a component of the Global Ocean Observing System (GOOS) (see last GLOSS Implementation Plan: https://unesdoc.unesco.org/ark:/48223/pf0000217832).

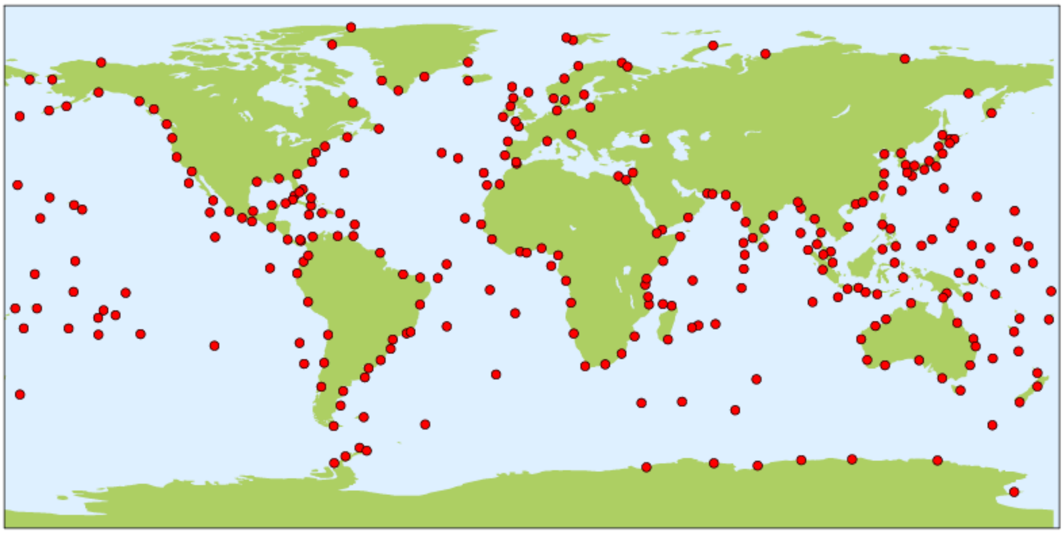

Figure 2: GLOSS Global Core Network

The GLOSS 'Global Core Network' (GCN) is composed of approximately 300 evenly-distributed sea level stations around the world for long-term climate change and oceanographic sea level monitoring (Figure 2). The success of GLOSS depends on the voluntary participation of countries and national bodies, that provide the financial resources for operation and maintenance of the networks, with the support and coordination of the scientific and oceanographic community gathered by the GLOSS program. As defined in the last GLOSS Implementation Plan, several GLOSS sea level data centres were established to facilitate the smooth flow of sea level data and associated metadata to national and international data centres, scientists and operational users:

- **Real time:** IOC Sea Level Station Monitoring Facility: Flanders Marine Institute (IOC/SLSMF) (VLIZ, Belgium) - focused on real time data transmission with low latency providing essential support for tsunami monitoring, and indicating which sites are active worldwide (more than 1000 stations, Figure 3). Data is not quality controlled and not intended for research purposes.
- **Fast Delivery Centre** (fast higher frequency data): University of Hawaii Sea Level Center (UHSLC, USA) - hourly and daily resolution data provided on a 1 month lag with basic quality control. Also provide a research quality dataset on a 1 year lag. Operates 73 tide gauges worldwide, many in remote locations.
- **Delayed mode** high frequency data and ancillary variables: British Oceanographic Data Centre (BODC, UK) - responsible for assembling and distributing the final version of hourly and sub-hourly data from GLOSS sites, along with metadata, and other variables such as meteorological data.
- **Mean sea levels**: Permanent Service for Mean Sea Level (PSMSL), Mean Sea Level (MSL) data: National Oceanographic Centre (NOC, UK) - monthly and annual mean data for monitoring long term changes in sea level, also provides estimates of relative mean sea level trends
- **GNSS at Tide Gauges:** GNSS at Tide Gauge Data Centre (SONEL), vertical land movement information: University of La Rochelle and CNRS/INSU (France) - collates land movement information measured at GNSS sites worldwide and information on geodetic ties between GNSS receivers and nearby tide gauges. Distributes GNSS solutions from a number of centres, and combines these with PSMSL sea level trends to produce geocentric sea level trends.
- GESLA: The GESLA (Global **Extreme Sea Level** Analysis) initiative stems from the scientific interest and need to explore the magnitude and frequency of coastal extreme sea levels and their changes. This is done by collecting and formatting global tide gauge observations from different sources and data providers. The first and second GESLA datasets were released in 2009 and 2016, respectively, while a third update and extended version is nearly to be published within 2021. The second version is currently freely available through the GESLA website (www.gesla.org) and contains 1355 records and 39151 station-years, globally distributed. The scientific relevance of GESLA data set is reflected in more than 50 publications so far.

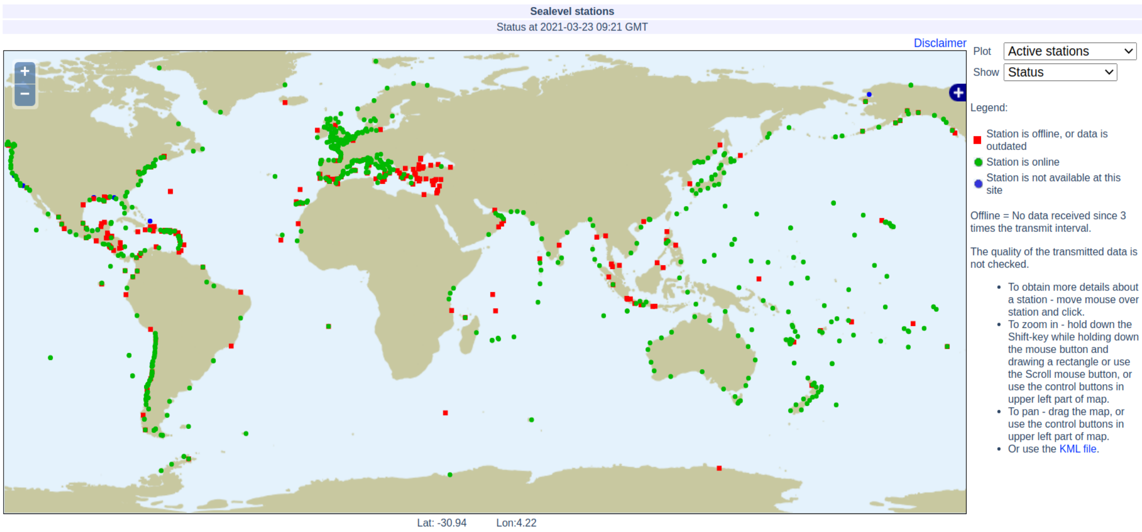

Figure 3: GLOSS IOC/SLSMF: >1000 stations contributing worldwide with raw (lower temporal sampling) data in real time.

4.2. CMEMS In Situ TAC

The Copernicus Marine Environment Monitoring Service (CMEMS) (https://marine.copernicus.eu/), led by Mercator Ocean International, is based on a distributed model of service production, relying on the expertise of a wide network of participating European organisations involved in **operational oceanography**.

The Service encompasses two kinds of production centres:

1. Monitoring and Forecasting Centres (MFC's), charged with maintaining numerical models of the ocean. There are seven MFCs: six for regional seas and one for the global ocean
2. Thematic Assembly Centres (TAC), which are tasked with the collection of ocean observations, both in situ (water column) and satellite observations.

The In Situ TAC (INS TAC) is the component of the Copernicus Marine Service which ensures a consistent and reliable access to a range of in situ data **for the purpose of service production and validation**. The mission of the INS TAC is to provide integrated in situ products built from in situ observations acquired from outside CMEMS data providers to fit the needs of Copernicus Marine Service internal and external users. It is a distributed centre composed of 6 regional centres working closely with the EuroGOOS ROOS's (Regional Operational Observing Systems) and a Global centre well connected to the JCOMM networks (Joint WMO-IOC Commission for Oceanography and Marine Meteorology). INS TAC also works closely with the SeaDataNet infrastructure, that coordinates a network of European National Data Centres which archive observation data acquired within scientific campaign and EMODnet, which is a network of organisations supported by the EU's integrated maritime policy.

In Situ TAC has two main objectives:

1. To collect multi-source, multi-platform, heterogeneous data, perform consistent quality control and distribute it in a common format (NetCDF) and in near-real-time (within 24 hours) to the CMEMS Marine Forecasting Centres (MFC), for assimilation into their numerical ocean models.
2. To supply the MFCs and downstream users with re-processed 25-50-year products in delayed mode. In addition to the near-real-time products, these delayed-mode products are useful for model validation or assimilation in ocean reanalysis and climate studies.

Regarding sea level measurements, in 2017 INS TAC started a fruitful collaboration with the EuroGOOS Tide Gauges Task Team that led to a “Recommendations for Copernicus Marine Service (CMEMS) on standard NetCdf format for tide gauge data” document which was implemented at INS TAC on 2018 (NetCdf_Recommendations_forCMEMS_EuroGOOSTGTT_October_2017.pdf). A Near Real Time Quality Control (NRTQC) package has been developed by Puertos del Estado based on GLOSS quality control procedures and it has been distributed to INS TAC through git. This NRTQC is being applied in IBI and MED regions.

Sea level data are gathered from providers in real time and distributed in a homogeneous NetCDf format through the CMEMS catalogue. More than 400 tide gauges are included, mainly in European Seas (Figure 4). Presently, there is no reprocessed sea level product, but it is planned for the next CMEMS phase (starting on 2022).

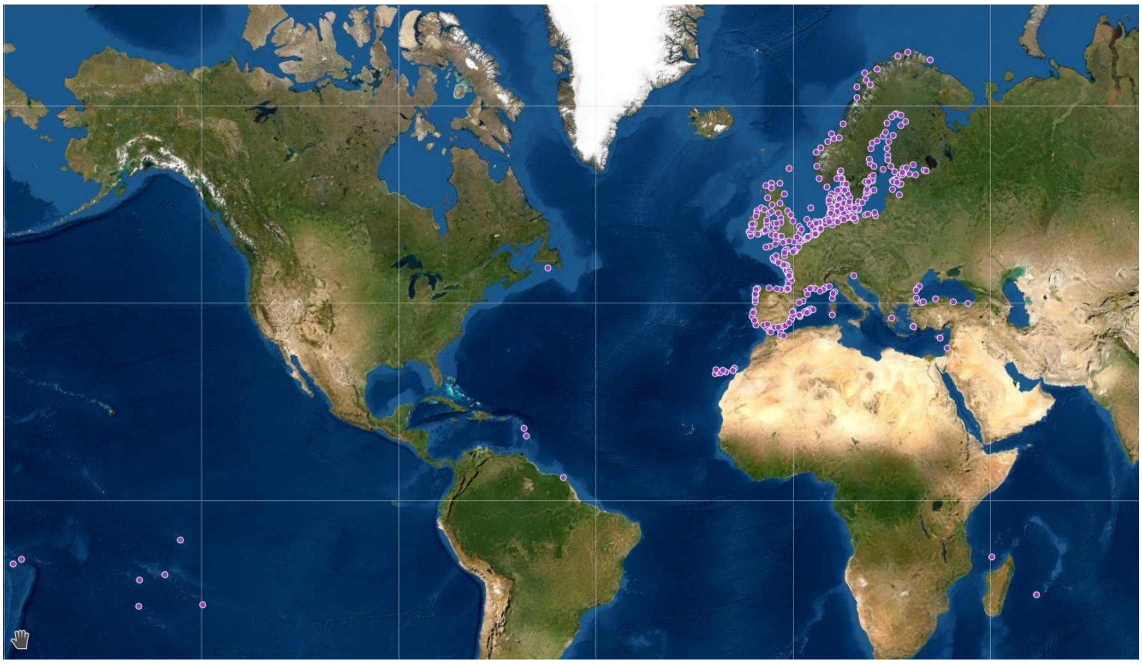

Figure 4: Sea level stations managed by the CMEMS In Situ TAC (February 2021).

4.3. SeaDataNet

SeaDataNet (SDN) is a pan-European infrastructure for managing, indexing and providing access to marine data sets and data products, acquired by European organizations from research cruises and other observational activities in European coastal marine waters, regional seas and the global ocean. Founding partners are National Oceanographic Data Centres (NODCs), major marine research institutes, UNESCO-IOC, and ICES. The SeaDataNet network has developed its network of data centres and infrastructure with standards, tools, and services during many EU projects (SeaDataNet, SeaDataNet2, Ocean Data Interoperability Platform, and SeaDataCloud). SeaDataNet develops, governs and promotes common standards, vocabularies, software tools, and services for marine data management, which are freely available from its portal and widely adopted and used, e.g. most of the EMODnet thematics projects largely adopted and support the SDN standards.

SDN common standards for the marine domain have been developed and are maintained, collaborating with European and international experts, adopting and adapting ISO and OGC standards, and achieving INSPIRE compliance, where possible. These standards comprise:

- INSPIRE compliant marine metadata profiles of the ISO 19115 – 19139 standards for data sets (CDI = Common Data Index) and research cruises (CSR= Cruise Summary Reports);
- marine metadata formats for data collections (EDMED = European Directory of Marine Environmental Data), research projects (EDMERP = European Directory of Marine Environmental Research Projects), monitoring programs and networks (EDIOS = European Directory of Initial Ocean-observing Systems), and organizations (EDMO = European Directory of Marine Organizations);
- SDN controlled vocabularies for the marine domain, with international governance, user interfaces and web services. These follow the W3C SKOS specification for encoding the data dictionaries and taxonomies served. Vocabularies for parameters, units, platforms, etc are maintained by BODC (NERC Vocabulary Service - NVS)
- standard data exchange formats (SDN ODV ASCII, SDN NetCDF with CF compliance) as applied for download services. These interact with other SDN standards such as the Vocabularies and Quality Flag Scale;
- standard QA-QC procedures for various data types, together with IOC-IODE and ICES.

SeaDataNet has focused, with success, on applying these standards for interconnecting data centres enabling the provision of integrated online access to comprehensive sets of multi-disciplinary, in situ and remote sensing marine data, metadata and products.

4.4. Tsunami Alert Device (TAD): JRC Data Server

The European Commission Joint Research Centre at ISPRA (JRC) holds a sea level data portal and web interface for tsunami applications and alerts TAD, based on real time tide gauge data from different sources and institutions, including those from the inexpensive and autonomous tide gauges (IDSL) specifically designed and installed by JRC in different countries for this application.

4.5. EMODnet Physics

EMODnet Physics is one of the seven domain-specific projects that through the effort done under the ur-EMODnet preparatory action (MARE/2010/02), and successive development (MARE/2012/10; EASME/EMFF/2016/006) and operational (EASME/2019/OP/0003) phases, have been successful in designing, organising and running operational services where ocean physics data and data products built with common standards can be found, viewed and downloaded in a way that is free of charge and free of restrictions of use.

EMODnet Physics integrates and makes available ocean physics data (Real Time, Near Real Time, Historical reprocessed & validated) and derived data products. Data products originate from institutes or other aggregating infrastructures (e.g. GDAC) and are redistributed, while EMODnet Physics also generates its own data products such as collections of data and elaborations. Moreover, it promotes standards for data, metadata and services. Data scope includes: temperature in the water column, salinity in the water column, wave direction and height, wind direction and intensity, sea currents direction and speed, sea level and trends, optical properties, sea ice, river outflow, acoustic pollution, and atmospheric - meteorological data.

Primary pillars under EMODnet Physics are Copernicus CMEMS-INS TAC and EuroGOOS, both for operational data, and SeaDataNet for archived and validated data. These European integrators bundle and make available data originating from many oceanographic observing networks and activities. In addition, EMODnet Physics derives data and data products from other European and international providers, such as ICES, GLOSS (PSMSL, SONEL, IOC-SLSMF), TAD, GDAC (Coriolis), OCEAN SITES, etc. It links all these ocean data sources into a single discoverable database in which metadata are harmonized, for instance, all involved organisations are described and coded in the SeaDataNet directory for marine organisations (EDMO), while vocabularies exist for parameters, sea regions, platforms, instruments, etc.

Data are harmonized in formats, where possible. The data flows and processing result in common metadata, data files for time series, data files for gridded products, maps and map layers, and catalogues. Depending on the data age (from Real Time to archived), different technologies and formats are used for retrieving data and products from the mentioned sources. And different levels of QA-QC might have been applied on the data, depending on the data age.

Data policy is to be open and free as much as possible. The distribution of data and data products is done by different technical services, using GeoServer, THREDDS, and ERDDAP. The EMODnet Physics portal provides functionalities for discovery, access and viewing of data and products in maps and graphics for time series and trends.

As anticipated, EMODnet Physics is targeting the Sea Level data by integrating more than 400 European tide gauge stations, the 290 Global Sea Level Observing Systems (GLOSS) core network, and more than 1.300 stations contributing MSL data to GLOSS-PSMSL. EMODnet Physics makes available a relative sea level trend product and a sea level anomalies product based on the PSMSL collection, and an absolute sea level trend product based on SONEL. In the PSMSL product, the MSL trends measured by tide gauges are local relative (RLR) MSL trends, as opposed to the global sea level trend. These trends are not corrected for land changes. Tide gauge stations measure instead local sea level, which refers to the height of the water as measured along the coast relative to a specific point on land. The absolute sea level is processed by using the geodetic data from the GNSS stations and SONEL serves as the GNSS data assembly centre for GLOSS. Other tide gauge datasets already federated into EMODnet Physics are the GLOSS Fast Delivery Tide Gauge Data from the UHSLC and the Tsunami Alert Device JRC Data Server.

5. Analysis of gaps and duplicities on tide gauge data portals

At the present time many portals are providing information on tide gauge data. The present status and simplified data flow is shown in Figure 5. It is important for the end-users and for the scientific community to know where to search for the relevant sea level data depending on the targeted applications. It is also critical for tide gauge data portals (providing sea level records and products) or data catalogues (referencing information and metadata on tide gauges) to know if they are up to date.

In this section, we have done an inter-comparison of 13 data portals or catalogues to estimate where are the duplicates and the gaps in terms of sea level information. This is a preliminary analysis focused only on the number and identification of stations in each data portal. A more comprehensive study should be accomplished in the future to assess other aspects such as duplicities on quality control, different versions of data and data products, etc. This will be briefly described and discussed in section 6.4.

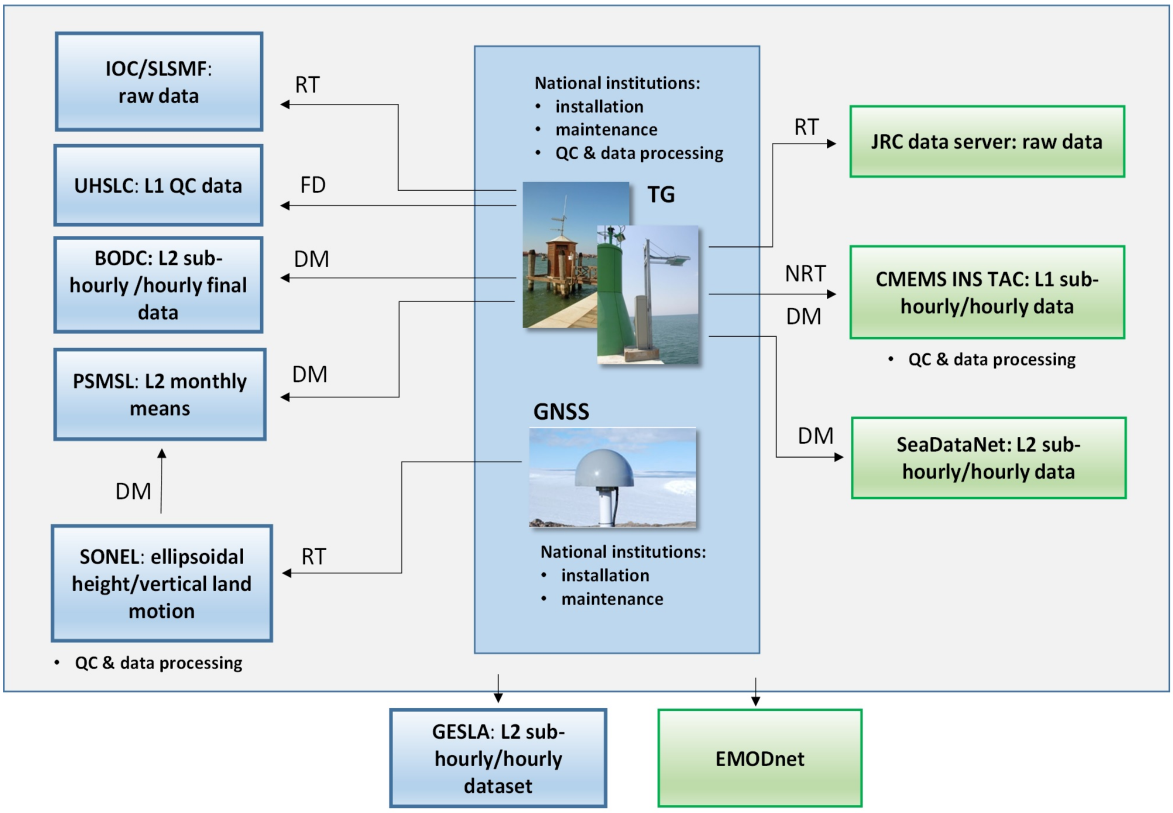

Figure 5: data portals, datasets and data flow of tide gauges and associated GNSS data at the time of drafting this report. GLOSS official data portals (only sea level data) in blue, European data portals (for several oceanographic variables: CMEMS INS TAC and EMODnet, or for sea level: JRC Data Server) in green. QC stands for Quality Control (L1, preliminary, L2, final), RT: Real Time, NRT: Near-Real Time (>15 min), DM: Delayed Mode. Applications and individual descriptions can be found in the text. This is a simplified, but close to reality data flow description (additional links between GLOSS data portals, for some stations; some networks are not directly connected to all GLOSS data portals either, e.g.: RONIM network). Some data portals compile information from other data portals as well as from national data providers, such as GESLA dataset and EMODnet Physics portal.

In addition to all data integrators described in section 4, the most complete and comprehensive metadata information or metadata catalogues are the IOC-GLOSS *Sea level Station Catalog* (SSC), the *GLOSS Core Network* catalogue and the *European Tide Gauge Inventory* (still in development in the framework of EuroSea project, see section 7.1). Table 1 summarizes the number of entries in every data portals or catalogues analysed, at the time of drafting this report.

As can be inferred from section 4, the content on data portals is highly variable in terms of number of stations, temporal sampling, transmission latency, length of the records, quality control and metadata information. Besides, at present, because of uncertainties in geographical coordinates (frequently lacking enough resolution) and of different naming conventions and ID's it is not possible to make an accurate site by site inter-comparison. Table 1 shows the number of entries found, including duplicates, for these reasons.

So in order to make an approximate first estimation of missing and common stations in different data portals and data catalogues, we used a 1x1 degree cells to inter-compare their content. This is summarized in the Figure 6, where we can see that the IOC/SSC is the most complete in terms of number with a reference to 1660 pixel-stations. The data portals PSMSL and EMODnet provide access to more than 1000 pixel-stations.

Code	Name	#data
ssc	*Sealevel Station Catalog*	2364
gcn	*GLOSS Core Network*	290
eutgi	*EuroGOOS Tide Gauge Inventory*	502
psmsl_rlr	Permanent Service for Mean Sea Level (RLR)	1548
psmsl_met	Permanent Service for Mean Sea Level (METRIC)	812
uhslc_rq	University of Hawaii Sea Level Center (Research Quality)	689
uhslc_fd	University of Hawaii Sea Level Center (Fast Delivery)	286
bodc	British Oceanographic Data Center	216
slsmf	Sea Level Station Monitoring Facility	1100
cmems	Copernicus Marine Service	865
emodnet	European Marine Observation and Data Network	3621
gesla_public	Global Extreme Sea Level Analysis	1276
gesla_private	Global Extreme Sea Level Analysis	77

Table 1: Number of stations contained by the main portals or catalogues. In italic are the metadata portals which are not providing any sea level data, but only metadata.

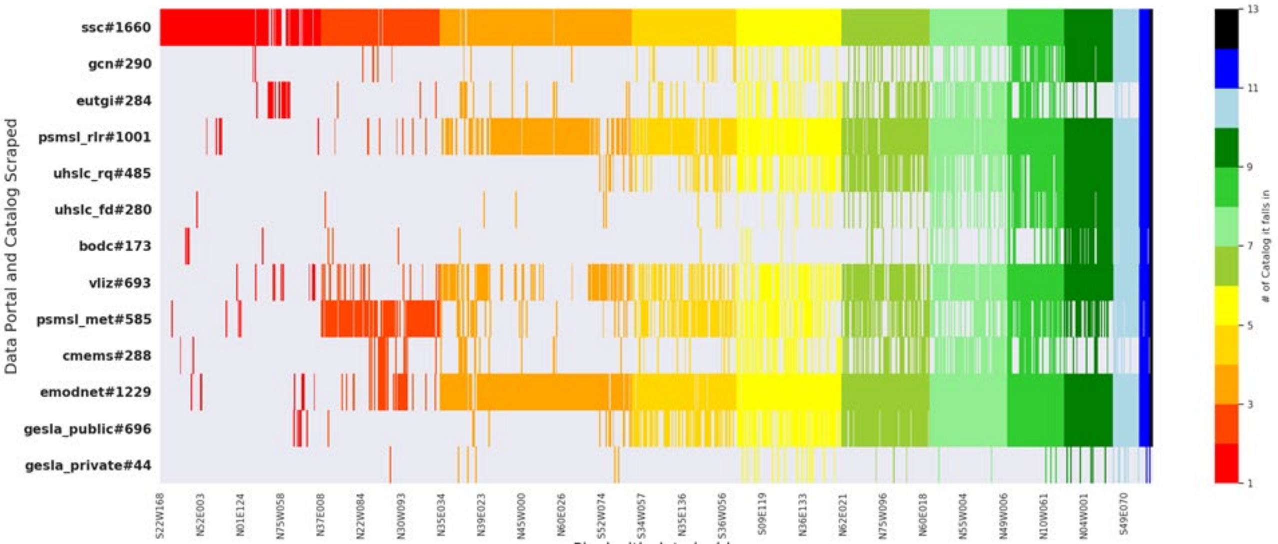

Figure 6: Bar code representation of the content of studied data portals and catalogues. The X-axis represent the coastline at 1 degree pixel where at least one station is present. The Y-axis is the name of the data portals studied (the # represent the total number of pixel where stations are present). The colour bar indicates the number of times a pixel is present in a data portal (red bars if only present in one portal and in black, stations in that pixel are present in them all).

A very important conclusion of this exercise is that, without a unique ID and/or name for the stations, it is nearly impossible to make an accurate inter-comparison between data portals. This is one of the basic recommendations of international programs and projects described in section 3, still not implemented in the tide gauge network. Nevertheless, based on this preliminary analysis we end up with some recommendations aligned with the principles already presented in previous sections for all types of data in oceanography:

- It is necessary to define new unique identifiers in the tide gauge network

- A common and shared definition of what a 'site' or 'station' is, would be mandatory to solve the need of new unique ID's in the tide gauge network
- Every data portal must use a common or well-known vocabulary for variable names and all kinds of metadata information, including country, data provider, etc.
- The use of a common reference global metadata tide gauge catalogue is critical for the community. The most complete metadata catalogue to date is the one maintained by the IOC/SLSMF, the IOC/SSC, that cross links 2369 stations between the different GLOSS repositories (UHSLC, PSMSL, SONEL, IOC/SLSMF). This could be therefore a good starting point for improvements in the global network as well as in any regional initiative.
- If possible every data portal should provide a simple ascii or html table summarizing the content of the portal for their tide gauges, as the one provided by PSMSL: https://www.psmsl.org/data/obtaining/

6. Improvement of existing data flow and enhanced coordination between data portals and data providers

In July 1-2 2020, the EuroGOOS Tide Gauge Task Team organized a meeting with representatives of the different data portals described in section 4. The objective was to take a first step to strengthening their links and cooperation, in order to face the main weaknesses of the tide gauge network in Europe, in terms of data access and exchange. These challenges are in many aspects shared with GLOSS and should therefore be confronted together. One of the main outputs of this meeting was the establishment of the following EuroGOOS TGTT/GLOSS technical subgroups to progress on:

- Review the definition of tide gauge/station/site (led by Guy Wöppelmann)
- New ID definition and assignment (led by Elizabeth Bradshaw)
- Agreement on minimum mandatory metadata and common vocabularies and definition (led by Marta Marcos)

It is important to emphasize that these issues are not new and have been tackled by GLOSS in the past. However, the evolution of the network to fulfil new applications and actors, the increasing demand of tide gauge data for operational purposes and the hundreds of new stations around the globe have brought out the convenience of allocating additional resources and funding, not only for maintenance of the networks by the Member States, but also for the data portals to adapt to the new technical requirements and data flow. This will in the end allow to progress to real inter-operability and, ultimately, to a better service to the users. This effort is also today more important in a context of integration with other observational platforms (e.g. GOOS) and with models and satellite observations (e.g. Copernicus), one of the objectives also of EuroSea project.

The first EuroSea Tide Gauge Network Workshop was held in January 12-14 2021, with more than 160 delegates registered from and outside Europe. Among other topics, the workshop aimed to produce recommendations on improvements of existing data flow and coordination between data portals, with a specific session. The following recommendations with respect to this topic were provided:

- There is a need for consistency of data with the oceanographic community and the global sea level community;
- Contribution from non-European attendees is welcome, especially if their data are only used operationally at present. We want to work closely with data suppliers to develop a clear path for them to follow as they start to build data repositories;
- There is a need to solve data policy issues to access data in several countries (e.g. Greece, Russia, North of Africa).
- The European network consists of hundreds of stations, most used for operational and sea level related hazards applications. Many of these already contribute to CMEMS/EMODnet and GLOSS/IOC SLSMF. However, the abovementioned problems lead to duplicities and gaps between data portals that complicates cross-agencies links and data exchange.
- A possible practical first step would be starting first with only these active operational stations within EuroSea project.
- Importance of progress on the EuroGOOS TGTT – GLOSS established subgroups working on: i) new criteria for site definition; ii) new unique ID assignment; iii) agreement on minimum common mandatory metadata, with common vocabulary. Based on existing work from GLOSS and not reinventing what is already done
- Support from OceanOPS and WMO/WIGOS global initiatives is appreciated. Need to define the scope of this collaboration (European network/GLOSS);
- Communication with data providers should be strengthened, and very important for maintaining up-to-date homogeneous metadata information in the region
- According to what has already been established for the oceanographic community, FAIR principles and traceability are required, including credit to data providers as much as data programs, to help on sustainability of the network
- GLOSS is concerned about the "GLOSS Core Network stations" status and possible needs of funding (38 in Europe)

With respect to data policy/access, Intergovernmental actions to promote data exchange across continents would be needed. As an example: the EMOD-PACE https://www.emodnet.eu/en/emod-pace initiative of DGMARE and Chinese Gov. is starting unlocking access to data and in EMODnet Physics, that have started receiving DM sea level data from 6 Chinese coastal stations.

In what follows, the status of the work started by the three EuroGOOS TGGT/GLOSS technical subgroups, and their recommendations, is presented.

6.1. Review tide gauge station/site definition

The question of what a tide gauge station and/or site is, has become pivotal to build database models and metadata exchange schemes across data centres in an interoperable system. At first sight, it might appear odd that such definitions cannot be found in existing manuals and textbooks (as far as our review extends). However, all the ingredients are actually available, although implicitly, for instance in the IOC Manuals, to establish direct definitions of these "objects". Our task was thus to assemble these ingredients into a comprehensive set of definitions useful in the present framework, and likely beyond.

Three conceptual levels are distinguished, starting with the simplest level of **"tide gauge**", then extending to "**tide gauge station**", and finally dealing with "**tide gauge site**". A "**tide gauge**" is considered in its most basic meaning as the instrument (gauge) that records sea level at a given location. The next level includes available information about the metrological connection between the zero value of the instrument and an external reference. This external reference gives access to the concept of "**tide gauge station**"; it is the so-called tide gauge station datum, a.k.a. tide gauge datum. The tide gauge datum is attached to the land upon which the tide gauge is grounded. It is defined as a vertical distance below the top of a physical benchmark nearby; hence its spatial extension is limited, local. The vertical distance is arbitrarily chosen and documented by the tide gauge operating agency. Since it is an arbitrarily chosen datum, a data centre could define and document its own tide gauge datum or vertical distance from the top of the benchmark.

At this level, an important practical international recommendation is missing. That is, the tide gauge datum should be defined by a set of benchmarks, not only one. The rationale of a local network of benchmarks is to prevent destruction, and thus ensure the datum continuity required to build a long time series at a station. Consequently, we propose to define a "**tide gauge site**" as an extension of the above tide gauge station definition by including additional benchmarks. All the tide gauge benchmarks define the tide gauge datum and serve to ensure its local stability in terms of vertical ground displacements. The geographical extension of the site is thus determined by the set of benchmarks and their relative stability. As long as the set of benchmarks spread further in distance from the tide gauge location, and still show relative stability, the tide gauge station or tide gauge site spatial fingerprint further extends. At its limit, the site is adjacent to another site, which behaves differently in terms of vertical ground motion.

Some comments follow from the above definitions. A GNSS station can be defined in a similar way as a tide gauge station, provided the GNSS antenna reference point is connected to a benchmark. Furthermore, if the benchmark to which the GNSS antenna is connected belongs to the local network of tide gauge benchmarks, the GNSS station can be considered as part of the tide gauge site. This consideration also applies to an additional tide gauge nearby. If both tide gauges share a common benchmark, they can both be considered to belong to the same tide gauge site. In addition, both tide gauges (in the same tide gauge site) experience the same vertical land motion (by definition). One can also expect that both will record the same long-term changes in sea level. However, they may not record the same high-frequency sea levels as these usually show shorter spatial fingerprints. For instance, wave set up can be rather different at several hundreds meters distance.

6.2. Agreement on minimum mandatory metadata and common vocabulary

Accurate and complete metadata contributes to making tide gauge data Findable, Accessible, Interoperable, and Reusable. GLOSS data centres have worked and agreed in the past in a comprehensive list of fields including affiliation, maintenance history, site location information, datum definition, quality control and processing, etc. Examples are shown in UNESCO/IOC. 2020 [3]. While this information needs to be maintained and updated with input from data providers, it is the objective of this section to provide recommendations on a minimum set of mandatory fields. Ideally these should be attached to the time series in adequate and agreed standard formats, and according to predefined and agreed vocabulary.

Minimum metadata to be included can be classified as site essential metadata, variable essential metadata and data essential metadata.

Site essential metadata include the following items:

- Site information, including a site identifier, name of the station, country (ISO3166 name), geographical coordinates (WGS84) and URL if available.
- Data provider with indication of the name of the provider, country, URL and European Directory of Marine Organisations (EDMO) record ID.
- Sensor type, following GLOSS standards agreed among data centres. That includes general type of sensor (radar, float, etc.), the instrument characteristics and the operational history

- Contribution to international programs

Variable essential metadata include:

- Short name of the variable: “SLEV” [1]
- Standard Name: “water_surface_height_above_datum” [2]
- Long name of the variable: “Water surface height above datum” [2]

[1] P09 Medatlas vocabulary
[2] It could be based on the community-supported Climate and Forecast (CF) specification, which supplies a standard vocabulary and some metadata conventions

Data essential metadata include:

- Time reference: relative to UTC recommended and with indication of the format (ISO8601).
- Units of observations. with mm recommended (following GLOSS conventions). Note: CMEMS INS TAC convention: meters, with mm resolution
- Temporal sampling. Recommended to be unique within the same data record. Upgrades of stations or data recorders to higher temporal samplings should imply a different data record.
- Default of null value used (e.g. 99999 or NaN or NC_FILL default value recommended by Unidata-NetCDF)
- Method of data processing such as filtering, sub-sampling, averaging, etc. (also possible to indicate unknown).
- Datum information, including:
  - TG Datum: e.g. Chart Datum, local reference, harbour datum, national geodetic reference...)
  - Tide Gauge Benchmark (TGBM) Name and description, TGBM distance to TG
  - TGBM height above TG datum
- TGBM ellipsoidal height (if available)
- Vertical land movement estimation (if available)
- Tidal constants (if available)
- Level of quality control

Some of the fields recommended here will not be available or known in all cases. Tidal constants, for example, are not public in some countries, and they are also computed in different ways, so agreement in this aspect would be convenient. As mentioned above, standard formats and common vocabularies are essential. This is easy with formats like NetCDF, used by most data centres (GLOSS, CMEMS INS TAC...).

As described in section 4.2, CMEMS INS TAC adopted in 2018 the recommendations on standard NetCDF format and attributes prepared in coordination with the EuroGOOS Tide Gauge Task Team and representatives of GLOSS and SeaDataNet portals:
https://eurogoos.eu/download/NetCdf_Recommendations_forCMEMS_EuroGOOSTGTT_October_2017.pdf
The document includes agreed vocabulary for some of the metadata fields or NetCDF attributes recommended here.

Further work of all data centres would be necessary to agree on vocabulary and metadata.

6.3. New unique ID assignment

Data aggregators currently face many issues when assembling datasets from multiple sources. Tide gauge stations may be given different names on many websites, despite being the same location. Station names may be written in multiple languages, which could lead to duplicate data. Tide gauge operators may change the name or the identifier of the station and so websites collecting those data may treat it as a new location

Station catalog metadata [edit]	[station overview]
SSC ID	SSC-gptx
Station Name	Galveston Pier
Country	United States
Latitude	29.31
Longitude	-94.793
DateLastModified	2015-10-22 19:02:29

Linked codes *cached on 2021-03-19T10:18:47+01:00 this page will automaticaly reload after the cache is refreshed

Codes	Location	Latitude	Longitude	Sensors
IOC: gptx	Galveston_Pier_TX	29.31	-94.793	bwl
IOC: gptx2	Galveston_Pier_TX	29.31	-94.793	pwl
GLOSS: 217	Galveston (Pier 21), TX	29.31666	-94.8	Float, Pressure, Bubbler, Acoustic
UHSLC: 767	Galveston, P.Pier	29.287	-94.79	N/A
UHSLC: 775	Galveston, Pier 21	29.31	-94.793	N/A
PTWC: gptx	Galveston_Pier_TX	29.3100	-94.7933	pwc, pwl, bwl
PTWC: pptx	N/A			
PSMSL: 161	GALVESTON II, PIER 21, TX	29.31	-94.793333	N/A
PSMSL: 828	GALVESTON I, PLEASURE PIER, TX	29.285	-94.788333	N/A
SONEL_GPS: 1934	GALVESTON	29.28515600	-94.78924600	N/A
SONEL_GPS: 2973	GALVESTON	29.32786700	-94.77263600	N/A
SONEL_GPS: 676	GALVESTON	29.32988068	-94.73680897	N/A
SONEL_TG: 2355	GALVESTON	29.30974700	-94.79346800	N/A
SONEL_TG: 2356	GALVESTON	29.28515600	-94.78924600	N/A

Figure 7: Example of Galveston Pier station in the IOC Sea Level Station Monitoring Facility Sea level Station Catalog (SSC).

An example of this problem is shown in Figure 7. Although this station on first appearance looks like it may be one station (reflected in the single GLOSS ID), it has moved location significantly, possibly without the ability to link the two series to the same fundamental benchmarks. There has also been more than one sensor at the station. Conversion of position information from degrees minutes seconds to decimal degrees has lead to recurring decimal positions, which may imply greater precision of position than actually exists, and can create confusion when comparing metadata.

Because of this inability to uniquely identify tide gauge locations, it can be difficult to monitor how many stations are actively contributing to a network or to perform a gap analysis.

Having unique identifiers would help data aggregators confirm the level of processing that a dataset has undergone and offer products from the same location for multiple purposes. If the NRT data comes from the IOC Sea Level Station Monitoring Facility, a quality-controlled dataset from the tide gauge operators and monthly mean data from PSMSL, but they all use the same station identifier, a data aggregator can offer a user a variety of products from the same location, based on their requirements for timeliness and level of quality control. Users may wish to access all of the data from a particular location and compare the datasets available.

Having established that there is a requirement for unique identifiers, we also need to review what type of identifiers are needed. We not only need globally unique identifiers, but persistent ones too. It is not sufficient that a tide gauge operator may have assigned their own identifier to a station, as this may not be globally unique. We also need persistent identifiers that will resolve to a stable object/landing page. A persistent identifier (PID) can be assigned to any object (datasets, time series, locations, sensors, quality control processes) (Figure 8).

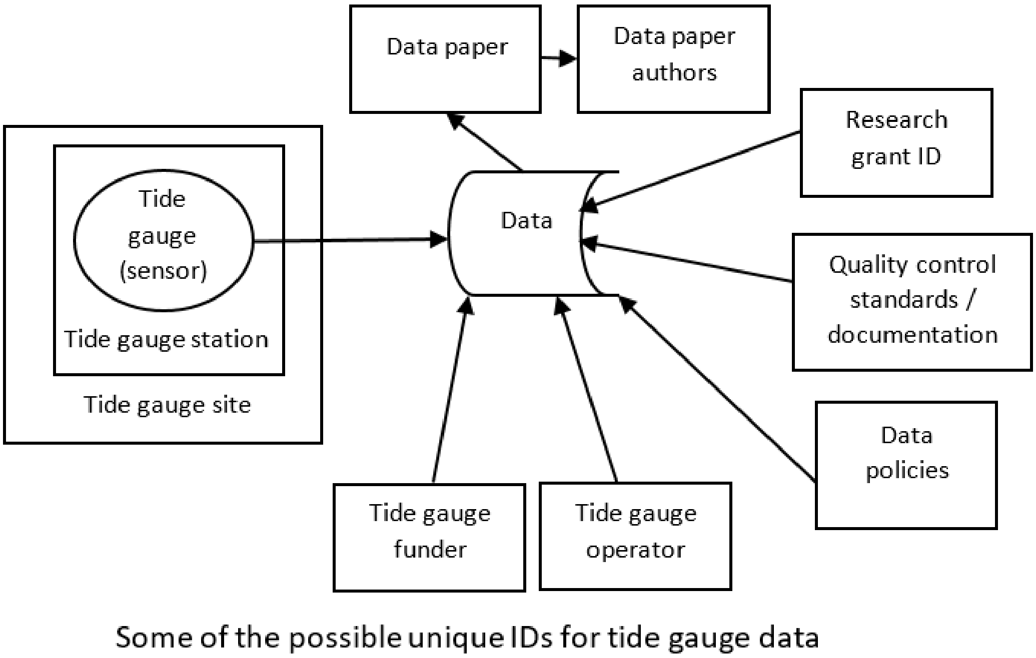

Figure 8: Examples of the possible unique IDs for tide gauge data

Tide gauge (sensor), site, station

As mentioned in section 6.1 above (Review tide gauge station/site definition) it has become essential to distinguish between the tide gauge (sensor), the site and the station. A unique identifier could contain a hierarchy of these concepts (the site ID could relate to the station ID) or the terms could be kept completely separate. This might be appropriate for the tide gauge, as an individual sensor may actually be removed from one location and then installed at another, so you would not want a site dependency introduced. This would also enable data to be traced to an individual instrument, if a calibration error were discovered at a later date, but the sensor had been installed at more than one location, so data from more than one site required recalibrating. Several methods for assigning identifiers are currently being trialled by ocean/earth science network operators. These include using persistent uniform resource locators (PURLs), Digital Object Identifiers (DOIs) and European Persistent Identifier Consortium (ePIC) PIDs (using the Handle resolution system - https://www.pidconsortium.net/). In 2018 PSMSL worked with the RDA WG on PIDs to develop a strategy to assign PIDs to sensors [7]. Tide gauge operators in France have been discussing the possibility of assigning DOIs to tide gauge sites and networks.

Tide gauge data

An identifier can be assigned to a dataset, which can vary in its granularity. It may be all the data from one sensor, one site or even an entire network. PSMSL are intending to create a DOI that covers the 2000 or so tide gauges. Operators in Spain have been investigating assigning a DOI to their dataset.

Sensor/instrument types

Research Resource Identifiers (RRID) are being developed that describe the tools used to carry out research, such as models of instruments used, not the specific instance of a sensor. This has primarily been used in the biosciences but could be adopted for tide gauge data.

Organizations

European tide gauge operators have already been encouraged through the AtlantOS project to create European Directory of Marine Organisations (EDMO) records, to describe their organisations. We may want to encourage operators, funders and other involved organisations to create an entry in the Research Organization Registry (ROR - https://ror.org/) as these are persistent identifiers. The schema does include research adjacent organisations, but further investigation is required to see if port authorities or exploration companies can be incorporated. Having persistent identifiers for organisations enables us to ensure that those involved in the development and maintenance of tide gauge networks and datasets are given proper credit for their work and are able to trace the impact of their contributions.

People

We would encourage individuals involved with the creation and distribution of tide gauge data to apply for an Open Researcher and Contributor ID (ORCID - https://orcid.org/), a persistent identifier for individuals who participate in research and innovation. This will enable us to ensure proper credit is given to those involved in the creation and publication of a dataset.

Projects

The Australian Research Data Commons (https://ardc.edu.au/) have developed a Research Activity Identifier (RAiD - https://www.raid.org.au/) which is a persistent identifier to connect the data, the tools used to collect those data, the people working on the project, the organisations they work for, through to the funders of the project. It may be possible to assign a permanent identifier to the GLOSS programme, and connect all activities and organisations involved under that unique identifier.

6.4. Roadmap for the global tide gauge network

The new strategy on tide gauge data flow, and possible future actions to progress, in a comprehensive and definitive way, in the items presented in this section, should be defined in the framework of the global tide gauge network. As a result of the conversations between GLOSS data centres and OceanOPS in March 2021, a roadmap has been presented for the next two years and is presented in Table 2.

Phase one	Phase two	Phase three
1-3 months	3-12 months	1-2 years
OCG Data Mapping and Metadata workshop IOC Observations Coordination Group and OceanOPS metadata harmonization and opportunity for networks to raise metadata issues and concerns	**Unique IDs for sensors, sites and stations** Following on from OCG workshop and GLOSS data centres meeting, liaise with OceanOPS to develop strategy for assigning unique IDs, using existing systems e.g. WIGOS IDs [8].	**ePIC IDs** Work with organisations that are able to mint ePIC PIDs to investigate assigning PIDs to instrumentation and platforms
GLOSS data centres review meeting Kick-off meeting for monthly GLOSS data centres meetings to discuss data integration and delivery	**GLOSS datasets with DOIs** Assign DOIs to GLOSS datasets where possible such as PSMSL and GESLA dataset	**PIDs for reference documentation** Work with IOC to obtain PIDs/DOIs for reference documentation such as GLOSS manuals
Conduct global survey of organisational/national PID systems Investigate systems in place/in development to see if they can be adopted/adapted and avoid duplication	**Unique IDs for organisations and people** At GLOSS biennial meeting (autumn 2021), request that GLOSS contributors request organisational unique identifiers such as EDMO and also ask individuals to consider obtaining ORCIDs.	**RORs** Investigate incorporating RORs in to existing systems, and ask suppliers/operators/funders to consider generating RORs. Establish if RORs are suitable for non-academic/research organisations

Table 2: Unique IDs development roadmap.

6.5. Quality control and data processing

As described in Section 5 and shown in Figure 4, there are numerous sea level data portals (including those of GLOSS) which receive and process European tide gauge data. Each data portal has its own specific submission criteria (e.g. regarding data format, sampling interval and accompanying metadata) to which suppliers must conform and this demands some level of quality control and data processing skills on the part of tide gauge operators. Data portals receiving raw, near real time observations such as the IOC's Sea Level Monitoring Facility, require minimal intervention on the part of tide gauge operators, apart from the establishment of initial data transmission links. Other GLOSS data portals, like BODC, UHSLC and PSMSL, receive delayed mode data, requiring TG operators to perform more complex data processing, including the application of rigorous quality control procedures, analysis techniques and in some cases filtering datasets to lower frequency intervals. Where TG operators lack these skills, their data are often restricted to the operational monitoring portal of the IOC SLSMF and are not available for scientific applications via other data portals.

To help address this skills gap, the IOC has recently dedicated a manual to quality control procedures [3] - IOC Manuals and Guides No. 83: Volume I: Quality Control of in situ Sea Level Observations: A Review and Progress towards Automated Quality Control (https://unesdoc.unesco.org/ark:/48223/pf0000373566, link verified 15/03/21 17:16). This publication provides guidance on sea level data processing techniques and sources of suitable software.

The types of quality control, the latency and the level of automation is variable and depends on the application and required products. For example, for tsunami warning, it is of utmost importance that the data are made available rapidly, and only lower temporal sampling data is required. In this case, quality control/interpretation is delegated to the expertise and knowledge of tsunami services operators. Automatic quality control procedures can be applied in near-real time (>15 min) for other applications, including use of tide gauge data in operational oceanography, and also to facilitate the work in delayed mode quality control and data processing, for research applications. However, some human intervention and inspection will always be required.

In order to facilitate the operational validation of numerical models with in situ data, CMEMS INS TAC follows a different approach to the one described for other data portals and applies NRT automatic quality control and basic data processing to all in situ data, including tide gauges. The procedures and methods applied follow the best practices described in the last IOC Manual No. 83 mentioned above [3], which are also the ones approved by EuroGOOS DATAMEQ Working Group and CMEMS. This NRT QC includes: spike detection and flagging, stability test, resampling and interpolation of short gaps and computation of filtered hourly values. By-products of this procedure include flagged original data sampling data (as provided by the national centers, in this case usually 5-15 min) and filtered hourly values.

As already mentioned, NRT QC is not sufficient for many applications that require access to the historical records. Also, there are important types of errors that are not yet implemented in existing NRT QC's, or that would be difficult to introduce in existing software packages (e.g.: datum shifts, clock malfunctions). Delayed mode quality control and data processing is necessary for the generation of a reprocessed sea level product from tide gauges that can be easily accessed and regularly updated for modellers, the altimetry community and scientists. As mentioned in section 4.2, CMEMS intends to implement this during its next phase, starting in 2022.

The need of high-quality historical records will require to increase the interaction with the original data providers, that will ultimately be able to provide information about the operation and history of the station. This interaction will also be critical to maintain a comprehensive metadata information, that should also be quality controlled and regularly updated. It could also be convenient to identify those institutions that already perform this effort for their national networks, and that could provide regular updates of the historical quality-controlled data to international data programs. However, when the institutions lack the resources and skills for this work, this could possibly be performed by one of the data centres mentioned here.

Nevertheless, the activities of the EuroGOOS Tide Gauge Task Team (such as the recent 1st EuroSea Tide Gauge Network Workshop) have revealed that some tide gauge operators would benefit from more formal face-to-face training in data processing and from the development of clear pathway for them to follow as they build data repositories. Consequently, the focus of the 2nd EuroSea Tide Gauge Network Workshop, to be held in November 2022, will be the development of new automatic QC algorithms and products from tide gauge data.

7. Actions in the European Tide Gauge Network

According to the roadmap presented in Table 4 for the global tide gauge network, there is now a plan for implementation of some of the recommendations of this manuscript, especially along the main discussed topics of new unique identifiers and metadata. Therefore, in principle, most of the work in the European network should rely now on the progress of the work at a global level. For example, if there is a new unique identifier defined with GLOSS, CMEMS, EMODnet or SeaDataNet will need to carry this information in their data files. If this information is managed at OceanOPS, and OceanOPS has the metadata database filled in by the operators of the GLOSS network, they will be able to provide API to get this information and update other data portals holdings.

Nevertheless, we will describe below some on-going initiatives that affect the European network, already identified as a priority by the EuroGOOS Tide Gauge Task Team, and partly to be accomplished in the framework of EuroSea and other projects. Their final implementation should however converge towards GLOSS requirements and decisions through the next years.

7.1. European tide gauge metadata inventory

As one of the actions started by the EuroGOOS Tide Gauge Task Team, in response to above mentioned recommendations, a New Tide gauge metadata catalogue is being developed by the Irish Marine Institute, which is a key deliverable for EuroSea WP3. The purpose of this is to develop a fundamental online, live, managed metadata catalogue registering all permanent tide gauges that are deployed along European and adjacent coastlines.

The objective of this task is to compile all relevant metadata from European tide gauges in a unique website, where national data providers can easily contribute and update their information, and that international programs could use for completion of their records. For the latter, these metadata will also be easily transferred to files easily ingested by automatic procedures (formats: XML, JSON, etc).

The tool is not yet finished (EuroSea project has just started in November 2019), but a proof-of-concept version already exists: http://eutgn.marine.ie/geonetwork/srv/eng/catalog.search#/home.

Developed with open-source applications, at this moment it makes use of a shared spreadsheet that feeds a database in the Marine Institute and a Geo Network application. It displays a map of the stations with a colour code related to the metadata content, that allows easy identification of where national data providers need to be contacted. The final place of the tool will be decided by the EuroGOOS Tide Gauge Task Team after consultation with GLOSS, CMEMS and EMODnet data portals, when finished. Before this, agreement on mandatory metadata, unique ID for tide gauges and common vocabularies definition should be ideally accomplished for a final useful tool at the end of the project. It should rely on existing metadata catalogues and contribute to future or global tools developed for the tide gauge network.

7.2. Implementation of recommendations on minimum metadata

Another action could be the implementation of the minimum mandatory metadata defined in section 6.2, and progress on the required common vocabulary for these fields with all data centres, as it was done for CMEMS in 2018. This will be possible as the EuroGOOS Tide Gauge Task Team/GLOSS subgroup on minimum metadata progress in the coming months.

Part of this effort has already started. As quoted in section 6.2 one of the essential metadata information associated with the sea level record is the knowledge of the records datum which is itself linked to the Tide Gauge primary Benchmark (TGBM). The information of the position of the TGBM in a geocentric reference frame is thus essential for many sea level applications (Absolute Sea Level Trend, World Height Datum, Satellite Altimeter calibration, ...). In 2020, SONEL, the GLOSS data assembly centre for GNSS at Tide Gauges, in the framework of a project of EuroGOOS with the Environmental European Agency (EEA), launched a campaign in 24 countries in Europe to collect this essential information. This campaign resulted in the identification of a total of 112 Tide Gauges with an Ellipsoidal Height of the TGBM that will be shared with CMEMS.

References

[1]. IOC, 2012. The Global Sea Level Observing System (GLOSS) Implementation Plan 2012. (IOC Technical Series No. 100). Intergovernmental Oceanographic Commission of UNESCO, Paris, 37pp. Available from https://unesdoc.unesco.org/ark:/48223/pf0000217832

[2]. IOC, 2006. Manual on sea-level measurement and interpretation. Volume IV: An Update to 2006. (eds. Aarup, T., Merrifield, M., Pérez Gómez, B., Vassie, I. and Woodworth, P.). IOC Manuals and Guides No. 14. Intergovernmental Oceanographic Commission of UNESCO, Paris, 80pp. https://unesdoc.unesco.org/ark:/48223/pf0000147773

[3]. UNESCO/IOC. 2020. Quality Control of in situ Sea Level Observations: A Review and Progress towards Automated Quality Control, Vol. 1. Paris, UNESCO. IOC Manuals and Guides No.83. (IOC/2020/MG/83Vol.1). https://unesdoc.unesco.org/ark:/48223/pf0000373566

[4]. Koop-Jakobsen K, Waldmann C, Huber R , Harscoat Valerie, Pouliquen Sylvie (2016). Data Harmonization Report. AtlantOS project. D7.1. https://doi.org/10.13155/47090

[5]. Harscoat Valerie, Pouliquen Sylvie, AtlantOS Wp7 Partners (representatives Of Networks And Integrators) (2016). Data Management Handbook AtlantOS. D7.4. https://doi.org/10.13155/48139

[6]. Tanhua T, Pouliquen S, Hausman J, O'Brien K, Bricher P, de Bruin T, Buck JJH, Burger EF, Carval T, Casey KS, Diggs S, Giorgetti A, Glaves H, Harscoat V, Kinkade D, Muelbert JH, Novellino A, Pfeil B, Pulsifer PL, Van de Putte A, Robinson E, Schaap D, Smirnov A, Smith N, Snowden D, Spears T, Stall S, Tacoma M, Thijsse P, Tronstad S, Vandenberghe T, Wengren M, Wyborn L and Zhao Z (2019) Ocean FAIR Data Services. Front. Mar. Sci. 6:440. doi: https://doi.org/10.3389/fmars.2019.00440.

[7]. Stocker, M, Darroch, L, Krahl, R, Habermann, T, Devaraju, A, Schwardmann, U, D'Onofrio, C and Häggström, I. 2020. Persistent Identification of Instruments. Data Science Journal, 19: 18, pp. 1–12. DOI: https://doi.org/10.5334/dsj-2020-018

[8]. Guide to the WMO Integrated Global Observing System (WMO-No. 1165), https://library.wmo.int/doc_num.php?explnum_id=10040.