Incorporating Discrete Unmanned Maritime System Data Collections Into NCEI Synthesized Data Products

Sharon Mesick
NOAA's National Centers for Environmental Information
Stennis Space Center, MS, USA
sharon.mesick@noaa.gov

Zhankun Wang
NOAA's National Centers for Environmental Information
Northern Gulf Institute, Mississippi State University
Stennis Space Center, MS, USA
zhankun.wang@noaa.gov

Alexey Mishonov
NOAA's National Centers for Environmental Information
ESSIC/CISESS-MD, University of Maryland
Silver Spring, MD, USA
alexey.mishonov@noaa.gov

Tim Boyer
NOAA's National Centers for Environmental Information)
Silver Spring, MD, USA
tim.boyer@noaa.gov

Huai-min Zhang
NOAA's National Centers for Environmental Information
Asheville, NC, USA
huai-min.zhang@noaa.gov

***Abstract* — In recent years, Unmanned Maritime Systems (UMS) have become widely used in hydrographic surveys and oceanographic studies, especially in coastal regions. NOAA's National Centers for Environmental Information (NCEI) has received an increased amount of data collected by UMS in the past 10 years. A variety of different types of data are typically collected simultaneously by UMS, which raises questions about the best practices for UMS data management. In this manuscript, we discuss enhancing the value of UMS data through integration into NCEI synthesized products. These products range from data-type specific databases like the Surface Underway Marine Database, and the World Ocean Database to analyses (value-added) products such as climatological mean fields and ocean time series of variables such as ocean heat and oxygen content. Data from Saildrones and Wave Glider missions demonstrate how data collected by Unmanned Surface Vehicles (USV) are integrated into the Surface Underway Marine Database. The integration of data from Unmanned Underwater Vehicles (UUV) systems - such as gliders - into the World Ocean Database is also described. Data management procedures and quality controls are discussed. The application of standardized methods adds long-term value to the original UMS, and improves the quality and reliability of the synthesized products. NCEI stewardship practices maximize NOAA's investment in environmental research, converting scientific insights into dynamic, usable information that inform strategy and decision making in government, academia, and the private sector.**

Keywords—Unmanned Maritime Systems, Saildrone, Glider, Unmanned Surface Vehicle, Unmanned Underwater Vehicle, Data Management, Data Synthesis, Essential Oceanographic Variables

I. INTRODUCTION

The NOAA National Centers for Environmental Information (NCEI) is the Nation's leading authority for environmental data. NCEI manages one of the largest archives of atmospheric, coastal, geophysical, and oceanic research in the world. NCEI contributes to NOAA's mission by developing new data products and services that span the science disciplines and improve data discovery, access, interoperability, reliability, and reuse. Both the data archive volumes and the user-requested volumes at NCEI have greatly increased in recent years [1]. A variety of data products have been developed at NCEI by extracting data from individual data archive packages and integrating them into synthesized, referential data products [2]. Each product contains archived data records broadly collected by NOAA and other US agencies, academia, commercial entities, and international partners. Most of the oceanographic data collections are archived as discrete data packages and are organized within the archives in a pre-defined structure, for example by cruises, projects/programs/experiments, collection methods, working groups, or data types. The data are frequently archived in their raw format, such as in engineering units in binary software package form, and/or without quality assurance and data quality control checks. During the archiving process, NCEI oceanographers and data scientists work with data submitters to extract and catalog metadata information to support online data search tools (e.g. NOAA OneStop data search platform, https://data.noaa.gov/onestop/). Rich metadata improves the public's ability to find and access the archived data collections. Users can be faced with a difficult and time-consuming challenge when seeking to find, access, and reuse data from multiple archive packages. NCEI develops synthesized data products to improve the user experience, increase the data reliability, and to add value to the data by placing the data into spatio-temporal context. In this process, NCEI extracts metadata and data from individual archived data packages, applies standard quality control (QC) procedures and organizes the data consistently in a uniform format. For example, both the Surface Marine Underway Database (SUMD) [3] and the World Ocean Database (WOD) [4] products each separately integrate millions of data sets into foundational products widely used as references in other contexts. When single deployments become part of a global dataset, their vital contribution to large-scale ocean monitoring becomes apparent.

Unmanned Maritime Systems (UMS) are defined as waterborne systems that can execute components of the ocean observing mission by utilizing sophisticated environmental

Supported and funded by National Oceanic and Atmospheric Administration.

awareness and without human input [5]. UMS include: Unmanned Underwater Vehicles [1] (UUVs) - unmanned, untethered, systems capable of autonomous submerged operation [2] ; and Unmanned Surface Vehicles (USVs) - untethered, self-propelled surface crafts ranging in size from human-portable systems to small boat-size vessels that are capable of autonomous, semi-autonomous, or remote-controlled operations. Ocean observations from UMS platforms are a recent addition to the NCEI archive collection and to NCEI synthesized products.

In this manuscript we focus on the methods to add Unmanned Maritime System (UMS) data to the NCEI archive collection and into NCEI synthesized data products, and discuss the value of doing so. In Section II, we describe the SUMD and the WOD as two example NCEI data products and then discuss the UMS data management workflow. Details of surface vehicle data integration into the SUMD and underwater vehicles into the WOD will be described in Section III and IV, respectively. Data usages of the SUMD and WOD will be discussed in Section V, followed by a summary on the advantages of data incorporation in Section VI.

II. UMS Data Management Workflow

A. *Surface Underway Marine Database*

The SUMD [3] is a comprehensive dataset of quality-controlled near real-time in situ ocean surface underway data collected by vessels and unmanned surface vehicles. Sea surface temperature (SST) and salinity (SSS) are the two essential data variables in SUMD, which are primarily collected by thermosalinographs (TSG). The TSG, usually installed inside and near the hull of a ship, are instruments used to measure ocean characteristics near the sea surface, typically in flow-through systems operating continuously while a vessel is underway. In addition to sensors measuring temperature and salinity (conductivity), other sensors can be optionally installed into the system for a wider range of measurements, such as a fluorometer and/or a turbidity sensor. Physical and biochemical measurements from other non-TSG sensors are also included in the SUMD when the data are collected concurrently with TSG SST and SSS, such as chlorophyll, dissolved oxygen, wind speed and direction, air temperature, relative humidity, micro-plastics, acoustic biomass etc. Fig. 1 shows the spatial distribution of the SUMD data colored by data sources. The data cover most regions of the global ocean. The database consists of data from over 460 vessels and unmanned surface vehicles from 1986 to the present.

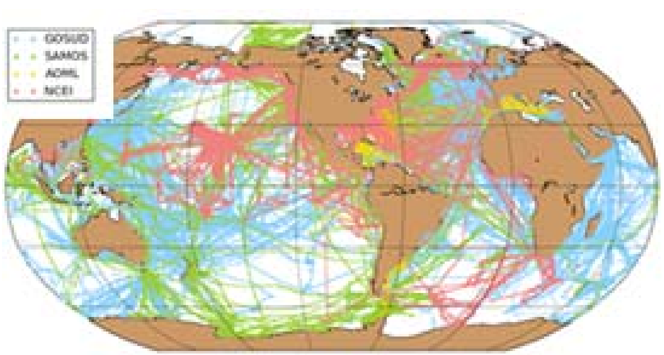

Fig 1 Graphical spatial distribution of SUMD data by data sources. GOSUD: Global Ocean Surface Underway Dataset; SAMOS: Shipboard Automated Meteorological and Oceanographic System data; AOML: Atlantic Oceanographic and Meteorological Laboratory thermosalinographs data; and NCEI: National Centers for Environmental Information data.

Compared to other available TSG underway datasets, the SUMD has several advantages [6]: (1) As one of the world's most complete TSG datasets, it contains all data from the different TSG data assembly centers, including the Shipboard Automated Meteorological and Oceanographic System (SAMOS) [7], Global Ocean Surface Underway Data (GOSUD) [8], and Atlantic Oceanographic and Meteorological Laboratory (AOML), and additionally includes historical data from NCEI archives. (2) When different versions of a dataset are available, the dataset with the highest resolution is always selected. (3) All data are processed using the same 11-step quality control procedures and criteria, and are flagged by using a two-level system to provide a well-organized, uniform quality-controlled dataset for the user community.

More than 380 million individual data records have been incorporated into the SUMD database as of this writing. The USV data account for about 1-2% of the SUMD data collection overall. We anticipate that the percentage of data from USV will increase dramatically over time as these systems move from research to sustained operations. USV have a unique advantage over large vessels in the ability to operate in shallow coastal waters and in difficult to reach locations.

B. *World Ocean Database (WOD)*

The WOD represents the world's largest collection of ocean profile data, containing over 17 million oceanographic casts dating from the 1870's through the present. The WOD [4] incorporates more than 20,000 different data sets received and archived at NCEI from various national and international sources. The data are collected from about 800 institutes around the world, by a range of more than 26,900 different platforms (including 16,092 profiling floats, 1,992 individual pinnipeds, 252 ice-drifting floats, 202 moored buoys, 200 individual gliders, and more than 8,000 ships).

Data undergo standardization and quality control procedures as described in the WOD documentation [4]. The WOD contains 29 depth-dependent physical and biochemical variables from 11 instruments types: Ocean Station Data (e.g., bottles), Mechanical Bathythermographs (MBTs), Expendable Bathythermographs (XBTs), Conductivity-Temperature-Depth (CTD), Undulating Oceanographic Recorders, Profiling Floats (e.g., Argo floats), Moored Buoys, Drifting Buoys (e.g., Ice-tethered platforms), Gliders, and Instrumented Pinnipeds (e.g., instrumented elephant seals). The WOD also contains a small subset of surface-only data, but its primary focus is ocean profile data. The WOD is the foundational dataset used in calculating

[1] Unmanned Underwater Vehicles (UUV) are sometimes referred to as Autonomous Underwater Vehicles (AUVs)

[2] Buoyancy gliders (e.g. Slocum gliders) use small changes in buoyancy in order to move up, down and forward in the ocean and are considered as a UUV in our data management workflow. Argo floats are not normally retrievable. Discussion continues as to whether Argo floats are formally included in the NOAA UxS definition of UUV.

the World Ocean Atlas (WOA) climatologies [9, 10] as well as the Global Ocean Heat and Salt Content anomalies [11-14]. Data from Unmanned Underwater Vehicles such as gliders are separated into vertical upward/downward casts and then incorporated into the WOD as profiles/casts [4].

C. *Unmanned Maritime Systems Data Workflow*

UMS are capable of executing pre-programmed missions without operator interaction, and are operated on a continuum from attended to fully autonomous. UMS are increasingly valuable in extending the foundational references because areas where UMS are often deployed are not well covered by other platforms. Incorporating UMS data into NCEI referential data products extends the usage and utility of the data far beyond the original purpose of the UMS deployment.

Fig. 2 illustrates the NCEI UMS data incorporation workflow. Data from both UUV and USV are archived at NCEI. Data are from different sources and from different types of platforms. In order to incorporate these disparate data into standard databases and longitudinal records, established protocols and best practices for data standardization and quality control are applied. Data standardization includes converting data into common digital formats and uniform measurement units; derivation of standard metadata including documenting observing platforms, instruments, and sensors, and application of standard data attribution. Data Quality Control (QC) checks may include individual checks for platform identification, time validation, location (e.g. landlock), range, climatology control, constant values, spikes, and gradients. QC flags are usually added to the database to record data quality suggestions. After data standardization and QC checks are completed, data from surface vehicles are integrated into the SUMD and underwater profiles are integrated into the WOD. Details are described in the following two sections.

Fig 2. NCEI unmanned maritime system data incorporation workflow.

III. Incorporating Surface Vehilce Data into SUMD

USV are an emerging data source for surface oceanographic measurements, especially in the coastal regions and remote areas, such as in the arctic, where traditional large vessels cannot easily reach. A variety of different types of data are typically collected simultaneously by USV during the mission to maximize the usage of the vehicle, which raises questions about the best practices for USV data management at NCEI. We have found that most data from USV can be considered as surface underway data and can be integrated into the SUMD. Table I shows the unmanned surface vehicles datasets that have been archived at NCEI. These data are either collected by wave gliders [15] or from Saildrones [16], which are the two most frequently seen commercial USVs in the NCEI archives. To date, most data submissions are from the Pacific Ocean (10 datasets), the North Atlantic (2 datasets) and the Arctic region (2 datasets). The arctic data are very valuable to fill in the observational gaps in SUMD as few surface in situ observations have been conducted in this region in the past.

TABLE I. Unmanned Surface Vehicles Data at NCEI

Platform		**Data Information**		
Type	***# Data Sets***	***Start and end dates***	***Region***	***Accession #***[c]
WG[a]	1	2011/07/29-2011/09/23	Arctic	88841
WG	2	2011/08/04-2011/09/29	North Pacific	209672
WG	1	2011/10/19-2014/08/21	North Atlantic	125198
WG	4	2011/11/18-2013/02/14	Pacific Crossing	114435
WG	2	2012/07/18-2012/10/18	North Pacific	209674
SD[b]	2	2015/05/23-2016/09/03	Bering Sea	187987
SD	1	2017/03/17-2017/04/25	North Pacific	209685
WG	1	2017/07/27-2017/09/27	North Pacific	209676
SD	1	2017/09/02-2018/05/18	North and Tropical Pacific	206573
SD	1	2017/09/02-2017/12/01	North and Tropical Pacific	206574
SD	1	2018/03/27-2018/04/22	South Pacific	209684
WG	1	2018/04/28-2019/10/27	South Pacific	209683
WG	1	2018/05/01-2018/06/26	North Pacific	188678
SD	1	2019/01/30-2019/02/26	North Atlantic	212656

[a.] Wave Glider; [b.] Saildrone.c. Data can be downloaded at https://www.nodc.noaa.gov/access/index.html by searching associated accession number

One characteristic of the USV observations is that many different types of data are collected simultaneously by different instruments and sensors mounted on the USV [15, 16]. For example, in the 2015-2016 Saildrone data [17] (NCEI accession # 187987), in addition to "traditional" underway oceanographic (temperature, salinity, dissolved oxygen, chlorophyll, CDOM, photosynthetic active radiation) and meteorological (air temperature, barometric pressure, wind, relative humidity) variables, non-traditional underway variables were collected, e.g. fish biomass and marine magnetics. Usually data collections are distributed by mission and data type to different groups for their usage and research. This mission-based distribution of data from the platform (or source), can make it difficult for users to find and access all the data from the same experiment/cruise, or by regions (regardless of type). In response, NCEI has expanded the SUMD data model to include additional types of underway data into the database. At this time no quality controls will be applied to non-traditional variables [3, 5]. NCEI is in the process of integrating all the unmanned surface vehicles data in Table I to SUMD.

IV. Incorporating Glider Data into WOD

Hundreds of different Autonomous Underwater Vehicles[3] (AUVs) have been designed over the past 50 years and some have become commercially successful and widely used in the ocean observing system community. AUVs carry a variety of sensors for navigation and ocean features mapping [18, 19]. We use glider data - one of the most frequently used types of AUVs - as an example to describe the integration of UMS underwater data into the WOD.

A. Gliders

A glider is an autonomous underwater vehicle (AUV) propelled by buoyancy force that moves from the ocean surface along a slant trajectory through the water column to a programmed depth and back to the surface while measuring oceanographic parameters [20]. Modern gliders carry various sensors to measure pressure, temperature, conductivity, chlorophyll a fluorescence, oxygen, transmissivity, and other key parameters [20-24]. Gliders can travel thousands of kilometers while making several hundred descents and ascents underway, thus achieving high vertical and horizontal resolution. Since gliders can be retrieved and reused, they represent one of the most cost-effective tools for oceanographic data collection. The annual operating cost of a glider is equivalent to a fraction of one ship-day [21].

Data collected by different glider platforms are integrated into WOD, as shown in Table II. Currently Seagliders produced by Kongsberg Underwater Technology Inc.; Slocum gliders manufactured by Teledyne Webb Research Co., and Spray gliders produced by Bluefin Robotics are the three dominant types of glider data in the NCEI archives with Slocum gliders provide approximately 72% of the glider casts in WOD (Table II). Data from the newly developed SeaExplorer glider manufactured by Alseamar was recently submitted and included in the WOD. Detailed information on gliders specifications and their functions can be found in [20-24].

Since 2010, profile data from Gliders comprises 25%-30% of all data integrated into the WOD. Each glider deployment has very high frequency cycling in a (usually) small geographic area. So while sheer numbers are high for glider casts, the geographic coverage is usually small for any given year. Those glider data can provide high resolution two-dimensional and even three-dimensional mapping of oceanographic features in a small region, which cannot be done by ARGO or other traditional observations. This is particularly important for fine scale and microscale studies. Furthermore, the locations where the gliders are deployed are usually in the continental shelf and slope regions, where other platforms cannot reach. So although the glider contribution is often regional/local the observations are very important for calculations such as ocean heat because the glider data are often the only data in the regions they cover.

NCEI serves as the central Data Management System for the Glider data by integrating data from different Glider Data Assembly Centers (DAC) into the archives. Those DACs include Integrated Marine Observing System (IMOS), Integrated Ocean Observing System (IOOS), European Gliding Observations (EGO), Global Temperature and Salinity Profile Programme (GTSPP) and others. Data from Gliders add significant contributions to the climatological mean fields' calculation and the long-term trend analysis, and play a supporting role (to Argo) in ocean heat content calculations.

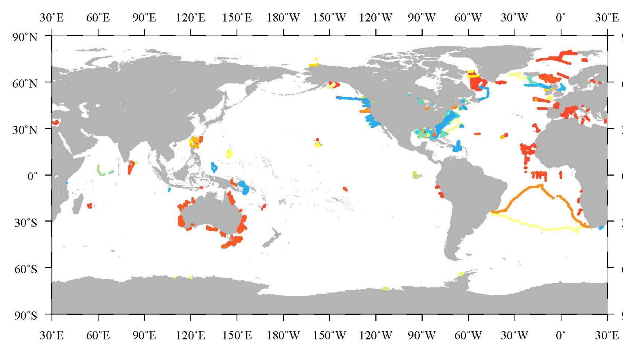

Fig. 3 Global coverage of the WOD glider data collection (as of Aug 2020). Different colors shows various data sources: orange – IMOS, Australia, red – EGO, Europe, blue – IOOS, etc.

TABLE II. WOD Glider Data & Glider Capabilities Based on Standard Packages

Glider	Casts/ % in WOD	Parameters			
		Maximum Depth, m	***Typical Speed, m/s***	***Maximum Range, km***	***Endurance days***
Seaglider	231,461 /13.36%	1,000	0.25	4,600	270
Slocum:	1,246,815/71.94%				
Alkaline		1,000	0.35-0.50	1,200	15-50
Rechargeable		1,000	0.35-0.50	3,000	30-120
Lithium		1,000	0.35-0.50	13,000	120-540
Spray	87,427 / 5.04%	1,500	0.25	4,700	180
Sea Explorer	1,806 / 0.10%	1,000	0.25-0.50	1,300	64
No platform info	165,503 / 9.55%				
Total	1,733,012				

[3] Autonomous Underwater Vehicles (AUVs) are sometimes referred to as Unmanned Underwater Vehicles (UUV)

B. Glider Data Sources

Gliders data integrated into the WOD are collected all over the world by several major Agencies. Figure 3 shows global geographic distribution of the glider data, color-coded according to data sources. This figure clearly shows that as currently implemented, the majority of the data are collected close to the contributing countries' economic zones.

Figure 4 shows the contribution of the glider data made by different programs as a percentage of the total glider data collection at NCEI. The major contributor of the glider data in WOD is IMOS, which operates a fleet of gliders measuring oceanographic parameters on shelf and boundary currents in Australian waters. It operates a number of Slocum gliders in the Coral Sea, East Australian Current off New South Wales and Tasmania, Southern Ocean; southwest of Tasmania, the Leeuwin and Capes Currents off South Western Australia and the Pilbara and Kimberly regions off North Western Australia[4].

Another major submitter is the IOOS[5] which collects data via the Underwater Glider Network Map and includes current and historical glider missions dating back to 2005 from its regional partners including: Gulf of Mexico (GCOOS), Southern California (SCCOOS), Northern Pacific (NANOOS), Central and Northern California (CeNCOOS), Great Lakes (GLOS), Mid-Atlantic (MARACOOS), the Atlantic Oceanographic and Meteorological Lab (AOML), and the US Navy.

The pioneer in glider data collection is the University of Washington team. UW Seagliders are the main source of the glider data in WOD for years 2002-2009. Today these data are submitted to NCEI via the IOOS channel. The Canada Department of Fisheries has also submitted a significant volume of glider data in 2010-2017 via GTSPP.

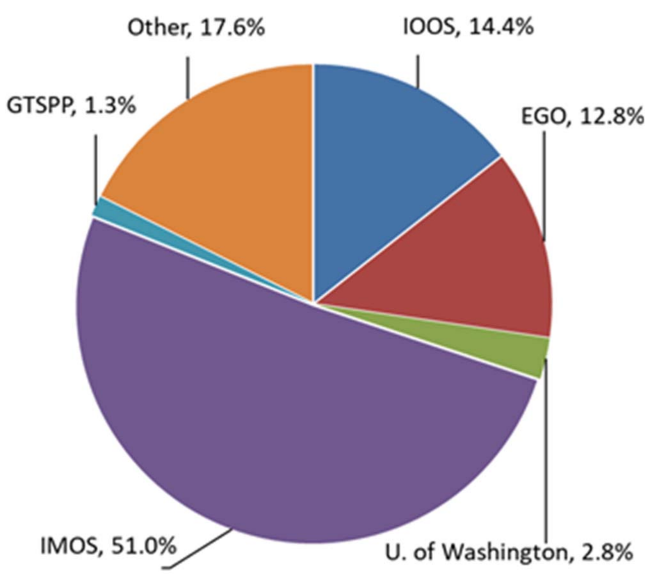

Fig. 4 Projects' contribution to the WOD glider data collection.

About 20% of the glider data in WOD has no identified program affiliation, which makes it difficult to credit data collecting/submitting agencies/institutions and adds to the challenge of applying standard quality control/assurance procedures.

Figure 5 shows the temporal distribution of glider casts in WOD over the period of data collection. NCEI noted a sharp increase in glider data submissions after the initial period of the technology development in 2002-2008. This reflects a consistent, growing interest in gliders as a convenient and versatile platform for oceanographic research. It is also worth noting that current data submission in 2020 is on its lowest level since 2008. Studies are underway to determine if there is any connection to the current pandemic situation [25].

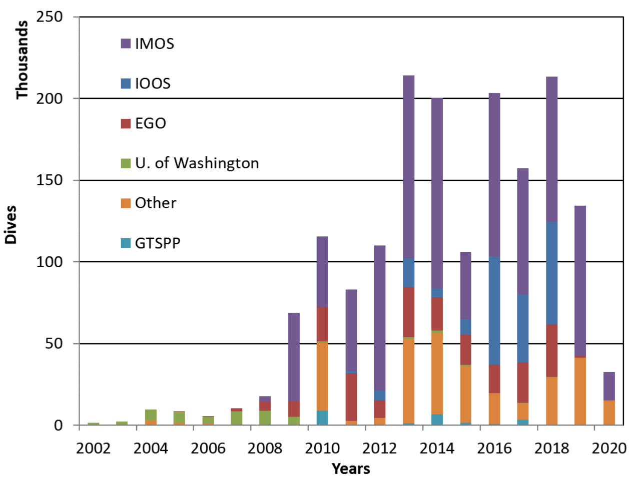

Fig. 5 Time distribution of the WOD glider data collection (as of Aug. 2020): IMOS, Australia; IOOS – Integrated Ocean Observing System; EGO (Everyone's Gliding Observatories) – European glider initiative; University of Washington; GTSPP (Global Temperature-Salinity Profile Program); and other.

C. Glider Data workflow

The workflow for incorporating glider data into the WOD varies slightly by Source/ Institution/ Project [4]. All glider data are submitted in the NetCDF format. Detailed specifications within NetCDF files are slightly variable for different data submitters and NCEI has developed several flavors of the conversion utilities tuned for each data submitter [4].

The most straightforward data process has been developed for the IOOS Glider Data Assembly Center (GDAC) submissions. In this model, participating institutions submit data to IOOS for organization into a standard data package including a single deployment of one glider, accompanied by standard metadata. The IOOS GDAC data package is automatically ingested to the archive, where metadata is generated, data packages are archived and a unique identifier is assigned. IMOS glider data collections are provided via the IMOS web-portal[4] and downloaded to NCEI quarterly-to-semiannually depending upon availability of the newly collected data. The data are organized by individual glider deployments. Similarly, glider

[4] http://imos.org.au/facilities/oceangliders/

[5] https://gliders.ioos.us/data

data collected by EGO (Europe) are stored on EGO web-portal and downloaded to NCEI annually.

All NCEI glider data goes through data format conversion and quality controls including data duplication and overlap checking, and real-time vs delayed time data comparison before WOD integration. For newly deployed gliders new codes are issued and added to the appropriate code tables for platforms, participating Institutions, Principal Investigators, and projects.

After merging new data into WOD, glider data are included in calculation of quarterly updates on the Ocean Heat Content, Salt Content, and steric sea level [4, 11, 12]. Data are available for public use by the broad scientific community and can be accessed via the standard WOD web-interfaces[6].

V. Data Usages of SUMD and WOD

Incorporating UMS data into NCEI's synthesized data products such as SUMD and WOD adds value to the data by extending the usage and utility of the data far beyond the original purpose of deployment. The utility is in the aggregation of all available historic and recent ocean observations across institutes, projects, countries, platforms, and instruments in a uniform format, with uniform quality control. Usage is extended by free public access to the synthesized foundational data products.

One UMS deployment can provide a huge amount of information about a finite ocean region over a finite time. The platform is usually deployed for a very specific research goal or as part of a larger monitoring effort. Incorporating the data from these platforms into synthesized products places the data into spatio-temporal context for reference. Now the one UMS deployment is accessible with all other data in the same oceanographic area adding the dimension of time: evolution of conditions over time brings greater understanding of ocean dynamics and environmental conditions; contributing to a long-term mean provides understanding as to whether the individual deployment is indicative of usual conditions or represents an outlier of the ordinary phenomena. Spatially, including a UMS deployment into an NCEI referential product allows for an understanding of how the specific measurements compare with surrounding ocean areas, those areas which directly affect the ocean conditions in the measured area, or similar ocean regimes along other coasts or in different ocean basins.

When single deployments become part of a global dataset, their vital contribution to large-scale ocean monitoring becomes apparent. The ocean observing system for temperature, for instance, is dominated spatially by Argo floats. The measurements from the global observing system are used to calculate ocean heat content and provide a measure of the Earth's Energy Imbalance (EEI). In particular, Gliders become increasingly valuable in this context, as boundary currents, shelf regions, and marginal seas - areas where gliders are often deployed - are not well covered by Argo floats. These areas can account for a significant amount of change in global ocean heat content [26].

SUMD serves as a significant resource for establishing matchups with satellite ocean surface salinity for validation and evaluation. Due to the high resolution of in situ data, these matchups may prove to be more informative than single observed data points. Underway data can provide continuous fine scale observations to capture the spatial variations of the surface ocean (e.g. ocean fronts, river plumes etc.). Moreover, about 50% of the SUMD data were collected in the coastal regions with depth less than 1000m where the variation of the SSS and temperature is larger and faster compared to open oceans. Satellite observations are usually more erroneous in the coastal regions. Coastal regions have more economic influences on human activities, such as tourism, fishing and aquaculture than the open oceans. SUMD becomes a critical and most reliable source for SSS research, particularly in the coastal regions.

SUMD is also an important data source for the NOAA Satellite Applications and Research (STAR) Sea Surface Density Product and is one of the sources for the NOAA National Centers for Environmental Prediction (NCEP) for climate models and products. NCEI will incorporate the SUMD data into the International Comprehensive Ocean-Atmosphere Data Set (ICOADS), which is the primary data source for the Optimum Interpolation Sea Surface Temperature (OISST) and the Extended Reconstructed Sea Surface Temperature (ERSST.) Both SST products provide monthly SST fields for climate change and global warming research. SST and SSS from SUMD are needed to improve our knowledge of earth's water cycle and climate and proven to be valuable for describing climate variability at seasonal to decadal time scales, improving estimates of evaporation and precipitation (E-P) budgets, and assessing numerical model performances.

WOD is the foundational dataset used to compute the WOA and to assess regional climatologies[7]. Data from WOD is also used for ocean climate studies by estimating the ocean heat, salt and oxygen content, steric sea level change and their long-term trends over time [13]. Since the first WOA release in 1982, the WOD and WOA products have been cited over 8,000 times in the literature. The sheer size of the citation count illustrates the vast utility of the WOD and the products derived from WOD (e.g., WOA) in earth system research. Additionally, because WOD and WOA have been utilized numerous times in various research projects, the datasets have proven to be very high quality, with feedback from users continually helping to improve the quality of the database. The near surface data from the WOD is also one of the main data sources for the construction of the Extended Reconstructed Sea Surface Temperature (ERSST). WOD is also widely used in oceanographic research, for example, data reanalysis, ocean heat content and salt content calculations, sea level assessment, deoxygenation research, sound speed calculations and mixed layer depth/barrier layer estimation etc. The WOD is used in study and monitoring, for research and public policy, for fields as diverse as biological regime changes, climate modeling, transport routing, fisheries, etc.

VI. Summary

The synthesis of oceanographic observations into NCEI foundational data products such as SUMD and WOC adds long-

[6] https://www.ncei.noaa.gov/products/world-ocean-database

[7] https://www.ncei.noaa.gov/regional-ocean-climatologies

term value to the original observations by improving data standardization, data quality control, public data access, and opportunities for data reuse.

As operational UMS deployments increase and the volume of data from UMS increases, NCEI will seek to build upon the successes found with standardized, automated archive data ingest systems such as the IOOS GDAC, and to extend these best data management practices to form a cohesive UMS data enterprise. The recent increase in UMS data archive submissions to NCEI, both from USV and UUV platforms adds new dimensions.

Synthesizing UMS data into foundational data products such as SUMD and WOD extends the foundational data reference to shallow coastal waters and difficult to reach areas typically inaccessible to established platforms such as ships and Argo floats. When single UMS deployments become part of a global dataset, their vital contribution to large-scale ocean monitoring becomes apparent.

Data synthesis at this scale requires rigorous, standardized practices to ensure data reliability. NCEI applies uniform data standardization and quality control procedures to UMS collections before product integration. Scientific data analysts apply subject matter expertise in long term data preservation and management to: a) improve metadata quality; b) convert data to common formats; and c) apply data quality control checks.

NCEI stewardship practices maximize NOAA's investment in environmental research, converting scientific insights into dynamic, usable information that inform strategy and decision making in government, academia, and the private sector.

Acknowledgment

The authors would like to thank the scientists and data providers for contributing their data collections to the NCEI archive, as well as data managers and our NCEI colleagues for stewarding the data for long-term preservation. Without this combined effort, data-management studies like this would not be possible. This study was partially supported by NOAA grants NA18OAR4170438 (Mississippi-Alabama Sea Grant Consortium), 363541-191001-021000 (Northern Gulf Institute) at Mississippi State University, and NA14NES4320003 (Cooperative Institute for Satellite Earth System Studies (CISESS)) at the University of Maryland/ESSIC. Additional support was provided by the National Centers for Environmental Information, National Oceanic and Atmospheric Administration, U.S. Department of Commerce.

References

[1] Peng, G., Ritchey, N. A., Casey, K. S., Kearns, E. J., A Privette, J. L., Saunders, D., & Ansari, S. (2016). Scientific Stewardship in the Open Data and Big Data Era-Roles and Responsibilities of Stewards and Other Major Product Stakeholders. D-Lib Magazine, 22(5-6), 1–1. doi: 10.1045/may2016-peng

[2] Durre, I., Yin, X., Vose, R. S., Applequist, S., & Arnfield, J. (2018). Enhancing the data coverage in the Integrated Global Radiosonde Archive. Journal of Atmospheric and Oceanic Technology, 35(9), 1753-1770.

[3] Wang, Z., NOAA National Centers for Environmental Information (2017). Quality-controlled sea surface marine physical, meteorological and other in situ measurements from the NCEI Surface Underway Marine Database (NCEI-SUMD). NOAA National Centers for Environmental Information. Dataset. https://accession.nodc.noaa.gov/NCEI-SUMD.

[4] Boyer, T.P., O.K. Baranova, C. Coleman, H.E. Garcia, A. Grodsky, R.A. Locarnini, A.V. Mishonov, C.R. Paver, J.R. Reagan, D. Seidov, I.V. Smolyar, K.W. Weathers, M.M. Zweng (2018). World Ocean Database 2018. A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87.

[5] Liu, Z., Zhang, Y., Yu, X., & Yuan, C. (2016). Unmanned surface vehicles: An overview of developments and challenges. Annual Reviews in Control, 41, 71-93.

[6] Wang, Z., Boyer, T., Zhang, H. M., Bayler, E. J., Biddle, M., Baker-Yeboah, S., & Zhang, Y. (2016, December). NCEI-TSG: A Global in situ Sea-surface Salinity and Temperature Database of Thermosalinograph (TSG) Observations. In 2016 AGU Fall Meeting. AGU.

[7] Smith, S. R., Briggs, K., Bourassa, M. A., Elya, J., & Paver, C. R. (2018). Shipboard automated meteorological and oceanographic system data archive: 2005–2017. Geoscience Data Journal, 5(2), 73-86.

[8] de la Villéon, L. P., Delcroix, T., Keeley, R., Carval, T., & Maillard, C. (2006). GOSUD: Global Ocean Surface Underway Data Project. European Operational Oceanography: Present and Future, 142.

[9] Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., Reagan, J. R., Seidov, D., Weathers, K., Paver, C. R., and Smolyar, I. (2018). World Ocean Atlas 2018, Volume 1: Temperature. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 81, 52 pp.

[10] Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., Garcia, H. E., Mishonov, A. V., Baranova, O. K., Weathers, K., Paver, C. R., and Smolyar, I. (2018). World Ocean Atlas 2018, Volume 2: Salinity. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 82, 50 pp.

[11] Talley, L. D., Feely, R. A., Sloyan, B. M., Wanninkhof, R., Baringer, M. O., Bullister, J. L., ... & Gruber, N. (2016). Changes in ocean heat, carbon content, and ventilation: a review of the first decade of GO-SHIP global repeat hydrography.

[12] Gaillard, F., Reynaud, T., Thierry, V., Kolodziejczyk, N., & Von Schuckmann, K. (2016). In situ–based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height. Journal of Climate, 29(4), 1305-1323.

[13] Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., ... & Zweng, M. M. (2012). World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. Geophysical Research Letters, 39(10).

[14] Boyer, T., Domingues, C. M., Good, S. A., Johnson, G. C., Lyman, J. M., Ishii, M., ... & Church, J. A. (2016). Sensitivity of global upper-ocean heat content estimates to mapping methods, XBT bias corrections, and baseline climatologies. Journal of Climate, 29(13), 4817-4842.

[15] Manley, J., & Willcox, S. (2010, May). The wave glider: A persistent platform for ocean science. In OCEANS'10 IEEE SYDNEY (pp. 1-5). IEEE.

[16] Mordy, C. W., Cokelet, E. D., De Robertis, A., Jenkins, R., Kuhn, C. E., Lawrence-Slavas, N., ... & Stabeno, P. J. (2017). Advances in ecosystem research: Saildrone surveys of oceanography, fish, and marine mammals in the Bering Sea. Oceanography, 30(2), 113-115.

[17] Meinig, C., Lawrence-Slavas, N., Jenkins, R., & Tabisola, H. M. (2015, October). The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance. In OCEANS 2015-MTS/IEEE Washington (pp. 1-6). IEEE.

[18] Nicholson, J. W., & Healey, A. J. (2008). The present state of autonomous underwater vehicle (AUV) applications and technologies. Marine Technology Society Journal, 42(1), 44-51.

[19] Goodman, L., & Wang, Z. (2009). Turbulence observations in the northern bight of Monterey Bay from a small AUV. Journal of Marine Systems, 77(4), 441-458.

[20] Webb, D. C., Simonetti, P. J., & Jones, C. P. (2001). SLOCUM: An underwater glider propelled by environmental energy. IEEE Journal of oceanic engineering, 26(4), 447-452.

[21] Eriksen, C.C., T.J. Osse, R.D. Light, T. Wen, T.W. Lehman, P.L. Sabin, J.W. Ballard, and A.M. Chiodi (2001). Seaglider: A long-range autonomous underwater vehicle for oceanographic research, IEEE J. Oceanic Eng., 26(4), 424-436.

[22] Wang, Z., DiMarco, S. F., & Polzin, K. (2018). Turbulence observations in the Northern Gulf of Mexico from a Slocum glider. In 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO) (pp. 1-6). IEEE.

[23] Rudnick, D.L., R.E. Davis, C.C. Eriksen, D.M. Fratantoni, and M.J. Perry (2004). Underwater gliders for ocean research, Mar. Tech. Soc. J., 38(2), 73-84.

[24] Sherman, J., R.E. Davis, W.B. Owens, and J. Valdes (2001). The autonomous underwater glider "Spray". IEEE J Oceanic Eng., 26(4), 437-446.

[25] UNESCO (2020). COVID-19 disruptions in ocean observations could threaten weather forecast and climate change predictions. IOC/UNESCO Press Release. https://en.unesco.org/news/covid-19-disruptions-ocean-observations-could-threaten-weather-forecast-and-climate-change

[26] Meyssignac, B., Boyer, T., Zhao, Z., Hakuba, M. Z., Landerer, F. W., Stammer, D., ... & Abraham, J. P. (2019). Measuring global ocean heat content to estimate the Earth energy imbalance. Frontiers in Marine Science, 6, 432.