An introduction and initial assessment of Uncrewed Systems Standards as a catalyst for data interoperability

Andrew Evans
Commander, Naval Meteorology and Oceanography Command (CNMOC)
Stennis Space Center, MS

Katharine Weathers
NOAA's National Centers for Environmental Information
Stennis Space Center, MS

***Abstract*—Uncrewed Underwater, Surface and Aerial Systems (UxS) are increasing in use and capability in the marine sectors. With much of the new emphasis on UxS sensor development and higher resolution data, data challenges have arisen to include findability, accessibility, interoperability, and reusability - the key to which is metadata. NOAA and Navy are developing a logical metadata model to serve as a base for unifying UxS data interoperability in the marine domain. The Logical Metadata Model has been presented to a mix of industry, government, and academic partners. It is constructed around modular components from different stages and aspects of the data lifecycle. These components include Sensor Metadata that feed into a Vehicle (platform) Metadata component that documents platform and sensor configuration, calibration, and data collection parameters. These components are assembled into the model's Core Metadata, which is comparable to a Mission Report; to create one, complete ISO 191** series metadata record paired with the data in a repository. Following the path of the data lifecycle, Curated Metadata is created once the raw data is processed and products are created. The Data Acquisition Metadata feeds into this Curated Data Metadata along with information such as the file names, file types, keywords, and usage constraints. By linking Data Acquisition and Processed metadata to each other, as well as to the scientific data, data become more discoverable and users can quickly leverage new analytical tools and algorithms as they become operational.**

Keywords—uncrewed systems; autonomous vehicle; UxS; data management; metadata; interoperability

I. Introduction

The Naval Meteorology and Oceanography Command (NMOC) - the Department of Defense's leading meteorological and oceanographic data provider - and the National Oceanic and Atmospheric Administration (NOAA) are two of the nation's leading authorities for environmental and oceanographic information. NOAA's National Centers for Environmental Information (NCEI) houses petabytes of data while striving to make research publicly discoverable, accessible, and reusable through domestic and international partnerships. NMOC's mission is to define the physical environment from the ocean to the space in support of the U.S. Navy and its international and domestic partners. NCEI has developed tools and best practices for International Organization for Standardization (ISO) metadata management and long-term preservation. Implementing metadata standards for uncrewed systems would enhance interoperability. This would lead to improved real-time data utility and long-term reusability, allowing for UxS data to answer today's environmental problems and tomorrow's climate predictions. Working together with the Commander, Naval Meteorology and Oceanography Command (CNMOC) under the Commercial Engagement Through Ocean Technology (CENOTE) Act of 2018, we are actively developing data standards and metadata requirements. Uncrewed systems present several data specific challenges ranging from data handling and management, multitude of different sensor and platform combinations, and efficient storage of data collected from a multi-sensor platform. As part of CENOTE, the Department of the Navy and NOAA to ensure agencies are effectively, efficiently, and responsibly exchanging data collected by unmanned maritime systems. The primary goal of this paper were to establish a set of metadata requirements and data standards that would improve and automate as much of the metadata collection as possible and develop a uniform data model. Metadata and data standardization will be critical as data repositories continue to grow in order to make data discoverable, searchable, and accessible. Standardization of file formats and data will be critical in the coming decades for machine-searchable data and utilization of AI/ML technologies on UxS data.

For decades, NOAA and the Navy have used uncrewed systems to collect physical oceanographic data and other data ranging from bathymetry and acoustics to synthetic aperture sonar data with modern sensors. UxS began with the Special Purpose Underwater Research Vehicle (SPURV) in the 1950s to collect data in the Arctic. SPURVs collected much of the same physical oceanography data that is collected today by using a vertical rake of temperature and conductivity sensors. From SPURV, UxS have grown from heavy and expensive equipment largely relegated to academic and government institutions to a multifaceted tool used for wide range of commercial, scientific and military applications. UxS size and performance have also continued to evolve from limited duration mini-submersibles requiring cranes and other specialized equipment to mission specific platforms that range from man-portable to sizes larger than some early manned submarines. Payloads carried on board UxS have continued to evolve from oceanographic sensors to a full suite of communication, sonar, and environmental sensors in addition to vehicle health monitoring systems. As platforms have continued to evolve collecting more data, operating for longer periods, carrying bigger payloads, and increasing autonomy, the need for standardization of metadata and format of collected data becomes more important.

Data standardization ensures that products and services derived from those data are both consistent and decipherable by others. While nautical charting and hydrographic standards have existed for as long as people have been mapping the world's oceans, it was not until the early 20th century that an international organization was formed to unify these standards. The International Hydrographic Organization (IHO) was founded as the International Hydrographic Bureau in 1919 for the development of hydrographic and nautical charting standards, which have been almost universally adopted. The IHO's mission established coordination of hydrographic activities between international organizations such as International Organization for Standardization, and national offices, in order to establish uniformity in nautical charts and documents, and methodology for carrying out and exploiting hydrographic surveys to support the safety of navigation and protection of the marine environment.

The International Organization for Standardisation (ISO) [1] was formed in 1946 and is comprised of standards organizations from 130 countries. In response to growing use of geographical information systems (GIS), ISO established Technical Committee 211 (ISO/TC211) in 1994 to develop standards for information concerning geographic objects and phenomena. In order to harmonize standards between the IHO, ISO and other organizations such as the NATO Digital Geographic Information Working Group (DGIWG) a formal agreement was established at the Committee on Hydrographic Requirements for Information Systems (CHRIS) meeting in 2000. This led to the development of IHO series S-100 standards which are aligned with the ISO 19100 series of standards for geographical information.

This document seeks to establish a baseline of best practices in order to align the UxS community with ISO 19100, IHO S-100, and other established standards for the collection of oceanographic and geospatial data. This baseline builds on metadata and data standards currently followed by groups in the UxS community and the proposed standardization will align them with the data strategy of the IHO and other national organizations. Establishing a common standard for UxS data collection and curation will foster improved interoperability, discoverability, and accessibility of the acquired data. The UxS metadata and data standard are a critical first step to achieving near real time data dissemination from semi-autonomous and fully autonomous collection platforms.

II. The Logical Data Model

The logical data model presented here (Figure 1) is a modular two-phase approach that is aligned to the lifecycle of UxS data. It is split between the data collection phase and the data curation/ archival stage in order to best capture both the scientific and engineering details of the data collection and to maintain a link to that information throughout the life of the data in the curation or archival phase.

A. Data Acquisition Metadata

The metadata for Data Acquisition (see Figure 2) focuses on capturing the necessary information of data acquisition or data creation with the goal to make data discoverable by the collection organization, location, type of data or sensor, and mission specific parameters. It consists of four sub groupings: core metadata, mission metadata, platform metadata, and sensor metadata. In an HDF-style file, these components can be stored as single XML file to accompany the data or packaged with the data as a character string attribute.

The core metadata assigns a unique identifier, and primarily provides a general overview of the data collected. This would be the overall point of contact, greater mission objective, and keywords. It also includes a citation for specification of the metadata convention used.

The mission metadata creates an in-depth record of who was collecting the data, where their mission area was, which organization was responsible for the mission, and any associated

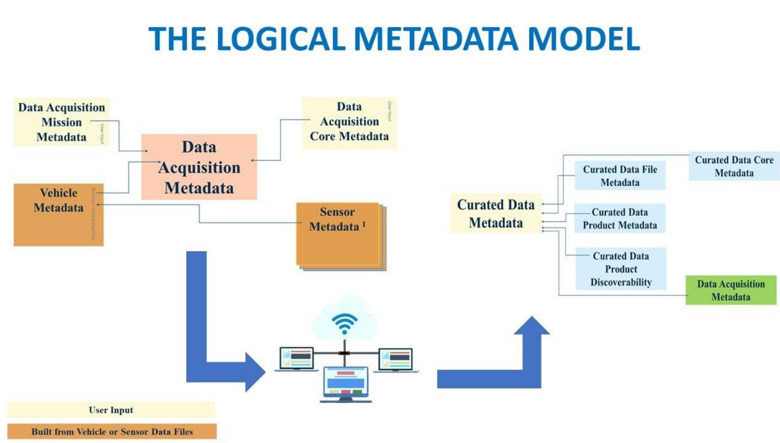

Fig. 1. A schematic of the Logical Metadata Model.

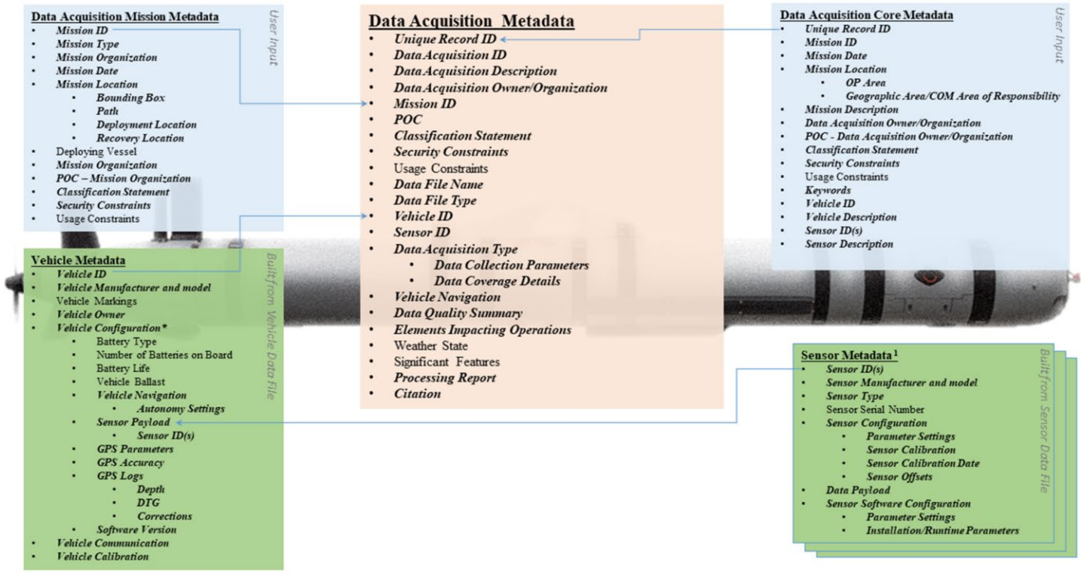

Fig. 2. A schematic Data Acquisition portion of the Metadata Model. Mandatory fields are in bold and italicized font.

security or classification constraints on the mission portion of the data collection.

The intention of the platform metadata and sensor(s) metadata are to capture the description and configuration details regarding identification, logs of missions, parameters for data collection such as units of measurements and coordinate system used to geo-reference the data, and calibration settings specific to both the sensor and the platform used. The sensor portion is subset from the platform portion so that can be repeatable for each sensor carried on the platform. The platform and sensor metadata sub-headings are based on the information collected in the platform and sensors configuration files.

B. Data Curation Metadata

The metadata for Data Curation (see Figure 3) focuses on capturing both the data acquisition metadata in addition to any processing information, attributed keywords or themes, and production information. This intended for the end user to be able to identify quickly whether that dataset is interoperable

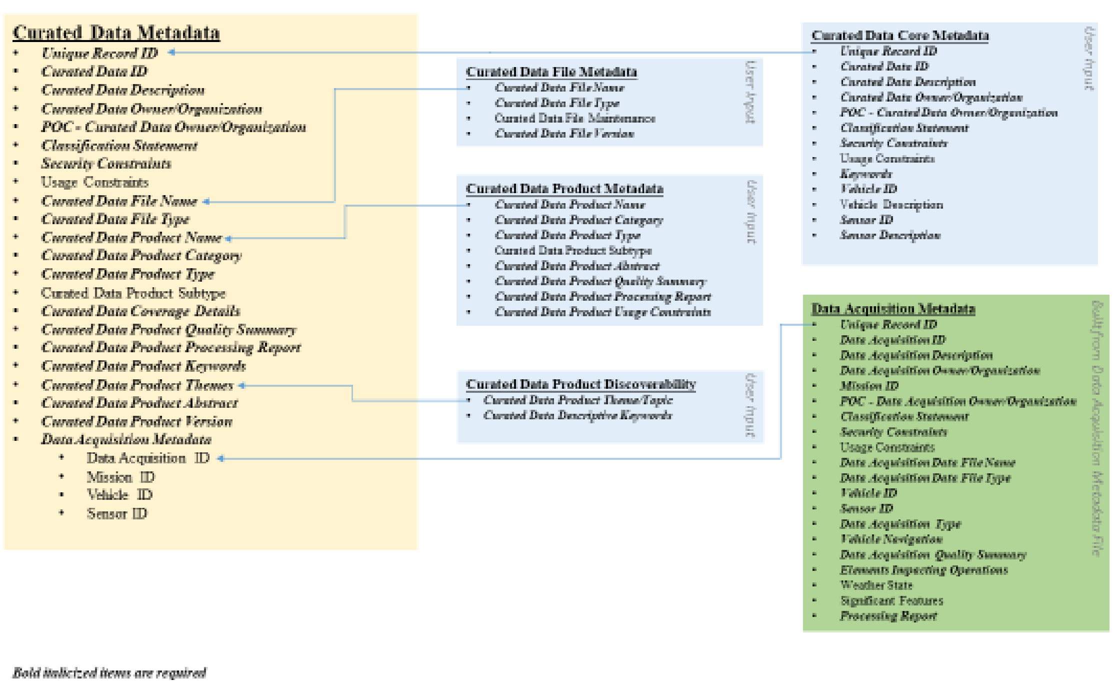

Fig. 4. A schematic of the Curated Data portion of the Metadata Model. Mandatory fields are in bold and italicized font.

with their system and applicable to their mission by linking it back to the acquisition process and platforms or sensors used. The metadata for data curation contains both a unique identifier for each dataset product and a citation to the identifier from the acquisition metadata in order to maintain a lineage of each product from the time of collection through processing and dissemination. The data curation metadata also includes standard metadata information such as the data owner, data abstract, and security constraints. Qualitative information such as the file types, product category, and quality summary are included in order to align with currently published metadata standards, enable AI/ML, and assist the end user in finding the most appropriate data for their needs.

C. An XML Template

An xml template has been created for use with uncrewed underwater systems data. This template could be used for either data acquisition or data curation metadata. This template is a full ISO 19115-2 standard metadata file. Fields can be pre-entered if they are foreseen to be static, such as organization, or can be coded such that the user knows what type of information needs to be input into the field, such as a Character String, Data String (and format), code list, etc. We designed the xml template be able to incorporate portions of other standards such as Sensor Model Language (SensorML) [2] metadata language. SensorML is a component of the Open Geospatial Consortium (OGC) set of Sensor Web Enablement (SWE) standards. Incorporation of the already widely adopted SensorML and SWE standards will enable rapid integration of UxS data into geospatial applications and development of best practices and automation of metadata capture.

With proper coding, automating the capture of the metadata we presented in this paper will both improve compliance and integrate into existing data workflows. Both human and machine can read XML and use XML encoding for data discoverability across systems reducing reliance on specific software. XML is also flexible in that it can be broken into components such as the citation with information about the sponsor organization, geographical bounds of a dataset, or platform specific information. Data users are able to leverage these components to tag datasets, which will improve searchability of data and structuring of data repositories for optimization. The XML template is stored either as a standalone file, which accompanies a dataset, or as a character string variable within an HDF data file to ease data transmission and transferal.

III. Use Case

In July of 2022, a small test case was completed with NOAA's National Ocean Service (NOS) in conjunction with the Mesophotic and Deep Benthic Communities (MDBC) Project. A REMUS 600 was launched and recovered collecting Synthetic Aperture Radar (SAR) data. Pre-mission, metadata fields were mapped from the ISO 19115-2 template to a database for integration during and post mission. Post-mission both scientific and engineering data were captured and reviewed. The data does currently validate within the ISO 19115-2 template.

While this pilot showcased that the template does capture the fields of interest to scientists and engineers of uncrewed systems, there are clear next steps. Another test is planned for October 2022. The database with new fields specified by the template will be in use. There is also hope that the vehicle health data can be more easily captured by use of non-propriety formats exported from the REMUS 600 software. Use of components and standardized vocabulary will better enable translation from the database to a valid metadata record as well.

Refining these issues will lead to a workflow that is optimized for the goal of testing an automated metadata delivery.

IV. Conclustion

While this Logical Metadata Model has received positive feedback from the government maritime representatives, and the uncrewed vehicle and sensor industry that it has been presented to, the continuing development of metadata standards should be considered an evolving process. We are soliciting feedback from engineers on the utility of XML with vehicle health data or if a separate file standard, such as SensorML would be beneficial for that data which is specific to the operators of UxS in real-time. There will be a continuation of real-world testing of the Model in the fall of 2022 where strengths, weaknesses, and feasibility will be assessed. In this talk we are soliciting feedback from the Marine Technology community on the Metadata Model, as well as sharing the lessons learned from the summer assessment.

References

[1] International Organization for Standardization website: http://iso.org

[2] Sensor Model Language (SensorML) website: http://ogc.org/standards/sensorml