# OPEN SCIENCE IN MARINE ROBOTICS

Simona Aracri[1*], Roberta Ferretti[1], Corrado Motta[1], **Fausto Ferreira[2]**, Marco Bibuli[1], Francesca de Pascalis[3], Angelo Odetti[1], Gabriele Bruzzone[1], Massimo Caccia[1]

[1] CNR-INM, Uos Genoa, Italy
[2] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[3] National Research Council Institute of Marine Sciences, Italy

*correspondence E-mail: simona.aracri@cnr.it

**Key words:** Marine Robotics; Open Science; FAIR data; Environmental Monitoring; Emerging Technology

## 1. Introduction

The concept of Open Science translates in sharing knowledge for progress and for the general public benefit. In essence, doing Open Science means sharing every step of the research process, from data, to protocols, to software, to results. The ultimate goal remains to share knowledge as quickly, broadly and effectively as possible.

The benefits of Open Science are multiple; for science itself, which becomes more transparent, verifiable, reproducible as well as faster and efficient, and therefore contributes to an acceleration in the process of creating the knowledge; for companies, which can take advantage of research results and, combining them with their specific expertise, offer more innovative products; for the whole society, in particular citizens and administrators, who can make decisions more objectively based on data (Fritz et al., 2019) and their scientific interpretation.

Since data are essential to the scientific process, Open Science efforts have focused on making data more openly available. Open data are data that may be accessed, used, and shared for any purpose without restrictions. However, making open data available is not enough. It has to be done in an intelligent way, for example making data available in an open repository and in a nonproprietary, standardized format. Whenever possible, a DOI should be created for the data together with a clear licensing information along with use constraints about the data (Ramachandran et al., 2021). All these good practices are part of the more general FAIR guiding principles for scientific data management (Wilkinson et al., 2016). Following FAIR principles, gathered data can be used in multiple fields of science not necessarily only within the collecting community, which is more likely to share the same vocabulary and background knowledge. Without FAIR research data, open science is simply impossible (Burgelman et al., 2019).

### 1.1. The key role of metadata

Research data needs metadata to become Findable, Accessible, Interoperable and Reusable (FAIR), by humans and machines (Wilkinson et al., 2016). The first step to render data FAIR is to establish a set of descriptive metadata, i.e. information to make the dataset findable by the wider scientific community. Usually includes information such as title, author, subjects, keywords, publisher, urls and are typically domain agnostic. Different standards are already consolidated for the descriptive metadata such as ISO19115 (ISO19115 standard), Dublin core (DCMI, 2022), and DataCite (DataCite Schema) among others.

In marine robotics, while there are a few efforts in standardization (Waldmann et al., 2021) mostly these are either for industrial applications such as Remotely Operated Vehicles (ROVs) (NORSOK Standard), military-originated (NATO, STANDARD ANEP-87) or simply best practice community efforts (Hörstmann et al., 2020). Moreover, these standardization efforts are related to the construction or operation of robots or their scientific instrumentation. On the other hand, much less attention has been paid to the need for metadata standards and their importance as a cornerstone for open data. While there are a few efforts such as Marine Regions (https://www.marineregions.org) for what concerns georeferencing of marine areas (e.g. for trials) or the vocabularies defined in the NERC Vocabulary Server (The NERC Vocabulary Server (NVS)), most marine robotics research groups are not using shared metadata standards (sometimes not even ISO 8601 (ISO, 2022) for Date & Time). This is detrimental to fruitful collaborations among different research groups as well as to the re-use of published open data. The Robotic Operating System (ROS) has become a standard middleware among the marine robotics community, but this does not include any metadata standardization and different groups attribute different terminologies to the same data (and metadata). Thus, a collaborative effort is needed to define shared and easy adoptable metadata standards and a shared vocabulary in the field of marine robotics. Without these, open data will suffer from lack of findability and interoperability violating two of the FAIR principles. This need is exacerbated by the open data mandates by several funding institutions such as the European Commission or UK Research and Innovation. Therefore, rather sooner than later the marine robotics community needs to address the lack of common metadata, shared vocabulary and data format to open their data to the public. One of the few examples available of a marine robotics open dataset properly documented in a data paper can be found in (Bernardi et al., 2022).

Working to set the backbone for marine robotics FAIR data protocols feeds into the broader ideal of creating interdisciplinary and international science. This concept is strongly supported in the European Open Science Cloud (EOSC) (EOSC Future H2020 Grant) resource, which, especially for Earth Science, is extended by projects such as RELIANCE (RELIANCE Grant H2020). In fact, RELIANCE aspires to manage the research lifecycle using a novel tool: research objects.

## 2. Outcome

### 2.1. Domain agnostic guidelines

A field deployment using a novel marine robot produces a set of data that can be divided into two macro groups, environmental and robotic. If the deployment has a robotic focus we tend to neglect the environmental data, consequently hindering its quality; on the other hand if the deployment has an observational meaning the robotic data are not stored, and if they are this happens following internal practices, which are highly variable depending on the research group. We propose a unified, yet, customised approach, as autonomous as possible, to value both data groups, rendering them mutually validating.During a given data campaign, with an innovative or traditional robotic platform, the descriptive metadata, registered for the environmental and for the device data are the same. The differences lay in the content of the metadata, i.e. data convention applied, name of the collecting platform operator and so on, but the fields populating the descriptive metadata are the same (**Figure 1**). Hence, being descriptive metadata domain agnostic, marine robotics can uptake the standards and conventions already available, for instance ISO19115. Nowadays the most reliable and stable format to share and store data is NetCDF - Network Common Data Form (Network Common Data Form, UCAR).

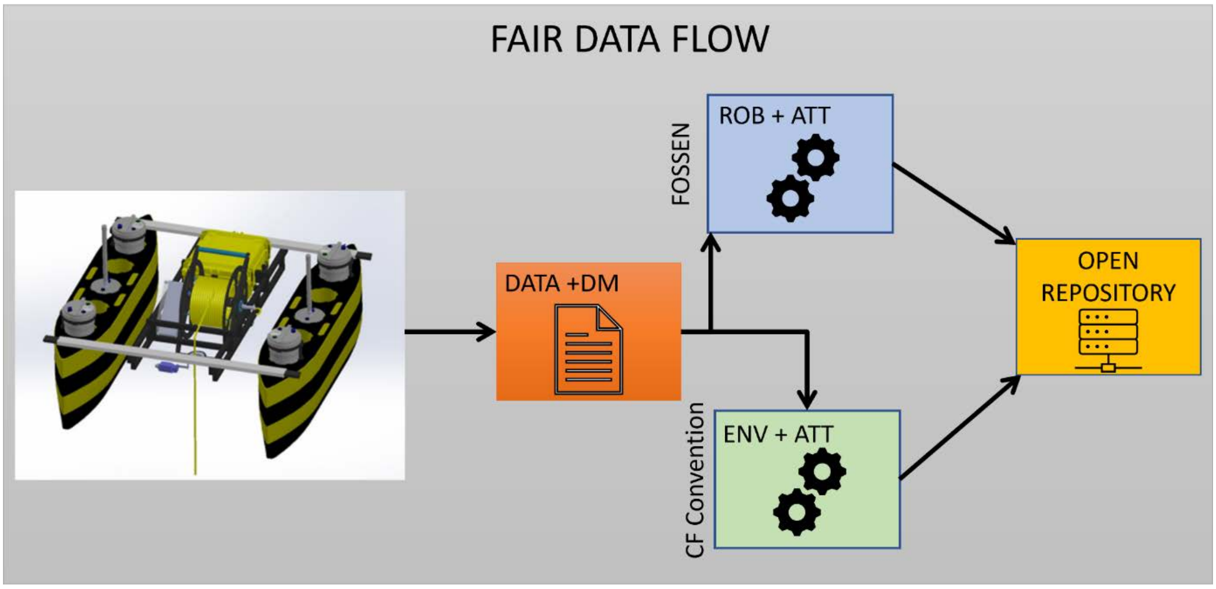


**Figure 1.** FAIR Data flow, from collection with a novel marine robotic device (SWAMP). Initially descriptive metadata (domain agnostic) are compiled for the whole dataset, then the variable specific metadata are compiled separately for robotic and environmental data, referring respectively to Fossen based and CF shared vocabulary. The joint dataset can then converge in a standardised format - e.g. NetCDF - and shared in an open repository.

### 2.2. Domain specific guidelines

The metadata describing the measured variables in Earth sciences are set, coming from a long tradition. It is also not uncommon to have to homogenise data coming from very different platforms, especially when studying long time series, which strongly encourages the implementation of data standards. In line with it, the CF (Climate and Forecast) (Hassell et al., 2017) Metadata Conventions offers a set of guidelines and, more importantly, a shared vocabulary to store and describe variables following the FAIR principles. Hence, when it comes to populating a NetCDF file following FAIR data principle the steps to follow are clear. Trying to do the same with marine robotic data is not straightforward. In fact, to the best of our knowledge, the only reference document is the book Guidance and Control of Ocean Vehicles by Fossen (1994). Merging the Fossen nomenclature and the data policies coming from Earth sciences, **Figure 1** shows the workflow to undertake to produce a combined FAIR dataset during a field campaign employing a novel robotic platform, such as SWAMP (Shallow Water Autonomous Multipurpose Platform) (Odetti et al., 2020).

## 3. Discussion

As shown, the need for metadata in marine robotics is urgent and an absolute necessity for open data. The work presented here is a step towards standard holistic data protocols joining environmental and marine robotic data. However, it needs to be shared and disseminated among the marine robotic community. One possibility to be explored is through the IEEE Oceanic Engineering Society (OES) Standards Committee. This Committee can be a good avenue to involve a broader community and make sure that any metadata standardization efforts are commonly agreed and adopted. By doing so, the findability and interoperability of marine robotics datasets will improve and open data will become more valuable. With an explosion in marine robotics applications and in the number of research groups advancing the state of the art and collecting large amounts of data, this need becomes even more important. Conveying the recommendations hereby presented will ultimately enhance the FAIR status of cutting edge marine robotics datasets.

**Acknowledgements:** This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060395 - Montenegrin Centre for Underwater Sensor Networks (MONUSEN). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work was supported by the European Regional Development Fund through the Interreg Italy-Croatia InnovaMare project (Partnership ID 10248782). The field campaigns with marine robots would not be possible without the extremely valuable work of the CNR-INM's technical team, thank you Giorgio Bruzzone, Edoardo Spirandelli and Mauro Giacopelli.

## References

Bernardi M, Hosking B, Petrioli C, Bett, J. B., Jones, AI Huvenne, V., Marlow, R., Furlong, M., McPhail, S., Munafò, A. (2022). AURORA, a multi-sensor dataset for robotic ocean exploration. The International Journal of Robotics Research 41(5): 461-469. https://doi.org/10.1177/02783649221078612

Burgelman, J-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., Schouppe, M. (2019). Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. Frontiers in Big Data 2: 43. https://doi.org/10.3389/fdata.2019.00043

DataCite Schema (2021). Available at https://schema.datacite.org/

ISO (2022). Date and time format. ISO 8601. Available at https://www.iso.org/iso-8601-date-and-time-format.html

DCMI (2022). Dublin core. Available at https://www.dublincore.org/specifications/dublin-core/

EOSC Future - Advancing European research through Open Science. H2020. Grant agreement ID: 101017536. https://doi.org/10.3030/101017536. Coordinated by ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS. EC signature date: July 5, 2021. Record number: 236671.

Fossen, T. I., (1994). Guidance and Control of Ocean Vehicles 494.

Fritz, S., See, L., Carlson, T. et al. (2019). Citizen science and the United Nations Sustainable Development Goals. Nature Sustainability 2: 922–930. https://doi.org/10.1038/s41893-019-0390-3

Hassell, D., Gregory, J., Blower, J., Lawrence, B. N., Taylor, K. E. (2017) A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1), Geoscientific Model Development 10: 4619–4646, https://doi.org/10.5194/gmd-10-4619-2017

Hörstmann, C., Buttigieg, P.L., Simpson, P., Pearlman, J., Karstensen, J. Waite, A.M. (2020) Towards a Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation. Paris, France, UNESCO, 33. http://dx.doi.org/10.25607/OBP-781

ISO19115 Geographic information — Metadata — Part 1: Fundamentals. Available at https://www.iso.org/standard/53798.html

Marine Regions. Managed by Flanders Marine Institute. Available at https://www.marineregions.org/. Accessed December 7, 2022.

NATO, STANDARD ANEP-87, (2017). Digital Underwater Signalling Standard for Network Node Discovery & Interoperability. Edition A Version 1. NATO STANDARDIZATION OFFICE (NSO). Available at https://nso.nato.int/nso/nsdd/main/standards?search=ANEP-87

Network Common Data Form (NetCDF) by UCAR/Unidata. https://doi.org/10.5065/D6H70CW6

NORSOK Standard, U-102 (2016). Remotely Operated Vehicle (ROV) Services.

Odetti, A., Bruzzone, G., Altosole, M., Viviani, M., Caccia, M. (2020). SWAMP, an Autonomous Surface Vehicle expressly designed for extremely shallow waters, Ocean Engineering 216: 108205. https://doi.org/10.1016/j.oceaneng.2020.108205

Ramachandran, R., Bugbee, K., Murphy, K. (2021). From open data to open science. Earth and Space Science 8: e2020EA001562. https://doi.org/10.1029/2020EA001562

REsearch LIfecycle mAnagemeNt for Earth Science Communities and CopErnicus users in EOSC (RELIANCE) - Extending the research-enabling capabilities of the European Open Science Cloud. H2020. Grant agreement ID: 101017501. https://doi.org/10.3030/101017501. Coordinated by INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK. EC signature date: December 1, 2020. Record number: 232563

The NERC Vocabulary Server (NVS) (2021). Available at https://vocab.nerc.ac.uk/. Accessed December 7, 2022.

Waldmann, C., O'Reilly, T., Jiang, F., Ferreira, F., Xiang, X., McCummins, R., Pallayilline, V., (2021). About the Value of Standards for Ocean Technology. OCEANS 2021: San Diego – Porto, 1-5. https://doi.org/10.23919/OCEANS44145.2021.9705984

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18