

FAIR data and Marine Robotics

R. Ferretti, S. Aracri, M. Bibuli, G. Bruzzone, M. Caccia, A. Odetti (CNR-INM)

C. Motta (Arcadia SIT)



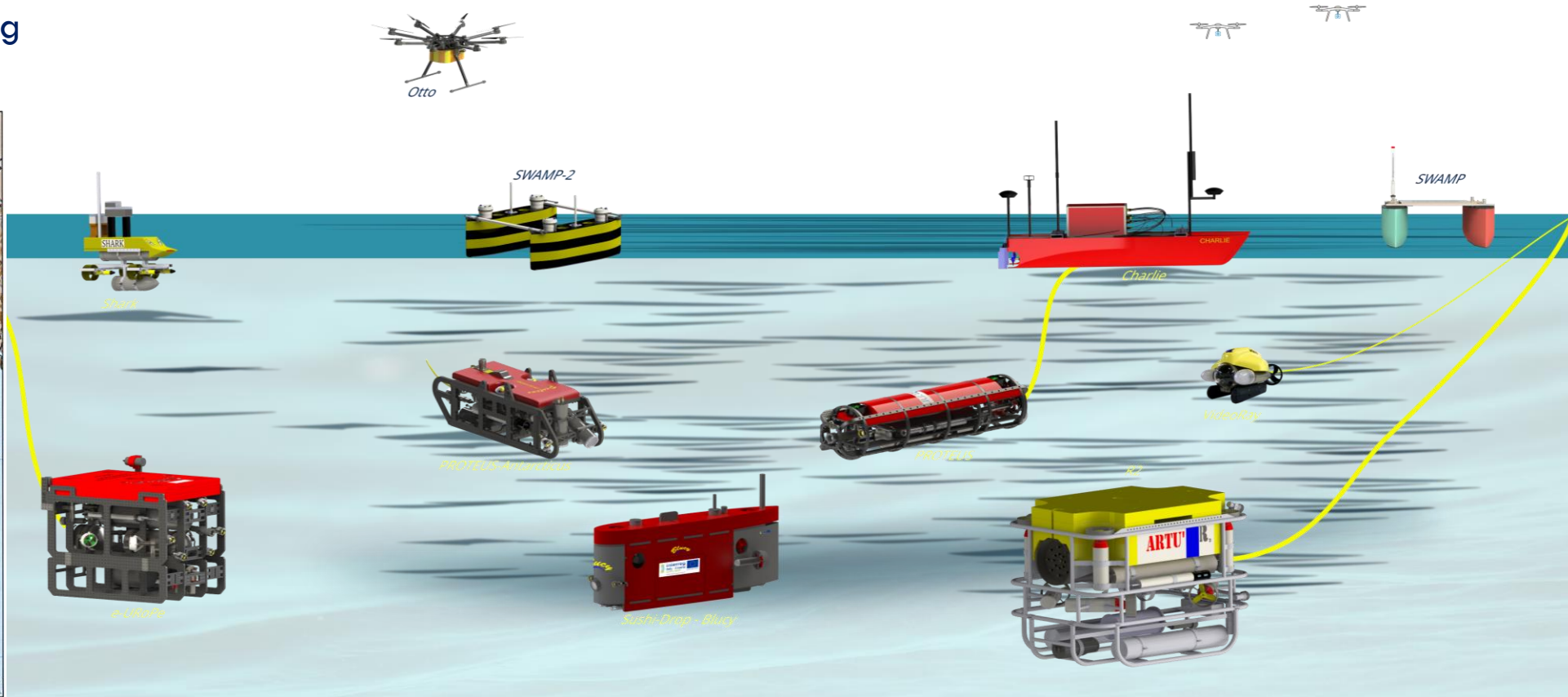
Marine robotics @ CNR-INM Genoa: the fleet



INstitute of Marine engineering
– National Research Council



Roma (Headquarter)
Roma – ARTOV
Genoa
Palermo

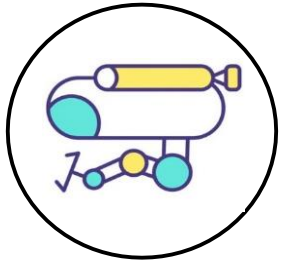


Marine Robots and Tools: Design, Construction, Control
Sensor integration, Field testing, Exploitation

roberta.ferretti@cnr.it



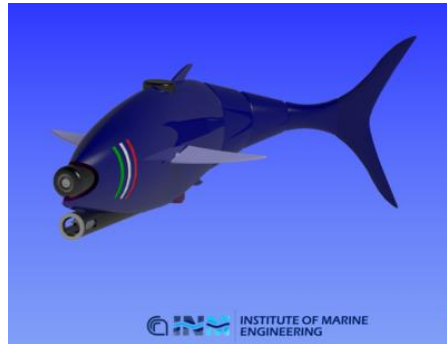
Marine robotics @ CNR-INM Genoa: state-of-the-art



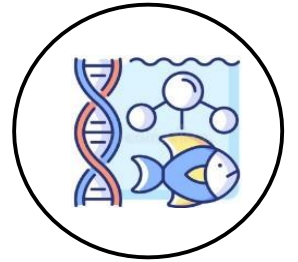
ROBOTICS RESEARCH



Bio-inspired marine robotics PERSICO



Lagoon environmental monitoring

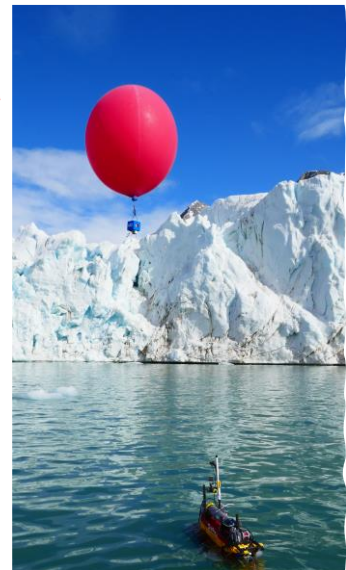


MARINE SCIENCE APPLICATIONS

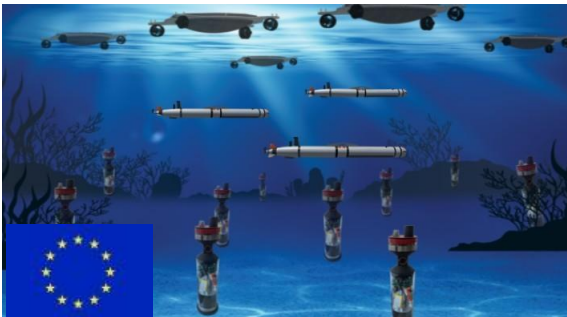


MARINE ENGINEERING INNOVATION

Artica and Antartica missions



Autonomous sensor networks MONUSEN



Coastal tourism, Marina and harbour monitoring

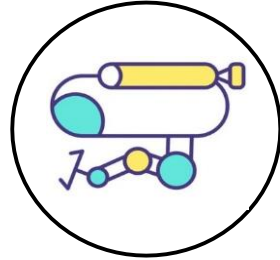


roberta.ferretti@cnr.it





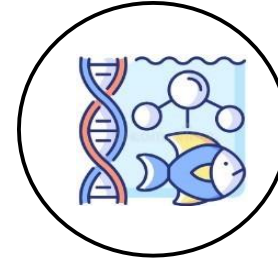
The NEW paradigm: SERVICE-ORIENTED APPROACH



ROBOTICS RESEARCH



MARINE ENGINEERING
INNOVATION



MARINE SCIENCE
APPLICATIONS

INTER-DISCIPLINARITY



MULTIPLE END-USERS

DATA-CENTERED
VISION

A new way of thinking to our research activities from a **data perspective**



Marine
Environmental
data



Marine
Robotic data

- Enhancement in the development of innovative robotics platforms
- Allow experiment replicability and results comparability as well as availability and reuse of dataset for GNC algorithms testing and validation
- Contribute to the global ocean observing effort



2021
2030 United Nations Decade
of Ocean Science
for Sustainable Development



Findable

Metadata and data should be findable for both humans and computers

Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

F

A

I

R

Accessible

Once found, users need to know how the data can be accessed

Reusable

The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

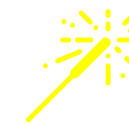


Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

[FAIR guide](#), Nature, March 2016



Designed to support knowledge discovery and innovation both by **humans** and **machines**



What are we working on - The importance of metadata



Descriptive metadata domain agnostic

include information such as title, author, topics, keywords, publisher, URL, etc. (ISO19115 standard)

Global Metadata is the term used to identify descriptive metadata in the NetCDF files.

Use metadata domain specific

clear description of the actual content of the data using standard naming conventions for variables, units, missing values, etc. (CF convention)

Variable metadata is the term used to identify use metadata in the NetCDF files



What are we working on - Findability



We agreed upon a set of minimum **mandatory and optional global attributes** to be used in our datasets → ACDD and ISO19115 compliant

← → ↻ 🔒 corradomotta.github.io/FAIR-Data-in-Marine-Robotics/html/index.html

🏠 FAIR data in marine robotics

Search docs

🏠 » Welcome to FAIR data in marine robotics's documentation! [View page source](#)

📄 FAIR metadata

Global Metadata

Attributes:

📘 Title	<input type="text"/>	keywords :	unmanned marine vehicles,marine robotics,autonomous systems
📘 keywords	unmanned marine vehicles,marine roboti	institution :	CNR-INM
📘 PI name	<input type="text"/>	platform :	SWAMP
📘 PI institution	CNR-INM	title :	Naval maneuver test in Venice
📘 License	Creative Commons	conventions :	ACDD-1.3,CF-1.6
📘 keywords vocabulary	<input type="text"/>	date_created :	2022-10-10T11:56:52.534874
📘 Data center (URL)	<input type="text"/>	summary :	Testing naval maneuver in Venice with SWAMP. Turning and zig-zag data collected.
📘 standard_name_vocabulary	<input type="text"/>	creator_name :	Ferretti Roberta, Bibuli Marco, Motta Corrado
📘 processing_level	<input type="text"/>	product_version :	1
📘 geospatial_vertical_min	<input type="text"/>	project :	INNOVAMARE
📘 geospatial_vertical_resolution	<input type="text"/>	processing_level :	raw data

Path to telemetry:

<https://github.com/CorradoMotta/FAIR-Data-in-Marine-Robotics>



What are we working on – Reusability



We are working to identify a minimum set of **use metadata** that are necessary and sufficient to clearly describe the marine robotic data, enabling the use of these data by potential stakeholders and avoiding a “bad use” due to misunderstanding

xarray.DataArray 'NGC_latitude' (index: 764)

45.44 45.44 45.44 45.44 45.44 45.44 ... 45.44 45.44 45.44 45.44 45.44

Coordinates:

index (index) int64 0 1 2 3 4 5 ... 759 760 761 762 763

Indexes: (1)

Attributes:

source : MicroStrain 3DM-GX5-35

long_name : latitude

standard_name : latitude

units : degree_north

coverage_conten... physicalMeasurement

comment : Latitude is positive northward; its units of degree_north (or equivalent) indicate this explicitly.



<https://github.com/CorradoMotta/FAIR-Data-in-Marine-Robotics>

Use (variable) metadata for marine robotic data is a particularly critical argument open to discussion.





Interoperability both on the **syntactic** and the **semantic** levels is assured by the use of **standard**

HOW STANDARDS PROLIFERATE:



<https://xkcd.com/927/>

Syntactic interoperability: two or more systems to communicate and exchange data



Open and Self describing data format (i.e. NetCDF – CF)



Semantic interoperability: the data is not only exchanged between two or more systems but also understood by each of them



Controlled vocabulary (collection of commonly agreed terms searchable online) and **ontology** (specifications of entities in a domain and their relationships)



What are we working on - Interoperability



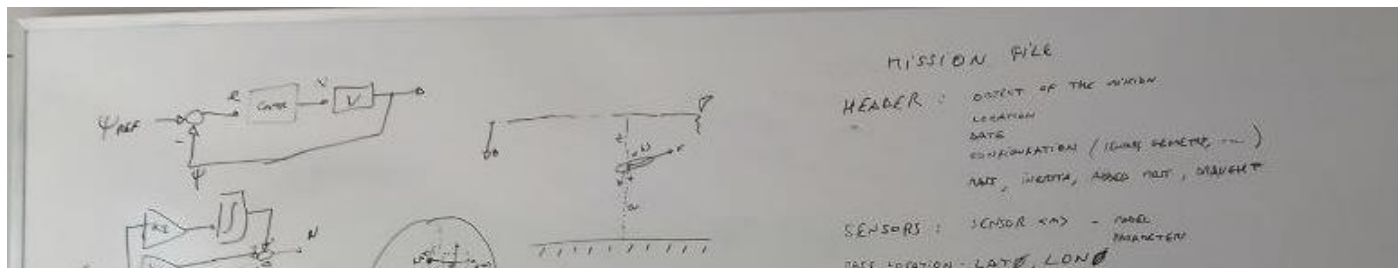
First attempt to define a shared vocabulary for data in marine robotics

CNR-INM-GE-TR-2022-XX

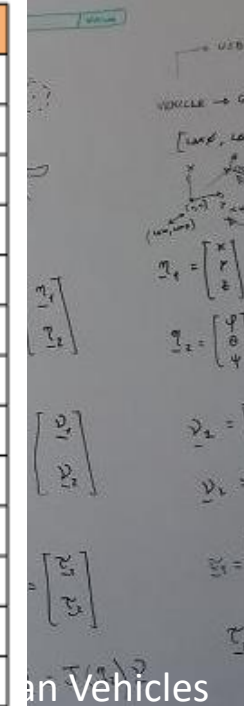
Marine Robotics FAIR Data Nomenclature

Massimo Caccia, Roberta Ferretti, Simona Aracri, Corrado Motta, Marco Bibuli

Abstract



Log Name	Comment	Long Name	Standard Name	Unit	coverage_content_type
date	Date in format...	date			physicalMeasurement
time	Time in format...	time	time	s	physicalMeasurement
latitude	Latitude is...	latitude	latitude	degree_north	physicalMeasurement
longitude	Longitude is...	longitude	longitude	degree_east	physicalMeasurement
xgps	x indicates...	projection_x_coordinate	projection_x_coordinate	m	auxiliaryInformation
ygps	y indicates...	projection_y_coordinate	projection_y_coordinate	m	auxiliaryInformation
roll	Roll rotation...	platform_roll	platform_roll	degree	physicalMeasurement
pitch	Pitch rotation...	platform_pitch	platform_pitch	degree	physicalMeasurement
yaw	Yaw is a...	platform_yaw	platform_yaw	degree	physicalMeasurement
heave_acceleration	Heave...	platform_heave_acceleration_down		m s-2	auxiliaryInformation
lcCtdDepth	Depth is...	depth	depth	m	physicalMeasurement
lcCtdTemperature	Sea water...	sea_water_temperature	sea_water_temperature	degree_C	physicalMeasurement
lcCtdConductivity	Conductivity...	sea_water_electrical_conductivity	sea_water_electrical_conductivity	S m-1	physicalMeasurement



an Vehicles



A framework for FAIR robotic datasets

www.nature.com/scientificdata

Free and Open-Source Software (FOSS) to render marine robotic dataset FAIR-compliant in automated way

- python scripts
- Jupyter notebooks and modules

scientific **data**



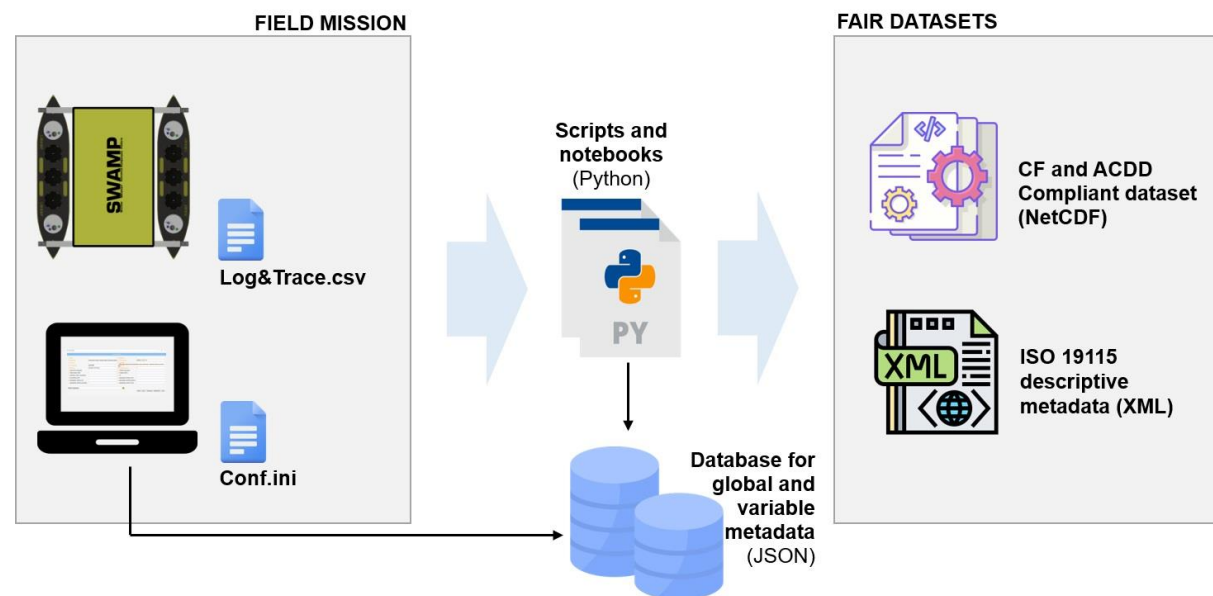
OPEN ARTICLE **A framework for FAIR robotic datasets**

Corrado Motta^{1,2,5}, Simona Aracri^{1,5}, Roberta Ferretti^{1,5}, Marco Bibuli¹, Gabriele Bruzzone¹, Massimo Caccia¹, Angelo Odetti¹, Fausto Ferreira^{1,3} & Francesca de Pascalis⁴

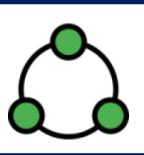
It is essential to publish and make available environmental data gathered by emerging robotic platforms to contribute to the Global Ocean Observing System (GOOS), supported by the United Nations - Decade of Ocean Science for Sustainable Development (2021–2030). The transparency of these unique observational datasets needs to be supported by the corresponding robotic records. The data describing the observational platform behaviour and its performance are necessary to validate the environmental data and repeat consistently the in-situ robotic deployment. The Free and Open Source Software (FOSS), proposed in this manuscript, describes how, using the established approach in Earth Sciences, the data characterising marine robotic missions can be formatted and shared following the FAIR (Findable, Accessible, Interoperable, Reusable) principles. The manuscript is a step-by-step guide to render marine robotic telemetry FAIR and publishable. State-of-the-art protocols for metadata and data formatting are proposed, applied and integrated automatically using Jupyter Notebooks to maximise visibility and ease of use. The method outlined here aims to be a first fundamental step towards FAIR interdisciplinary observational science.



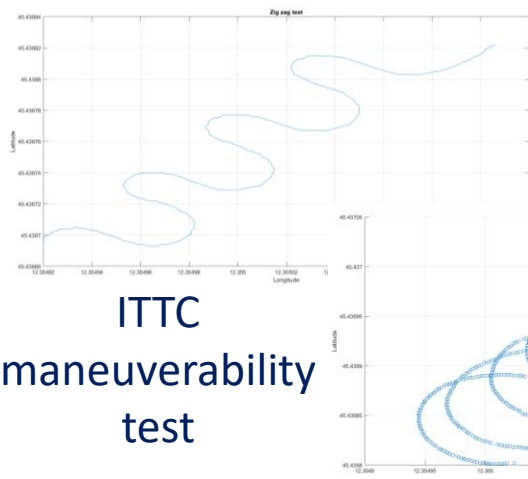
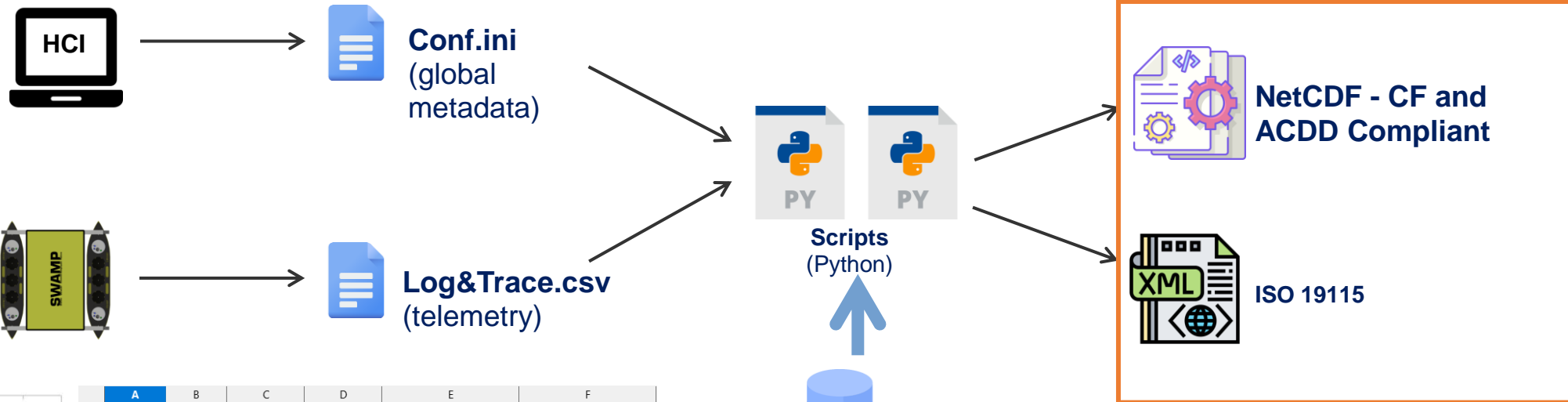
<https://www.nature.com/articles/s41597-023-02495-3>



<https://doi.org/10.5281/zenodo.8256384>



Final Goal: marine robotics data «FAIR by default»



	A	B	C	D	E	F
1	date	time	latitude	longitude	projection_x_coordinate	projection_y_coordinate
2	date	time	latitude	longitude	projection_x_coordinate	projection_y_coordinate
3	20220915	75849250	45.43682500	12.35510700	-57.23	31.66
4	20220915	75849250	45.43682500	12.35510700	-57.23	31.66
5	20220915	75849500	45.43682400	12.35510500	-57.34	31.50
6	20220915	75849500	45.43682400	12.35510500	-57.34	31.50
7	20220915	75849500	45.43682400	12.35510500	-57.34	31.50
8	20220915	75849500	45.43682400	12.35510500	-57.34	31.50
9	20220915	75849500	45.43682400	12.35510500	-57.34	31.50
				1.35510000	-57.79	31.11
				1.35510000	-57.79	31.11
				1.35510000	-57.79	31.11
				1.35510000	-57.79	31.11

Database for global and variable metadata (JSON)

- global_metadata.json
- variable_metadata.json

<https://doi.org/10.5281/zenodo.7825000>

roberta.ferretti@cnr.it





Summary



The Importance of Ocean Understanding: A solid understanding of the ocean's functioning is more critical than ever before. This is driven by initiatives like the **UN Decade of Ocean Science** for Sustainable Development and advancements in marine technology, which provide access to groundbreaking observations.



Role of Innovative Marine Robotics: Innovative marine robots, such as autonomous underwater vehicles (AUVs) and surface vehicles (ASVs), are filling observational gaps in marine sciences. They can collect a growing amount of original data, especially in **challenging environments** where conventional data collection methods are often ineffective. These robots play a crucial role in expanding our knowledge of the ocean.



CNR-INM contribution: The design of innovative robotic platforms and the development of the Free and Open Source Software (FOSS) framework for the implementation of **FAIR principles** in managing **environmental and robotic data** is fundamental. This approach facilitates greater accessibility and sharing of data, improving cooperation and collaboration in the field of marine robotics and contributing significantly to global marine research and conservation efforts.



Credits: @radofoto

Thank you for your attention!

Roberta Ferretti, PhD
National Research Council
Institute of Marine Engineering
Genoa, IT



Questions?

Back up slides

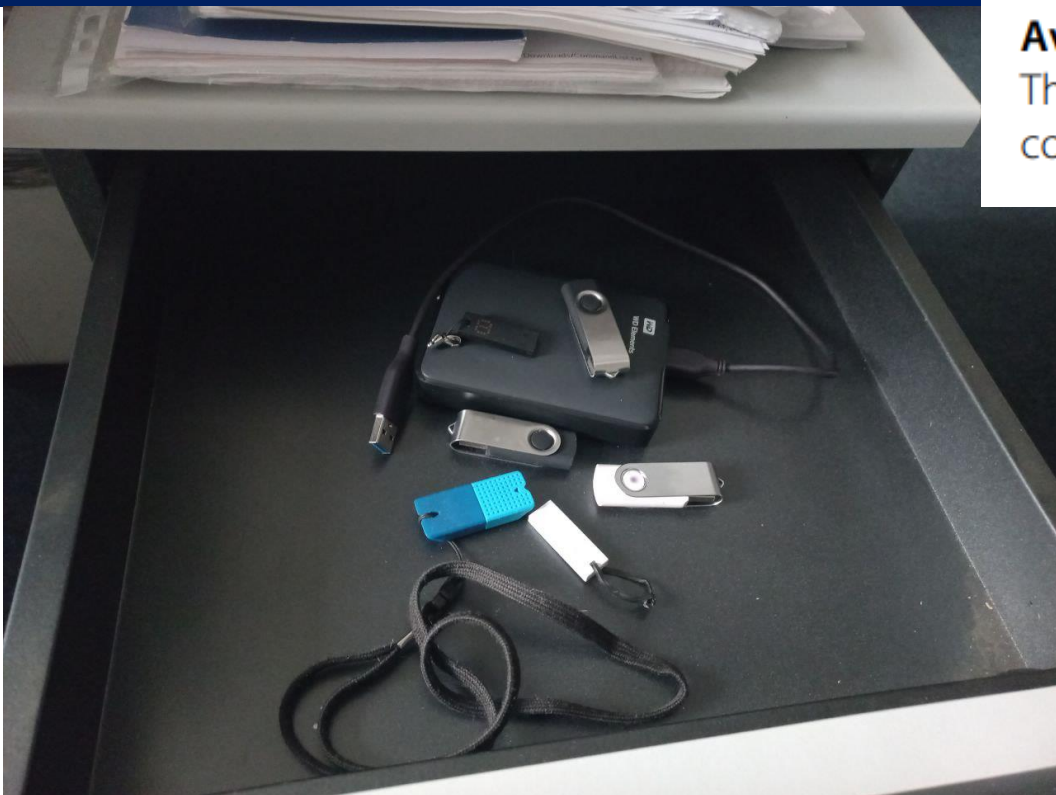


Data Accessibility



Availability of data and materials

The datasets generated or analyzed during our survey are available from the corresponding author upon reasonable request.



Let's try to ask the authors to provide the data

Data requests to authors are successful in about **40%** of cases but...

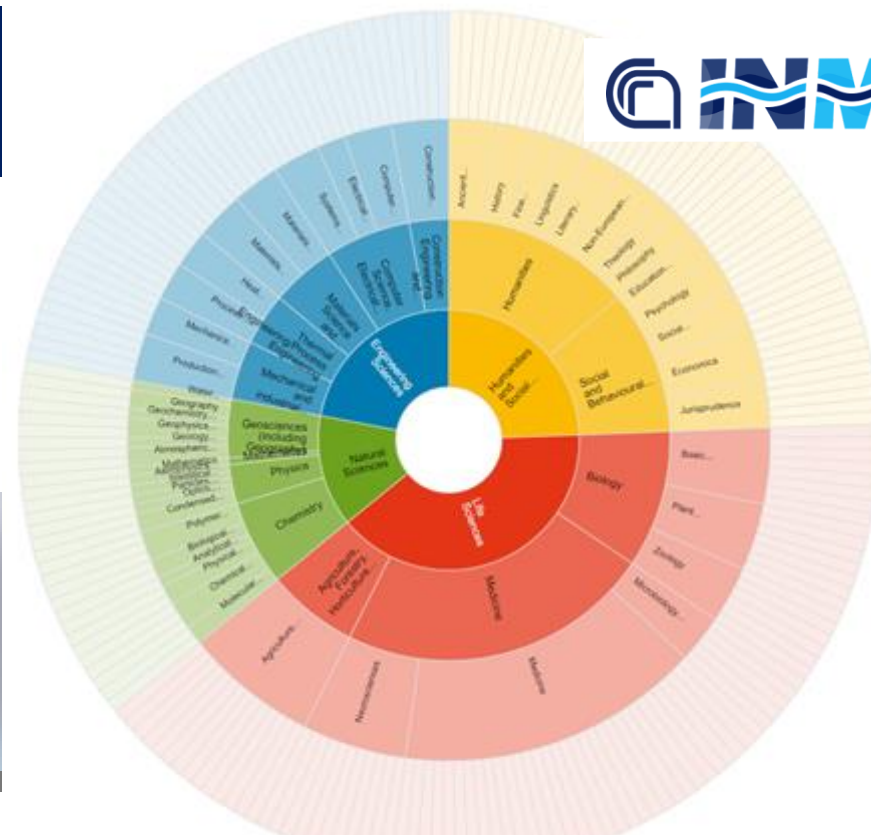
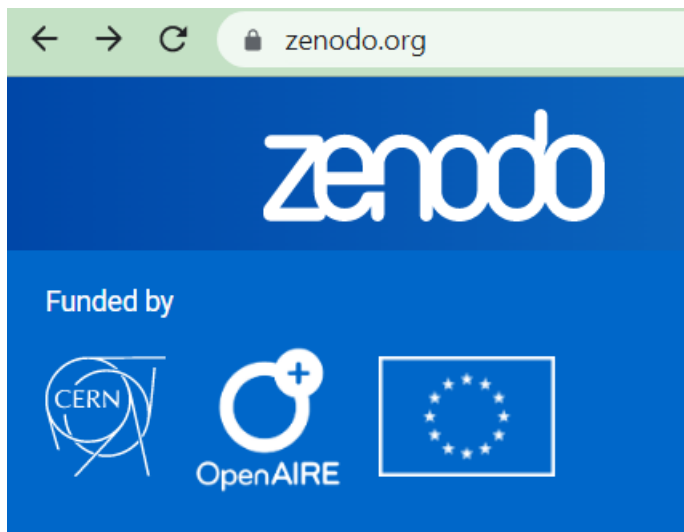


DOI
<https://doi.org/10.1038/s41597-021-00981-0>

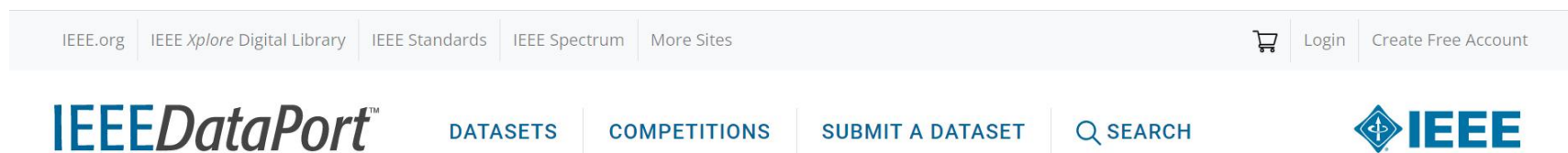


Data Accessibility

Deposit the data in a **trusted repository** or in a **data journal**



<https://www.re3data.org/>





Interoperability both on the **syntactic** and the **semantic** levels is assured by the use of **standard**

HOW STANDARDS PROLIFERATE:



<https://xkcd.com/927/>

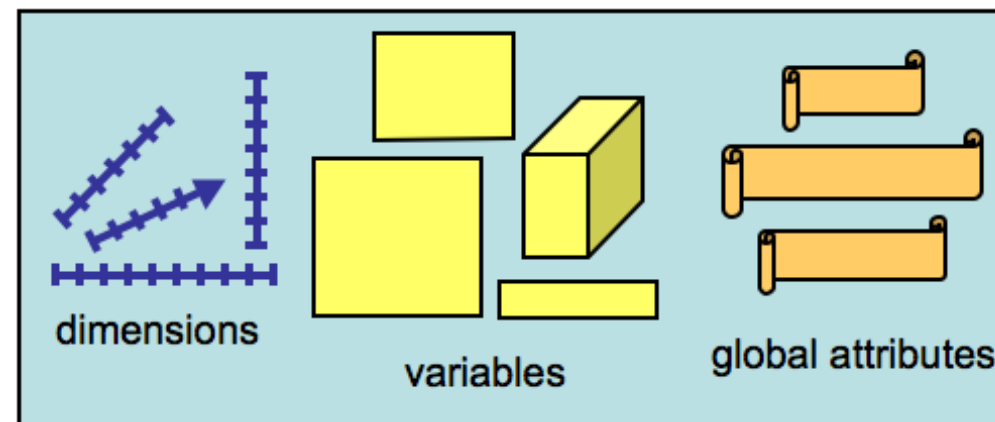
Syntactic interoperability: two or more systems to communicate and exchange data



Open and Self describing data format (i.e. NetCDF – CF)



NetCDF (**Network Common Data Form**) is a set of software libraries and **self-describing, machine-independent** data formats that support the creation, access, and sharing of array-oriented scientific data and metadata.



<https://www.unidata.ucar.edu/software/netcdf/workshops/2010/datamodels/NcFile.html>



Reuse is enabled by

- a clear statement of licence of use

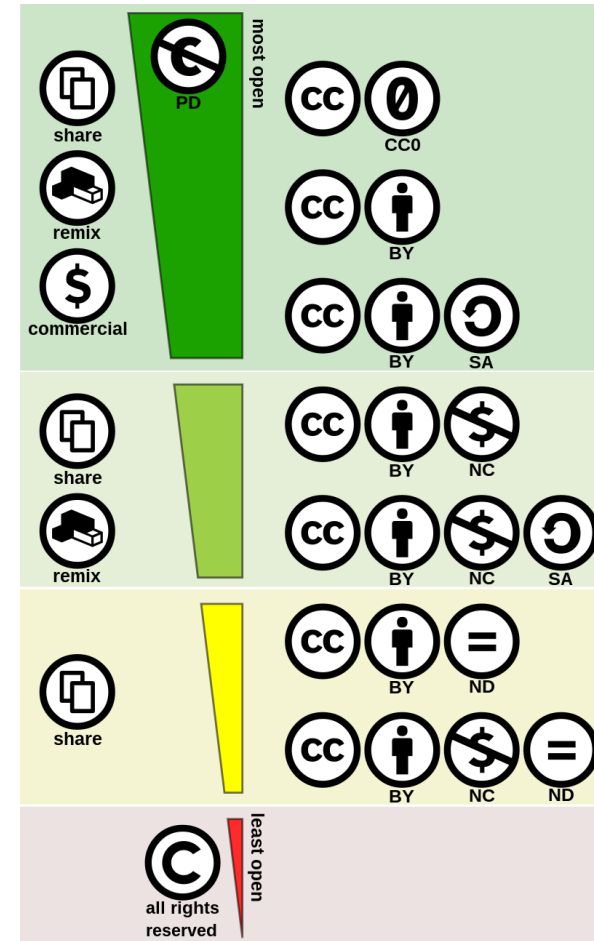
«as open as possible, as close as necessary»

- «use metadata» : clear description of the actual content of data using standard naming conventions for variables, units, missing values, etc...

Use metadata are **domain specific**

? Use metadata for robotic variables

FAIR
≠
OPEN





- Metadata are data providing information about data that make them findable, trackable and (re)usable.
- **Descriptive metadata** usually includes info such as title, author, subjects, keywords, publisher, urls, etc. They are mainly **domain agnostic**.
- Several standards exist: Dublin Core; ISO 19115; DataCite

A **persistent identifier (PID)** is a long-lasting reference to a resource. That resource might be a publication, dataset or person

There are different PID types for different kinds of resources: **DOI** for objects (publications, data, software) and **ORCIDs** for people (researchers, authors, contributors).

Many repositories will assign a PID of the former type when an object is deposited.



SWAMP – Shallow Water Autonomous Multipurpose Platform



Design, development and field testing of a new Autonomous Surface Vehicle for harsh environment

- **Adaptation to critical environments:** Remoteness, Difficult access, Shallow water and High risk for operators and scientists
- **Portability:** lightweight, small dimensions
- **Reconfigurable – Modular - Multipurpose:** able to host and easily integrate different payload for different missions

