

NSF EarthCube Workshop for Ocean Time Series Data, 13th September 2019

FAIR data principles and initiatives

JUSTIN BUCK, NOC/BODC
STEVE DIGGS, UCSD/SIO



Outline

FAIR?



Ongoing
activity &
projects



Ocean
FAIR
data

The FAIR guiding principles for data

FAIR principles

(<https://www.force11.org/group/fairgroup/fairprinciples>)

Findable

Accessible

Interoperable

Reusable

<https://www.nature.com/articles/sdata201618>



National
Oceanography Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL



British Oceanographic
Data Centre
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

FAIR principles

(<https://www.force11.org/group/fairgroup/fairprinciples>)

Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1 the protocol is open, free, and universally implementable.
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

<https://www.nature.com/articles/sdata201618>

A selection of key FAIR developments

- Lorentz workshop (2014), introduction of the term FAIR
- Wilkinson et al. (2016), Scientific data paper
<https://www.nature.com/articles/sdata201618>
- European Commission (2018), turning FAIR into reality
https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
- Wilkinson et al. (2018), GO-FAIR paper on FAIRness metrics
<https://www.nature.com/articles/sdata2018118>
- Nature editorial (2019), Nature backs the Enabling FAIR Data initiative and requires authors to deposit data in community repositories.
<https://www.nature.com/articles/d41586-019-00075-3>
- Tanhua et al. (2019), Ocean FAIR services OceanObs'19 white paper
<https://www.frontiersin.org/articles/10.3389/fmars.2019.00440/full>

Why make ocean data FAIR

“Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.”

Wilkinson et al. (2016)

On-going FAIR activities and projects



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL



**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

International FAIR initiatives and projects related to FAIR



Enabling FAIR data project



[Laura and John Arnold Foundation](#) grant to a coalition of groups representing the international Earth and space science community.

Convened by the [American Geophysical Union \(AGU\)](#), to develop standards that will connect researchers, publishers, and data repositories in the Earth, space, and environmental sciences to enable [FAIR](#) data on a large scale.

Goal to accelerate scientific discovery and enhance the integrity, transparency, and reproducibility of this data.

<http://www.copdess.org/enabling-fair-data-project/>

Enabling FAIR data project – Findable and Accessible

Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1 the protocol is open, free, and universally implementable.
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

Recommendations to achieve findable and accessible:

- Assign DOIs to datasets
- Use ORCID identifiers to identify people in metadata

GO-FAIR

GO FAIR is a bottom-up, stakeholder-driven and self-governed initiative that aims to implement the [FAIR data principles](#)

It offers an open and inclusive ecosystem for individuals, institutions and organisations working together through [Implementation Networks](#).





The Earth is our Lab

Europe's Environmental Research Infrastructures



13 ENV RIs | 37 partners | launch 01 January 2019 | envri-fair.eu

ENVRI

Community Building
Common framework
ENVRI Reference Model

ENVRIplus

Common developments
Shared solutions for RI
Service portfolio

ENVRI-FAIR

Enabling RI for FAIR
FAIR compliant services
Link ENVRI to EOSC



The Earth is our Lab

Europe's Environmental Research Infrastructures



13 ENV RIs | 37 partners | launch 01 January 2019 | envri-fair.eu

ENVRI-FAIR mission

Share experiences and find common solutions to

- data sharing,
- accessibility,
- interoperability

Develop joint services

Expose services to EOSC

Prepare ENVRI-hub



**EUROPEAN OPEN
SCIENCE CLOUD**



ENVRI-FAIR Work package 9

Marine domain research infrastructures



13 ENV RIs | 37 partners | launch 01 January 2019 | envri-fair.eu

Goals:

- Develop the FAIRness of marine RIs
- Liaise with other work packages and domains for common approach and solutions
- WP4 the FAIR landscape
- WP3 data innovation
- WP5 technical solutions

Status:

- Deliverable 9.1 accepted that analyses the current FAIRness of the marine RIs

Ocean FAIR data



**National
Oceanography Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

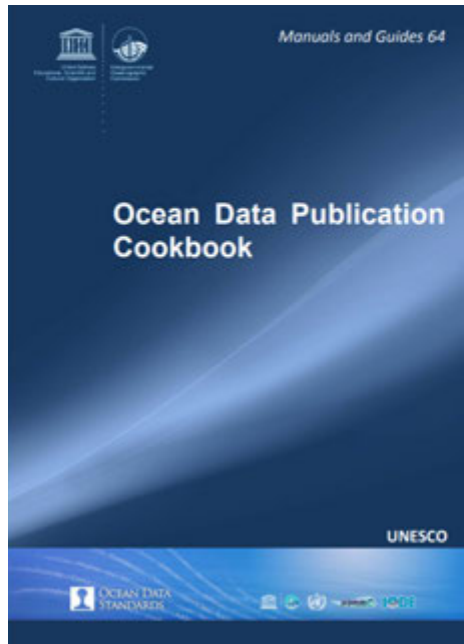


**British Oceanographic
Data Centre**
NATURAL ENVIRONMENT RESEARCH COUNCIL

NERC SCIENCE OF THE
ENVIRONMENT

Making ocean data FAIR - Use of DOIs

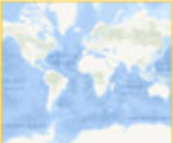

Ocean data publication cookbook Ifremer DOIs on “dynamic data”



Argo part of the integrated global observation strategy

Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)

Click to download the data DATA



Date 2019
Author(s) Argo
Contributor(s) Akosova Fumshko, Akkadadi Turki, Aranda Pascual, **Andre Klein**, Arhan Michel, Ahmadpoora Agus, Babin Marie, Babin Sorin, Balcerano Daniele, **Barringer Matt**, Barre Nicolas, Boutevassian M., Bilbouch Mathias, Bikhhi Pedro Velloz, Bingham Clare, **Bernard Vincent**, **Brig Henry**, Blain Stephane, Bonnel Olaf, Bontius Arjo, Boss Emmanuel, Bourles Bernard, Bowe Amy, Boyer Tim, ...

DOI 10.17824/182
Publisher SEANOE
Keyword(s) float, Argo, global ocean observing system, ocean circulation, in-situ, ocean pressure, sea water salinity, sea water temperature, multi-year, weather climate and seasonal observation, global-ocean

Abstract Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection. The array provides 100,000 temperature/salinity profiles and velocity measurements per year distributed over the global oceans at an average of 3-degree spacing. Some floats provide additional bio-geo parameters such as oxygen or chlorophyll. All data collected by Argo floats are publicly available in near real-time via the Global Data Assembly Centers (GDACs) in Brest (France) and Monterey (California) after an automated quality

Making ocean data FAIR – Interoperable metadata

Vocabularies

- Use of controlled dictionaries for the description of data and metadata
- Adam's talk in this workshop

Also, posters on vocabularies during OceanObs'19

Linked data (Adam's talks)/Sensor Web

- Machine readable data and metadata



- Marine SWE profile
<https://odip.github.io/MarineProfilesForSWE/>

Making ocean data FAIR – Interoperable services

ERDDAP (Kevin O'Brien's talk)

Sensor Web

- Sensor Observation Services



ERDDAP > List of All Datasets

1408 matching datasets, listed in alphabetical order. View page: 1 (current) 2

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Access-ible	Title
	set	data	graph			public	* The List of All Active Datasets in this ERDDAP *
data			graph			public	AMSRE Model Output, obs4MIPs NASA-JPL, Global, 1 Degree, 2002-2010, Monthly
data			graph	M		public	AMSRE Model Output, obs4MIPs NASA-JPL, Global, 1 Degree, 2002-2010, Monthly, L
		data	graph		files	public	AN EXPERIMENTAL DATASET: Underway Sea Surface Temperature and Salinity Aboa
							2007-2010
	set	data	graph			public	Animal Telemetry Network (ATN)
data			graph	M		public	Aquarius Sea Surface Salinity, L3 SMI, Version 5, 1.0°, Global, 2011-2015, 3-Month
data			graph	M		public	Aquarius Sea Surface Salinity, L3 SMI, Version 5, 1.0°, Global, 2011-2015, 7-Day
data			graph	M		public	Aquarius Sea Surface Salinity, L3 SMI, Version 5, 1.0°, Global, 2011-2015, Daily
data			graph	M		public	Aquarius Sea Surface Salinity, L3 SMI, Version 5, 1.0°, Global, 2011-2015, Monthly
data			graph		files	public	Audio data from a local source.
	set	data	graph		files	public	Audio data from a local source.
data			graph	M	files	public	AVHRR Pathfinder Version 5.3 L3-Collated (L3C) SST, Global, 0.0417°, 1981-present, D
							Day Composite)
data			graph	M	files	public	AVHRR Pathfinder Version 5.3 L3-Collated (L3C) SST, Global, 0.0417°, 1981-present,
							Nighttime (1 Day Composite)
data			graph			public	AVISO Model Output, obs4MIPs NASA-JPL, Global, 1 Degree, 1992-2010, Monthly
data			graph	M		public	AVISO Model Output, obs4MIPs NASA-JPL, Global, 1 Degree, 1992-2010, Monthly, L



ABOUT US RESEARCH SOFTWARE PROFESSIONAL



Quick start
User Guide
Demo
Changelog
References
Credits



Standardized, web-based upload and download of sensor data and sensor metadata

Making ocean data FAIR – Reusable data formats

NetCDF formats

- Attribute Convention for dataset discovery (ACDD)
- Climate-Forecast NetCDF
- NetCDF-LD (emerging)



Established Community formats

- SeaDataNet (NetCDF & ASCII)
- OceanSITES (NetCDF)
- Everyone's Gliding Observatories (NetCDF)
- Argo (NetCDF)
- NCEI NetCDF templates
<https://www.nodc.noaa.gov/data/formats/netcdf/v2.0/>
- WHP-Exchange (NetCDF, CF 1.7)
- ...

Making ocean data FAIR – Gaps in capability

The complexities and gaps in attaining FAIR are buried hidden the detail, e.g.

- The need to assign usage licences to vocabularies
- How to assign DOIs to data included in services or for dynamic data
- The use of ORCID identifiers in DOI metadata

There are now tools available to help with assessing FAIRness of data or services (GO-FAIR, Fairsharing tools on GitHiub, ...). These are an active area of development.

NSF (through EarthCube) is actively supporting researchers and data centres in the US to be more "FAIR", in every sense of the word.

Other potential criteria beyond FAIR e.g. discoverability, integrity, reproducibility,...

FAIR data and OceanObs'19

- Ocean FAIR services OceanObs'19 white paper
- Ocean FAIR services OceanObs'19 poster
- Best practices break out session (Thursday afternoon)
- Steve Diggs will be at the CCHDO/BCO-DMO booth (#109A)

