



LifeWatch ERIC e-Science for NIS research workshop | 20-21 May 2021

FAIRification process for RIs:

The case of the LifeWatchGreece Research Infrastructure

HCMR TEAM in the ENVRI-FAIR project:

Dimitra MAVRAKI - Data Manager (dmavraki@hcmr.gr)

Eva CHATZINIKOLAOS - Scientific Responsible (evachatz@hcmr.gr)

Emmanouela PANTERI – IT Manager (emmipan@hcmr.gr)



Open and FAIR data concepts

- ✓ *Researchers acknowledge the benefits of open data, but data sharing practices are still limited. **Attitudes towards data sharing are generally positive, but open data is not yet a reality for many researchers.***
- ✓ *In 2016, the **'FAIR Guiding Principles for scientific data management and stewardship'** were published in **Scientific Data**. The authors intended to provide guidelines to improve the **Findability, Accessibility, Interoperability, and Reuse** of digital assets as humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.*
- ✓ *FAIR Data and Open Data make clear **the importance of sharing data, information and access, allowing collaboration and saving costs and time between scientists or between other user communities and possible reusers.***

How do Researchers share data?

Of 13 methods stated, top 4 options for currently sharing data were:

1. Emailing data files (65%)
2. Cloud service e.g. Dropbox, Google Drive (59%)
3. Portable storage (35%)
4. Supplementary data (20%)

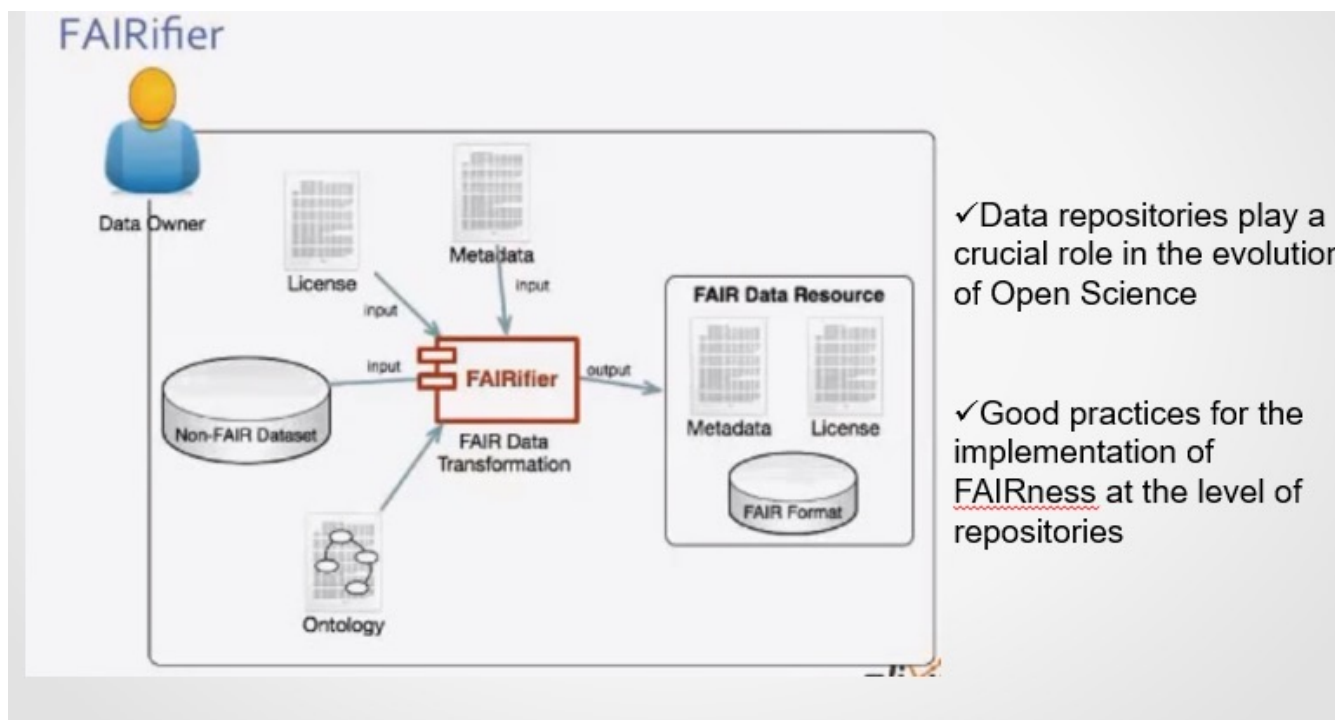
Formal repository (public / institutional) c.12%

[Jisc DAF studies](#)



FAIRness of Repositories and their (meta)data

- ❖ The [FAIR Data Principles](#) apply to metadata, data, and supporting infrastructure
- ❖ Trustworthy data repositories capable of curating FAIR data for researchers are a critical requirement for a European Open Science Cloud. ENVRI-FAIR connects the Environmental Research Infrastructure (ENVRI) community to the EOSC.



- repositories should be based on, or at least should support exporting metadata to, some **standardized metadata format** such as Dublin Core, DataCite or DDI.
- digital repositories should have a **search web page**, but could also support federated search via the SRU/W protocol.
- establishing qualified **links between data stored in the system** should be supported. “How to cite” option should be supported
- repositories should preserve the **usage license** associated with the data. Repositories should support Creative Commons licensees, although publishing data under some other well-known licenses should be supported by repositories as well. The information about license should be stored in metadata.
- repositories should support **publishing open, closed and embargoed data**.

**** Difficult to define FAIRness



- repositories **should preserve information about data provenance** which could be stored in metadata. Metadata could include: creator, institutions - publishers, source, mail address, publication year, production year, geo-location, data collector, data manager, distributor, editor, funder, producer, rights holder, sponsor, and supervisor. Also, repositories could support dataset versioning.
- humans and computers should be able to exchange and interpret each other's data. It means data should be **readable for machines** without the need for specialised or ad hoc algorithms, translators, or mappings. In order to ensure this, it is critical to use (1) commonly used controlled vocabularies, ontologies, thesauri and (2) a well-defined framework to describe and structure (meta)data.







GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data

email login ENGLISH

Home About

Hosted resources available through this IPT

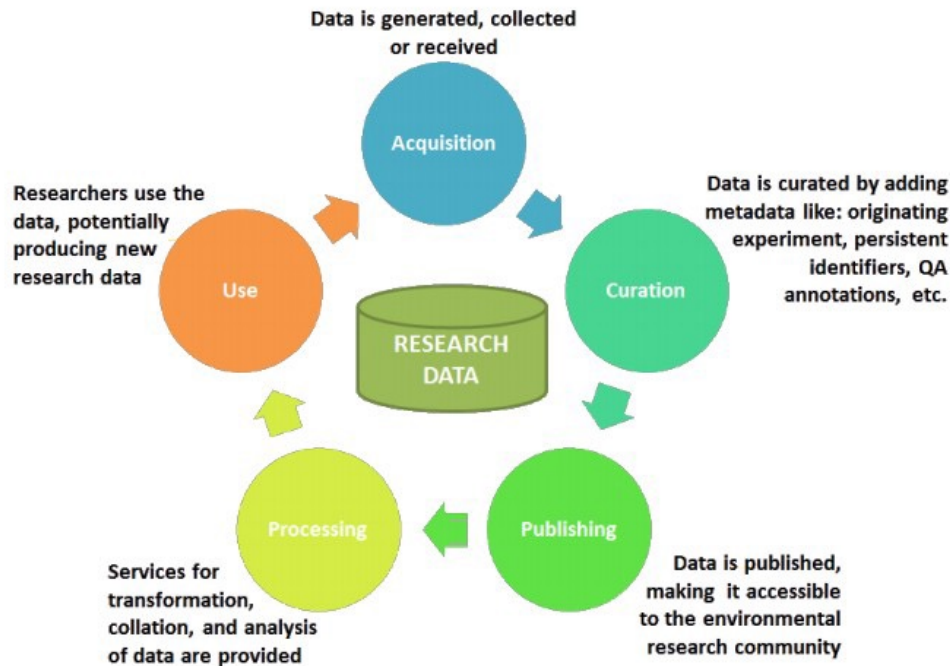
Filter:

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
	A. Vedel Taning 1918. A 7 Mediterranean Scopelidae (Saurus, Aulopus, Chlorophthalmus and Myctophum). Report on the Danish oceanographical expeditions 1908-10 to the Mediterranean and adjacent seas. Vol. II. Biology.	Hellenic Centre for Marine Research	Metadata-only	--	0	2017-12-18	2016-05-24	--
	Aegean Polychaetes	Hellenic Centre for Marine Research	Occurrence	--	2,215	2019-02-05	2019-02-05	--
	Benthic communities and environmental parameters in Amvrakikos Wetlands; Mazoma, Tsopeli Tsoukalio, Rodia and Logarou lagoons (September 2010 – July 2011)	Hellenic Centre for Marine Research	Sampling event	--	303	2019-02-05	2019-02-05	--
	Benthic communities and environmental parameters in three Mediterranean ports (Sardinia, Crete, Tunisia)	Hellenic Centre for Marine Research	Sampling event	--	540	2019-02-05	2019-02-05	--
--	Benthic Fauna of the Evvoia Coast and Evvoia Gulf.	Hellenic Centre for Marine Research	Sampling event	--	368	2019-01-25	2019-01-25	--
	Billards A. 1936. Les fonds de peches pres d Alexandrie. VI Hydroidea. Notes et memoires No 13	Hellenic Centre for Marine Research	Sampling event	--	24	2019-01-31	2019-01-31	--
--	BIOMAERL Maerl Biodiversity, Functional Structure And Anthropogenic Impacts (1996-1998).	Hellenic Centre for Marine Research	Sampling event	--	12	2019-02-05	2019-02-05	--
	Broch H. 1935. The fishery grounds near Alexandria. III Cirripeds. Notes et Memoires No 10.	Hellenic Centre for Marine Research	Sampling event	--	14	2019-01-31	2019-01-31	--

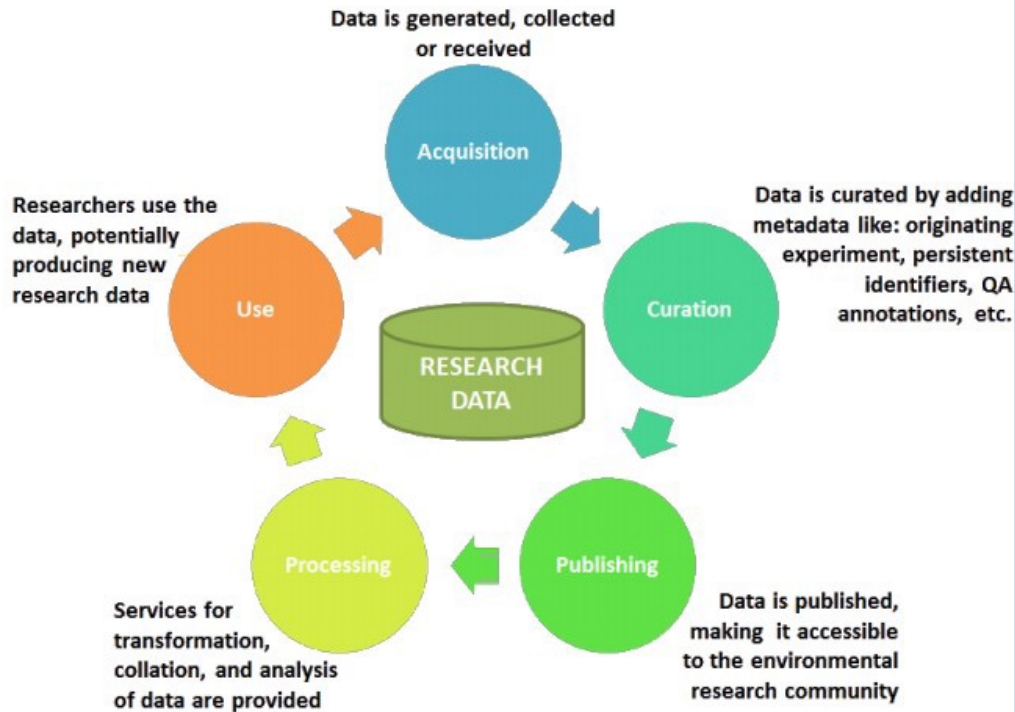
❑ Hosts a plethora of marine datasets (data and metadata)

❑ Already most of the FAIR principles are applied (such as rich metadata, standards, citation)

<http://ipt.medobis.eu/>



- ❑ LWG RI gives emphasis to metadata availability, even if the data are not available
- ❑ a metadata document with mandatory and suggested fields is created and distributed to all marine data providers prior to the submission of their data
- ❑ mandatory fields are: dataset title, citation, persons involved, abstract, key words, publications, external links
- ❑ metadata are always open to public (CC-Zero)



- ❑ LWG RI publishes its data through IPT (integrated publishing toolkit), a free open source software tool used to publish and share biodiversity datasets through the GBIF network
- ❑ in order to publish datasets through the IPT, data records are standardized in Darwin Core
- ❑ WoRMS taxonomic mapping is used, concerning locations, the Marine Regions application is used and for environmental information, the ENVO ontology
- ❑ detailed provenance, citation and usage license, always available

FAIRness of LifeWatchGreece RI



❑ **Findable** as it is a digital repository, and **Accessible** through its web page: <https://www.lifewatchgreece.eu/> and portal <https://portal.lifewatchgreece.eu/> after registration.

❑ **Interoperability** is guaranteed through commonly used controlled vocabularies and ontologies (e.g. Dublin Core Schema)

❑ **Reusable** through its Data Policy and Sharing Agreement which declares that LWG RI publishes open, closed and embargoed data, and in addition presents the rights and duties of data providers and data users.



*future challenges in respect to FAIR principles, **define as part of the LifeWatch community common :***

- **strategy and procedures** especially on ontologies used, on creation of services and on the data products re-use of. The lack of uniformity in data services and models from one repository to another, and in the richness and availability of metadata descriptions, makes integration and analysis of these data a manual, inaccurate and time-consuming task.
- **certification process,** to enter to the European network of trustworthy repositories enabling FAIR data

- ✓ It is very optimistic that LifeWatchGreece Research Infrastructure has already invested in FAIRness of data. Undoubtedly there are more steps to be done and we are engaged to fulfill them through our participation to the ENVRI-FAIR project.

- ✓ In addition, more effort shall be given to **define a good semantic model that could represent a consensus view in a particular domain, in our case for the marine domain.** Semantic models often contain multiple terms from existing ontologies and vocabularies. These conceptual models allow us to classify our data models and data items using the provided terms, concepts, and conceptual structures.

- ✓ It is proposed to **define a common set of indicators for all LifeWatch** Common Facilities (CFs) and Distributed Centers (DCs), in order to assess their FAIRness status. We could use the set of the proposed indicators by the RDA Working Group “FAIR data maturity model”.

- ✓ Repositories and services it is proposed to follow **a specific and common certification framework.**



Thank you