

FAIR Practices in Artificial Intelligence

- Skills for the European
- Open Science
- Commons

www.skills4eosC.eu

**Building a Sustainable Future for Open Science:
Skills4EOSC Competence Centres Workshop**

Recommendations from the AI
Research Community

Curtis Sharma, Elda Osmenaj, Ugo
Moschini, Valentina Pasquale, Lisana
Berberi

10 June 2025

Work Package 6

Professional Networks for Lifelong Learning

Mapping of Professional Networks

Starter Kits for Professional Networks

Data Stewardship Networks

Open Science Communities

Harmonisation and Creation New Professional Networks

FAIR for Health & Technology

FAIR for AI

Museum Curator Network for FAIR Data

Fellowship Programme

Top 10 FAIR Recommendations for ML/AI Research

Delphi Preparation phase

Desk research & Literature review



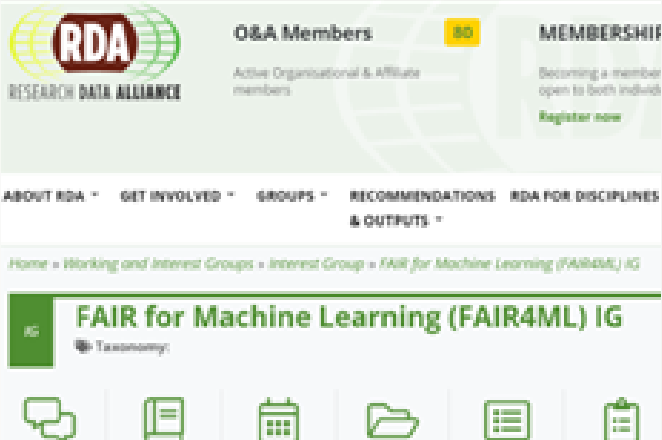
Drafting a preliminary list of practices

26

5	Metadata of the AI/ML model must remain accessible even in case the AI/ML model is no longer accessible.	This ensures the accessibility of metadata, even when the AI/ML model is no longer available.
6	The AI/ML model and its metadata must be retrievable by their identifier using a standardized communications protocol that is open, free, and universally implementable. Where necessary, the protocol should allow for an authentication and authorization procedure.	This allows the AI/ML model to be discovered by humans and machines, to be downloaded or directly invoked to conduct AI inference, enabling accessibility. The AI model interoperates with other

Expert's identification & engagement

(FAIR4ML) IG



Final list of recommended practices for Delphi






20

Recommended best practice to vote	More information/description on the best practice
1. ML/AI models should be trained on FAIR data.	To increase FAIRness of an ML/AI model, training should be thoroughly documented and made FAIR themselves.
2. Globally unique Persistent Identifiers should be assigned to ML/AI models and included in their metadata.	For instance, a DOI should be assigned to the IP and its metadata to make it findable for both humans and machines via indexing in searchable resources of model updates, new identifiers must be generated to distinguish the updated versions from the original.
3. ML/AI models should be described with rich metadata.	Rich metadata (descriptive keywords, contextual information, etc.) enhances findability and reuse.
4. Metadata for ML/AI models should remain accessible even when the ML/AI model is no longer accessible.	This ensures long-term preservation and access to metadata, even when the ML/AI model is no longer available.
5. ML/AI models and their metadata should be retrievable through persistent identifiers	This enables the ML/AI model to be discovered and machines, to be downloaded or directly invoked.

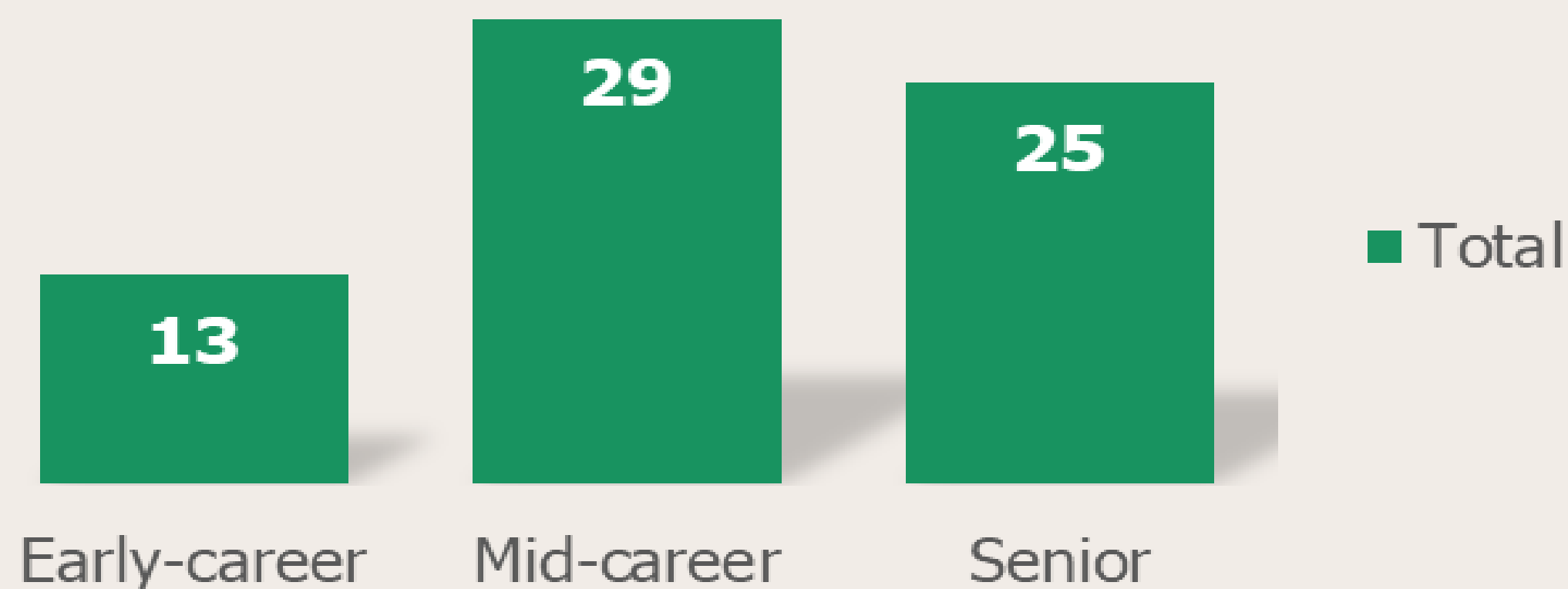
- Expert: I have extensive experien...
- Proficient: I am quite familiar w...
- Intermediate: I have a moderate u...
- Limited: I have a basic understand...
- None: I have never heard of the F...



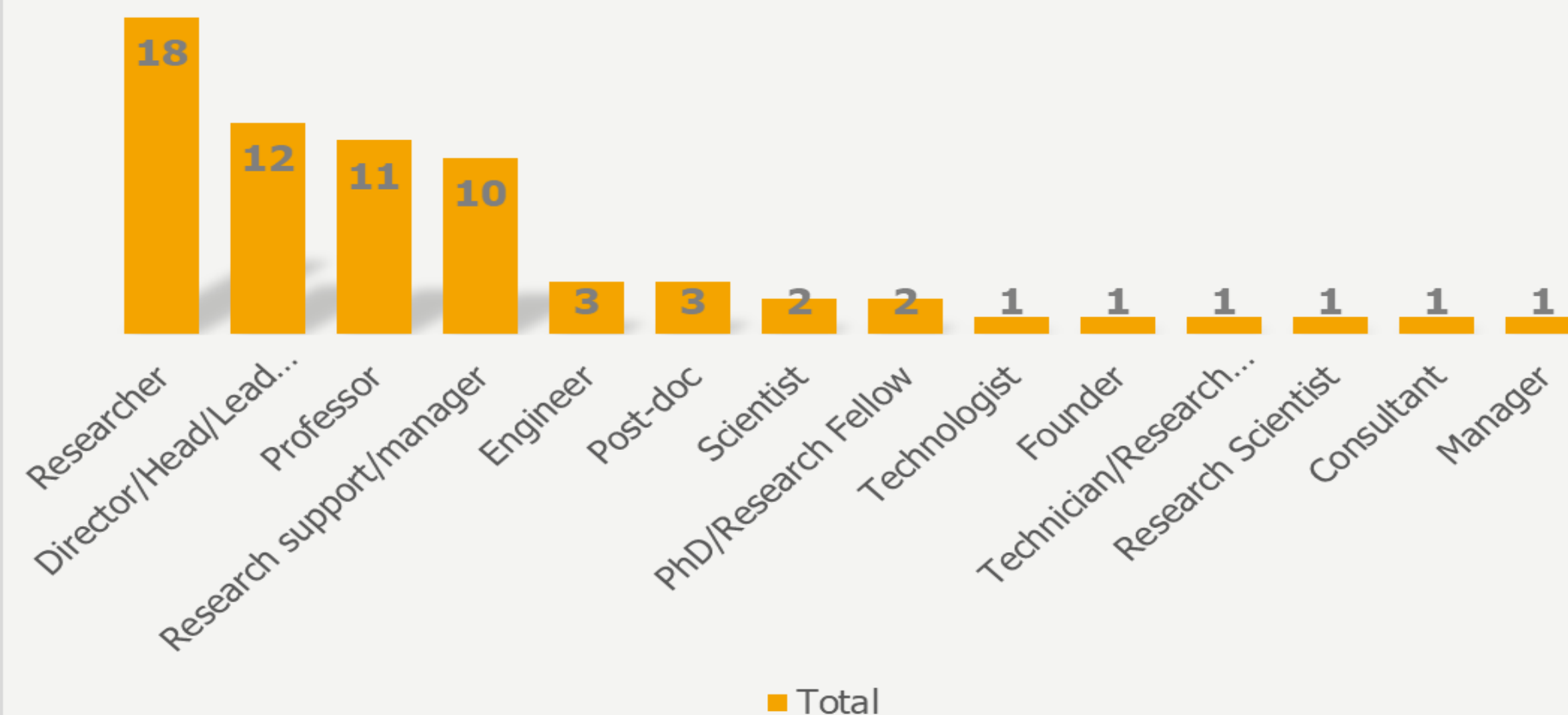
How familiar are you with FAIR principles?

		Answers	Ratio
Expert: I have extensive experience with the FAIR principles and have been involved in multiple projects where we have successfully applied these principles.		21	31.34 %
Proficient: I am quite familiar with the FAIR principles and have actively implemented them in my research/data management practices.		22	32.84 %
Intermediate: I have a moderate understanding of the FAIR principles.		22	32.84 %
Limited: I have a basic understanding of the FAIR principles.		1	1.49 %
None: I have never heard of the FAIR principles.		1	1.49 %
No Answer		0	0 %

Distribution across career stages



Analysis of roles in different categories



Top 10 FAIR Recommendations for ML/AI Research

FINDABLE

ACCESSIBLE

INTEROPERABLE

REUSABLE

2

a. **Globally Unique Persistent Identifiers** should be assigned to ML/AI models and included in their metadata.

F A

1

b. ML/AI models should be described with **rich metadata**.

F R

2

c. ML/AI models and their metadata should be retrievable through persistent identifiers using a **standardized communications protocol** that is open, free, and universally implementable.

F A

1

d. The ML/AI model's metadata should report **evaluation results**, e.g., accuracy, precision, recall, train vs. test error, etc.

R

1

e. ML/AI model metadata should **document any prerequisite on the input data format and structure**, including any required pre-processing, and output data format.

I R

1

f. ML/AI model **metadata** should use a documented, accessible, and ideally broadly used **format**.

I R

2

g. ML/AI model metadata should include **qualified references** to any related/required research objects, including code libraries, environments, code dependencies etc.

I R

1

h. ML/AI models should have **clear and accessible licenses** attached to them.

A R

2

i. Any **ethical considerations/analyses during model development** should be reported.

R

1

j. Clear **instructions on how to deploy the ML/AI model** should be provided.

A

6 practices reached consensus in round 1

4 practices reached consensus in round 2

Lkert Scale 1-5
1 = not recommended
5 = recommended

Consensus
At least 60% of responses concentrated on a given point

Top 10 FAIR Recommendations for ML/AI Research

Practice	Examples of Participants' comments
Globally unique Persistent Identifiers should be assigned to ML/AI models and included in their metadata.	This would facilitate citations and unambiguous references, especially in situations where foundation models are adapted/extended
ML/AI models should be described with rich metadata	Yes absolutely - ML is cursed right now in terms of providing details on what data they are trained on and what is the parameterisation of the model. This is essential.
ML/AI models and their metadata should be retrievable through persistent identifiers using a standardised communications protocol that is open, free, and universally implementable.	This would be ideal, but I am not sure if it will be technically possible given the diversity of model implementations and the bandwidth required for retrieval
The ML/AI model's metadata should report evaluation results, e.g. accuracy, precision, recall, train vs. test error, etc	This would be good because it allows for results reproducibility ... which is useful to advance research.
ML/AI model metadata should document any prerequisite on the input data format and structure, including any required pre-processing, and output data format.	This is important - perhaps more important than explaining the model itself as that makes it clear what are the precise inputs and outputs.
ML/AI model metadata should use a documented, accessible, and ideally broadly used format.	The format used should be easily obtainable from other formats, via software libraries, tools, web-app etc.
ML/AI model metadata should include qualified references to any related/required research objects, including code libraries, environments, code dependencies etc.	Yes, this is critical in order to ensure reusability of the model.
ML/AI models should have clear and accessible licences attached to them.	Absolutely ... Without licence information these materials are easily useless.
Any ethical considerations/analyses during model development should be reported.	A huge gap in today's practices
Clear instructions on how to deploy the ML /AI model should be provided	This is the key for R in FAIR

Practices with borderline Consensus

Practice	Participants' comments
Models should be trained on FAIR data	Where the data is outside the control of the researcher using it, it seems crazy to mandate that it should be FAIR for the user of the data. That is just a constraint on potential research.
References to all datasets (including versioning) used for training, validation, and testing of the ML/AI model must be included in the ML/AI model's metadata, via their Global Persistent and Unique Identifiers.*	Ideally yes, but may not always be possible? So, a general AI metadata can have "must have" fields and "optional" ones. This may become an "optional": good to have when possible
Model's metadata should include all information needed to fully reproduce the results, including (but not limited to) parameters like random number generator seed, hardware specifications, etc.*	Not known or not given
A sample FAIR dataset to be used with/test the model should be provided.*	Helpful, but not sure if essential
ML/AI model metadata should include the scope and known limitations to support responsible reuse in other contexts	Although this is relevant to expand the usability of the models, other aspects are of higher priority, in my opinion.

Eliminated Practices Round 1

Practice

Metadata for ML/AI models should remain accessible even when the ML/AI model is no longer accessible.

Standard metadata schemas and standard-compliant APIs should be used to make the ML/AI model interoperable with other models.

ML/AI models should be hosted on a dedicated online repository ensuring findability, accessibility, and interoperability.

ML/AI models should be made available also as platform-independent and general representation to enable interoperability and reuse.

ML/AI Models should be containerized to enhance interoperability and reuse.

ML/AI models should read, write and exchange data in a way that meets domain-relevant community standards.

AI/ML models should be associated with detailed provenance information, ideally by adding metadata referring to a specific ML-provenance language and including information about decision-making process for method selection.

Model lineage should be ensured by documenting that this model (version) in production is the output of a specific execution of ML/AI source code with specified hyperparameters and metrics.

Eliminated Practices Round 2

Practice

ML/AI models should be trained on FAIR data.

ML/AI model metadata should include the scope and known limitations to support responsible reuse in other contexts

References to all datasets (including versioning) used for training, validation, and testing of the ML/AI model must be included in the ML/AI model's metadata, via their Global Persistent and Unique Identifiers.*

Model's metadata should include all information needed to fully reproduce the results, including (but not limited to) parameters like random number generator seed, hardware specifications, etc.*

The ML/AI model and related datasets should comply with privacy protection laws and regulation: metadata should provide information on caveats for use and any warnings about unintended violations of privacy laws which may result from the use of the model.*

Metadata schemas for ML/AI models should make use of standard or ad hoc ontologies.*

Reporting of the ML process should follow community-backed guidelines and recommendations and provide the utilities/tools to convert data between the platform-specific format if applicable.*

A sample FAIR dataset to be used with/test the model should be provided.*

Next steps



Internal review and
submission to
Skills4EOSC
coordinators



Publish to
Skills4EOSC Zenodo
Community



Public co-creation



Stimulation and
uptake

Thank you - Questions



Curtis Sharma
4TU.ResearchData/TU
Delft



Elda Osmenaj
Istituto Italiano di Tecnologia
(IIT)



Valentina Pasquale
Istituto Italiano di Tecnologia
(IIT)



Ugo Moschini
Istituto Italiano di Tecnologia
(IIT)



Lisana Berberi
Karlsruhe Institute of Technology
(KIT)

- [“Top 10 FAIR Data Things” in Artificial Intelligence](#)
- [Lifecycle for Machine Learning](#)

Special thanks to the [FAIR4ML IG](#)
And all researchers for their time and contribution





This presentation is released under a CC-BY4.0 license

www.skills4eoscs.eu

in X zenodo



Funded by the European Union



UK Research and Innovation

co-funded by

Supporting



Skills4EOSC has received funding from the European Union's Horizon Europe research and innovation Programme under Grant Agreement No. 101058527 and from UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10040140].