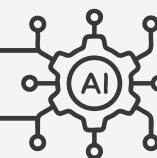


- Skills for the European
- Open Science
- Commons



QUICK GUIDE



## Top 10 FAIR Data Things for AI

## Authors

Curtis Sharma, TU Delft  
Elda Osmenaj, IIT

Ugo Moschini, IIT  
Valentina Pasquale, IIT  
Lisana Berberi, KIT

## Design

Aedeka Srl

Images & icons  
iStock, Flaticon

## Disclaimer

This is a digest of Deliverable “**D6.3 Top 10 FAIR Data Things for AI and Health Technology**” (<https://zenodo.org/records/15497622>) and was produced by the Skills4EOSC project. The project 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].

The information and views set out in this document are those of the author(s) and do not necessarily reflect the official opinion of the European Commission. Neither the European Commission guarantees the accuracy of the information included in this document. Neither the European Commission nor any person acting on the European Commission’s behalf may be held responsible for the use which may be made of the information contained therein.



## Copyright Notice

This work by Parties of the Skills4EOSC Project is licensed under a Creative Commons Attribution 4.0 International License: [creativecommons.org/licenses/by/4.0](https://creativecommons.org/licenses/by/4.0)



## DOI link

<https://doi.org/10.5281/zenodo.16738062>



## Table of Contents

---

<b>1. Introduction</b>	<b>&gt; 4</b>
------------------------	---------------

---

<b>2. How the 10 Practices were identified</b>	<b>&gt; 5</b>
Delphi Process Timeline	> 6

---

<b>3. The 10 FAIR Practices for ML/AI</b>	<b>&gt; 7</b>
Practice 1	> 7
Practice 2	> 7
Practice 3	> 7
Practice 4	> 8
Practice 5	> 8
Practice 6	> 8
Practice 7	> 8
Practice 8	> 9
Practice 9	> 9
Practice 10	> 9
The 10 Practices for FAIR AI	> 10

---

<b>4. Reflections from the community</b>	<b>&gt; 11</b>
--	----------------

---

<b>5. Conclusions and outlook</b>	<b>&gt; 12</b>
-----------------------------------	----------------

---

<b>6. References</b>	<b>&gt; 13</b>
----------------------	----------------

---





## 1. Introduction

This booklet presents a concise, practice-oriented version of the content developed in the **Skills4EOSC D6.3 deliverable titled “Top 10 FAIR Data Things for AI and Health Technology”**.

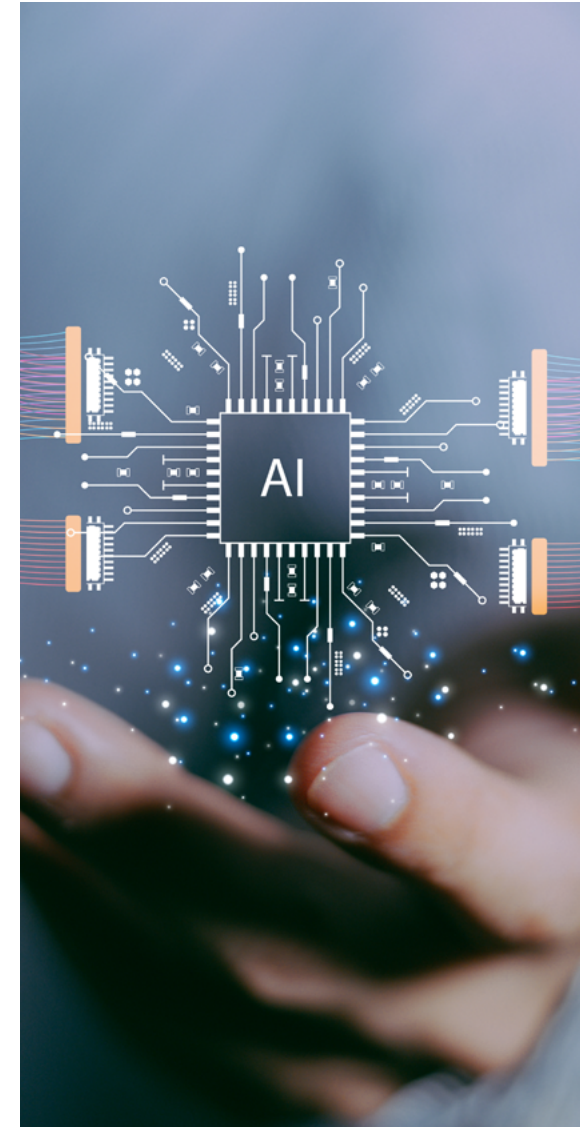
It focuses on the **Artificial Intelligence** section of that document.

It is intended as a quick-access resource for researchers, data professionals, and developers working to apply **FAIR principles (Findable, Accessible, Interoperable, Reusable)** in Artificial Intelligence workflows, especially those involving machine learning models and datasets.

The **10 practices** highlighted here were identified and validated through a collaborative, two-round Delphi study involving domain experts in ML/AI and FAIR. The aim is to share easy, practical steps to help make machine learning and AI models more FAIR. The materials of the Delphi study are available on Zenodo (link: <https://zenodo.org/records/16536643>).

In addition to the practices themselves, this booklet briefly summarises key reflections that emerged during a final community discussion, to provide context and indicate future directions.

**The work focuses on machine learning (ML) models** because they are widely used across disciplines and offer a practical entry point for developing FAIR implementation guidelines. This focus also enabled collaboration with ongoing, well-aligned initiatives. While the study did not distinguish between types of ML (e.g., supervised or reinforcement learning) or delve into more complex models like deep learning, the approach and outcomes are intended to be broadly applicable across different AI model types.





## 2. How the 10 Practices were identified

The **10 FAIR practices** presented in this booklet were identified through a structured two-stage Delphi study involving domain experts in **Machine Learning and Artificial Intelligence**.

In the first phase, a list of 20 recommended practices was compiled based on literature review and expert input. These practices were then evaluated in a first survey round, where participants rated their importance and provided suggestions.

**Six practices were approved immediately in Round 1.** Six were retained for further evaluation (having not quite reached consensus), and eight were discarded. In addition, six new suggestions emerged from participant comments and were added to the second round of voting.

**The second round included the six retained practices and the six new ones.** After this round, a total of 10 practices reached consensus and were selected as the final set of FAIR recommendations for ML/AI models. These practices are presented in the next section.

The entire process was carried out in close collaboration with experts from the FAIR and ML/AI communities, in particular the RDA **FAIR4ML IG** (FAIR for Machine Learning Interest Group), to ensure that the selected practices are both technically sound and practically relevant.







### 3. The 10 FAIR Practices for ML/AI

#### **PRACTICE 1**

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

Persistent identifiers (PIDs), such as DOIs, should be assigned to ML models and included in their metadata. This ensures that models can be reliably cited, found, and referenced by both humans and machines. PIDs are essential to support proper indexing, discovery, and traceability of models and their related components — including datasets, code, and configurations.

When models are updated or revised, new PIDs should be generated to reflect these changes and ensure clarity across versions.

#### **PRACTICE 2**

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

Metadata should include descriptive keywords, contextual information, and other relevant attributes to support findability and reusability.

Clear and complete metadata helps others understand what the model is, how it works, and under what conditions it can be reused — promoting transparency and trust across the AI workflow.

#### **PRACTICE 3**

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

The use of standardised, open, and free communications protocols ensures technical accessibility through widely adopted tools and services. Where necessary, the protocol should support authentication and authorisation procedures, enabling access control without compromising interoperability.



#### PRACTICE 4

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

Evaluation results such as accuracy, precision, recall, and train vs. test error should be included in the model's metadata.

This documentation supports transparency, allowing others to understand how the model was assessed, its limitations, and under what conditions it can be meaningfully reused or compared to other models.

#### PRACTICE 5

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

Providing this information helps ensure that others can correctly prepare data and interpret results when attempting to reuse or test the model, promoting transparency, reproducibility, and reusability.

#### PRACTICE 6

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

Using documented and accessible formats — ideally based on widely adopted standards — allows metadata to be interpreted by both humans and machines.

As with metadata for other digital outputs, this enhances discoverability, interoperability, and long-term usability.

#### PRACTICE 7

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

Explicit references to relevant research objects — such as code libraries, execution environments, or dependencies — should be included in the model metadata.

These references should use persistent identifiers and follow best practices for linking resources. Where possible, semantic





descriptions of the relationships among different workflow components should be included to support reusability, reproducibility, and collaborative development.

## **PRACTICE 8**

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

Licensing terms should be explicitly stated and easily accessible.

Clear licensing is essential to ensure legal clarity and encourage responsible reuse, redistribution, and adaptation.

Examples include open licenses such as Open Responsible AI Licenses (OpenRAIL), which are designed to support transparent and ethical sharing of AI artefacts.

## **PRACTICE 9**

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

Ethical risks — such as bias, discrimination, limitations, or unintended impacts — should be documented to promote responsible reuse.

Attention should also be given to privacy, surveillance, and security concerns, helping ensure responsible and inclusive reuse of the model.

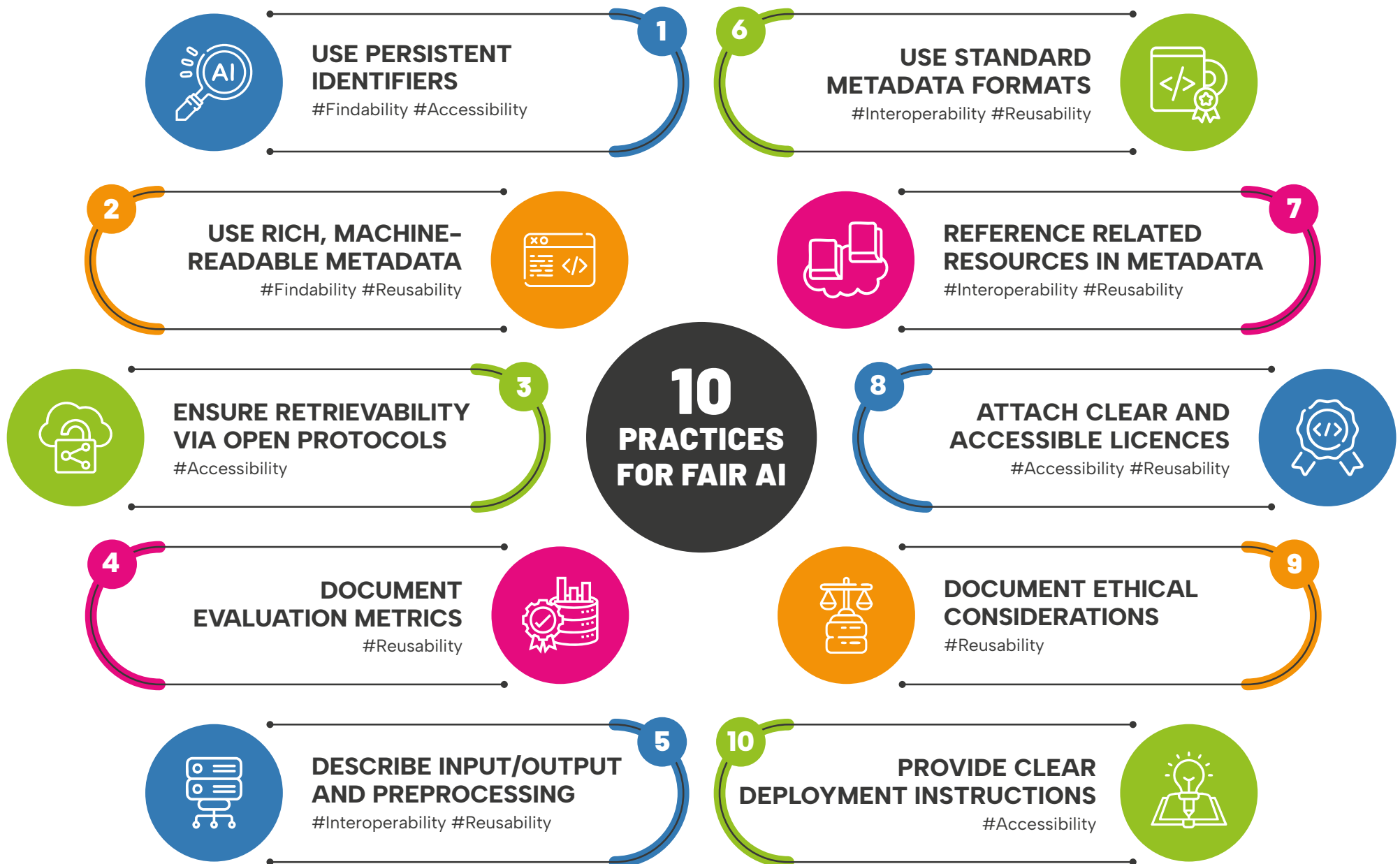
## **PRACTICE 10**

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

Deployment instructions ensure that users can effectively utilise the model in various environments, thus enhancing reproducibility.



## THE 10 PRACTICES FOR FAIR AI





## 4. Reflections from the community

To complement the Delphi study, a short online discussion was held with a small group of ML/AI and FAIR experts (survey participants). Participants were asked to reflect on the final list of 10 selected practices, sharing their impressions through structured input and open conversation.

### **Feedback was gathered around four key questions:**

- Which practices are already partially adopted in current workflows
- Which are easier to implement
- Which are more difficult to implement
- Which will require additional support for future uptake

Responses showed a clear distinction between well-established practices — such as using persistent identifiers and including licensing information — and more challenging ones, including documenting ethical considerations, model provenance, and model deployment. These latter practices were seen as important but harder to define, standardise, or apply without additional tools or shared frameworks.

During the open discussion, participants noted that many of the FAIR practices require not only awareness but also concrete examples, technical guidance, and supportive infrastructures. The need for common standards, shared evaluation frameworks, and lightweight tooling was emphasised as critical to bridge the gap between intention and implementation. In particular, issues related to ethics and transparency were seen as important, but often lacking clear operational pathways.

### COMMUNITY DISCUSSION

#### ML/AI Expert

While structured data simplifies FAIR compliance, broader metadata implementation and agreement on the most important metadata to report remains challenging, often due to social and organizational barriers (understanding the benefit and finding time for proper implementation) rather than technical ones.



#### ML/AI Expert

Detailed reporting of model performance, along with established guidelines, significantly enhances model robustness and comparison with other models.



#### ML/AI Expert

Including performance is very useful, but many details should be considered as models are affected by small variations for example seed (values used to ensure consistency and variability in processes and models) used during training and applying identical sets of metrics may not be the most appropriate.





## **5. Conclusions and outlook**

The 10 practices presented in this booklet are not intended as a definitive checklist, but as a practical starting point to support the adoption of FAIR principles in Artificial Intelligence.

They reflect what is achievable today with limited resources, and what is needed to build more transparent, shareable, and reusable ML/AI models.

Their effective implementation will require not only awareness but also supportive tools, practical examples, and lightweight guidance tailored to real-world use cases. Community-driven initiatives and experimentation will play a key role in refining and disseminating these practices across domains and organisations.

This work invites further collaboration, feedback, and adaptation. It is an open call to move from principle to practice — together.





## 6. References

Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). *Introducing the FAIR Principles for research software*. *Scientific Data*. <https://doi.org/10.1038/s41597-022-01710>

Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). *Preparing, conducting, and analyzing Delphi surveys: Crossdisciplinary practices, new directions, and advancements*. *MethodsX*, 8, 101401. <https://doi.org/10.1016/j.mex.2021.101401>

Castro, L. J., Psomopoulos, F., Serrano-Solano, B., Sharma, C., Shiferaw, K. B., Solanki, D., & Zhang, Y. (2023). *Lifecycle for FAIR Machine Learning*. Zenodo. <https://doi.org/10.5281/zenodo.10407265>

Chen, Y., Huerta, E. A., Duarte, J., Harris, P., Katz, D. S., Neubauer, M. S., Diaz, D., Mokhtar, F., Kansal, R., Park, S. E., Kindratenko, V. V., Zhao, Z., & Rusack, R. (2022). *A FAIR and AI-ready Higgs boson decay dataset*. *Scientific Data*. <https://www.nature.com/articles/s41597-021-01109-0>. <https://www.nature.com/articles/s41597-021-01109-0>

Cobey, K. D., Haustein, S., Brehaut, J., et al. (2023). *Community consensus on core open science practices to monitor in biomedicine*. *PLoS Biol.*, 21(1), e3001949. <https://doi.org/10.1371/journal.pbio.3001949>

Duarte, J., Li, H., Roy, A., Zhu, R., Huerta, E. A., Diaz, D., Harris, P., Kansal, R., Katz, D. S., Kavoori, I. H., Kindratenko, V. V., Mokhtar, F., Neubauer, M. S., Park, S. E., Quinnan, M., Rusack, R., & Zhao, Z. (2022). *FAIR AI Models in High Energy Physics*. arXiv. <https://doi.org/10.48550/ARXIV.2212.05081>

Huerta, E. A., Blaiszik, B., Brinson, L. C., Bouchard, K. E., Diaz, D., Doglioni, C., Duarte, J. M., Emani, M., Foster, I., Fox, G., Harris, P., Heinrich, L., Jha, S., Katz, D. S., Kindratenko, V., Kirkpatrick, C. R., Lassila-Perini, K., Madduri, R. K., Neubauer, M. S., Psomopoulos, F. E., Roy, A., Rübel, O., Zhao, Z., & Zhu, R. (2023). *FAIR for AI: An interdisciplinary and international community building perspective*. *Scientific Data*, 10(1), 487. <https://doi.org/10.1038/s41597-023-02298-6>

Katz, D. S., Kavoori, I. H., Kindratenko, V. V., Mokhtar, F., Neubauer, M. S., Park, S. E., Quinnan, M., Rusack, R., & Zhao, Z. (2022). *FAIR AI Models in High Energy Physics*. arXiv. <https://doi.org/10.48550/ARXIV.2212.05081>

Osmenaj, E., et al. *FAIR Principles Implementation in ML/AI - Findings from Skills4eosc Delphi Study*. 19th International Digital Curation Conference (IDCC25) The Hague, the Netherlands. <https://zenodo.org/records/14932518>

Osmenaj, E., Moschini, U., Pasquale, V., Sharma, C. J. M., & Berberi, L. (2025). *Materials from the Delphi Study on Implementing FAIR Principles in ML/AI - Deliverable 6.3 (Skills4EOSC Project)*. Zenodo. <https://doi.org/10.5281/zenodo.16536643>

Ravi, N., Chaturvedi, P., Huerta, E. A., Liu, Z., Chard, R., Scourtas, A., Schmidt, K. J., Chard, K., Blaiszik, B., & Foster, I. (2022). *FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy*. *Scientific Data*. <https://doi.org/10.1038/s41597-022-01712-9>

Roy, A., & FAIR4HEP collaboration. (2022). *FAIR Principles for data and AI models in high energy physics research and education*. Proceedings of 41<sup>st</sup> International Conference on High Energy physics – PoS(ICHEP2022). <https://doi.org/10.22323/1.414.0240>

Sharma, C. J. M., Osmenaj, E., Moschini, U., Pasquale, V., & Berberi, L. (2025). *D6.3 Top 10 FAIR Data Things in Artificial Intelligence & Published recommendations on FAIR data in Health Technology*. Zenodo. <https://doi.org/10.5281/zenodo.15681158>



[www.skills4eosC.eu](http://www.skills4eosC.eu)

Supporting



Funded by  
the European Union



co-funded by

UK Research  
and Innovation

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].

cover image source: [freepik.com](https://www.freepik.com)