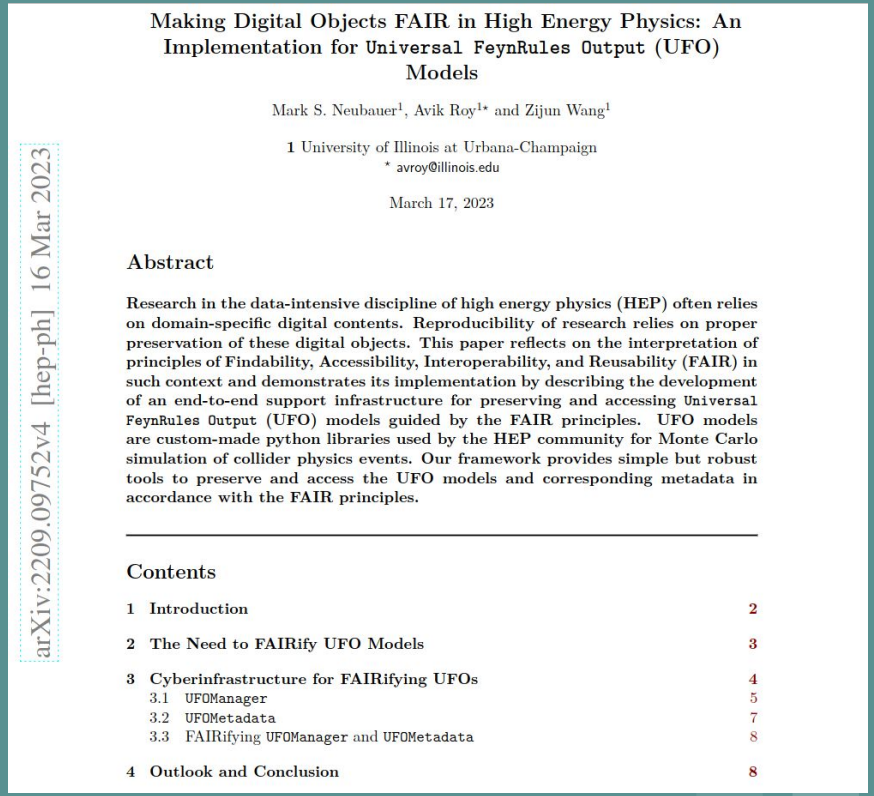
FAIR principles for Digital Objects in High Energy Physics

A Case Study with Universal FeynRules Output (UFO) Models

Avik Roy, Mark Neubauer, Zijun Wang
University of Illinois at Urbana-Champaign



Making Digital Objects FAIR in High Energy Physics: An Implementation for `Universal FeynRules Output` (UFO) Models

Mark S. Neubauer[1], Avik Roy[1*] and Zijun Wang[1]

1 University of Illinois at Urbana-Champaign
* avroy@illinois.edu

March 17, 2023

arXiv:2209.09752v4 [hep-ph] 16 Mar 2023

Abstract

Research in the data-intensive discipline of high energy physics (HEP) often relies on domain-specific digital contents. Reproducibility of research relies on proper preservation of these digital objects. This paper reflects on the interpretation of principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) in such context and demonstrates its implementation by describing the development of an end-to-end support infrastructure for preserving and accessing `Universal FeynRules Output` (UFO) models guided by the FAIR principles. UFO models are custom-made python libraries used by the HEP community for Monte Carlo simulation of collider physics events. Our framework provides simple but robust tools to preserve and access the UFO models and corresponding metadata in accordance with the FAIR principles.

Contents

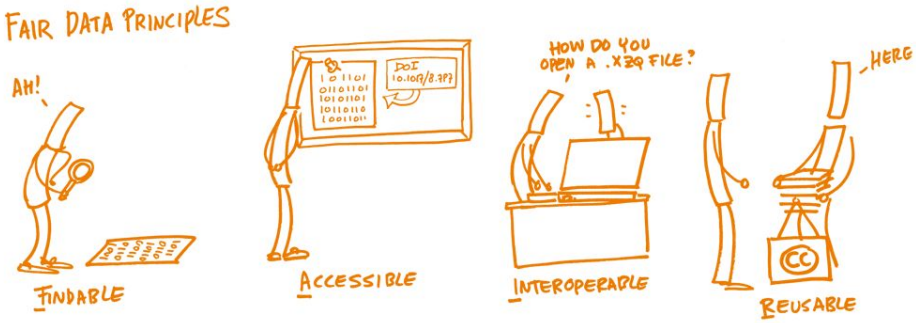
Details in our paper arxiv: 2209.09752 recently accepted at *SciPost Physics Codebases*
Git repositories: https://github.com/Neubauer-Group/UFOManager
https://github.com/Neubauer-Group/UFOMetadata/

CHEP, Norfolk, VA
May 09, 2023

The FAIR Principles

- To inspire scientific data management for reproducibility and maximal reusability[1]
- Originally proposed for scientific data
- Can be interpreted as guidelines to manage and preserve other Digital Objects (DOs) e.g. research software[2] , tutorials and notebooks[3] , AI and ML models[4]
- Different working groups working on FAIR guidelines for different DOs (e.g FAIR4RS, FAIR workflows, FAIR VREs)

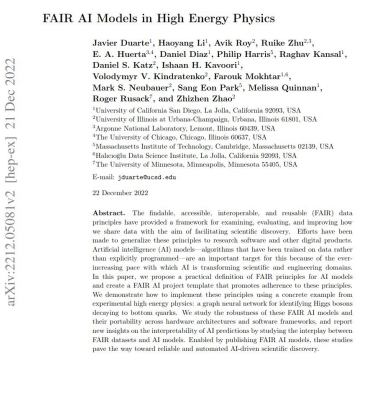
Findable:	locating DOs in a failsafe fashion
Accessible:	obtaining DOs along with their context, content, and format
Interoperable:	being usable across multiple computing platforms
Reusable:	specifying the context and extent of reusing DOs



Extending FAIR to Other Digital Objects

- Interpretation of FAIR depends on the nature of the digital content being FAIRified
- Dedicated work being done for interpreting and benchmarking the FAIR principles for a variety of digital objects
- Question: What kind of tools and processes are needed to make HEP DOs FAIR?
- Constraints:
  - FAIRification without interfering with the established practices of developing the DOs
  - Compatible with existing DOs as well as newer content
  - Preferably, an end-to-end software system with certain automations incorporated

arxiv: 2212.05081



FAIR AI Models in High Energy Physics

Javier Duarte[1], Haoyang Li[2], Avik Roy[2], Ruike Zhu[2,3], E. A. Huerta[3,4], Daniel Diaz[1], Philip Harris[5], Raghav Kansal[1], Daniel S. Katz[2], Ishaan H. Kavoori[1], Volodymyr V. Kindratenko[2], Farouk Mokhtar[3,6], Mark S. Neubauer[2], Sang Eon Park[2], Melissa Quinnan[1], Roger Rusack[7], and Zhizhen Zhao[2]

[1]University of California San Diego, La Jolla, California 92093, USA
[2]University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[3]Argonne National Laboratory, Lemont, Illinois 60439, USA
[4]The University of Chicago, Chicago, Illinois 60637, USA
[5]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[6]Haleakala Data Science Institute, La Jolla, California 92093, USA
[7]The University of Minnesota, Minneapolis, Minnesota 55405, USA

E-mail: jduarte@ucsd.edu

22 December 2022

Abstract. The findable, accessible, interoperable, and reusable (FAIR) data principles have provided a framework for examining, evaluating, and improving how we share data with the aim of facilitating scientific discovery. Efforts have been made to generalize these principles to research software and other digital products. Artificial intelligence (AI) models—algorithms that have been trained on data rather than explicitly programmed—are an important target for this because of the ever-increasing pace with which AI is transforming scientific and engineering domains. In this paper, we propose a practical definition of FAIR principles for AI models and create a FAIR AI project template that promotes adherence to these principles. We demonstrate how to implement these principles using a concrete example from experimental high energy physics: a graph neural network for identifying Higgs bosons decaying to bottom quarks. We study the robustness of these FAIR AI models and their portability across hardware architectures and software frameworks, and report new insights on the interpretability of AI predictions by studying the interplay between FAIR datasets and AI models. Enabled by publishing FAIR AI models, these studies pave the way toward reliable and automated AI-driven scientific discovery.

arXiv:2212.05081v2 [hep-ex] 21 Dec 2022

arxiv: 2210.08973

FAIR for AI: An interdisciplinary, international, inclusive, and diverse community building perspective

E. A. Huerta[1,2,*], Ben Blaiszik[1,3], L. Catherine Brinson[4], Kristofer E. Bouchard[5,6,7], Daniel Diaz[31], Caterina Doglioni[31], Javier M. Duarte[31], Murali Emani[11], Ian Foster[1,2], Geoffrey Fox[32], Philip Harris[33], Lukas Heinrich[33], Shantenu Jha[11,34], Daniel S. Katz[11,18,19,20], Volodymyr Kindratenko[17,18,19], Christine R. Kirkpatrick[21], Kati Lassila-Perini[22], Ravi K. Madduri[1], Mark S. Neubauer[17,23,27], Fotis E. Psomopoulos[35], Avik Roy[17,23], Oliver Rübel[5], Zhizhen Zhao[17,19], and Ruike Zhu[18]

[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, USA
[2]Department of Computer Science, University of Chicago, Chicago, Illinois 60637, USA
[3]Globus, University of Chicago, Chicago, Illinois 60637, USA
[4]Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, USA
[5]Scientific Data Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA
[6]Biological Systems & Engineering, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA
[7]Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California, 94720, USA
[8]Lund University, Department of Physics, Box 118, 221 00 Lund, Sweden
[9]School of Physics & Astronomy, The University of Manchester, Manchester M13 9PL, UK
[10]Department of Physics, University of California San Diego, La Jolla, California 92093, USA
[11]Leadership Computing Facility, Argonne National Laboratory, Lemont, Illinois 60439, USA
[12]Biocomplexity Institute and Department of Computer Science, University of Virginia, Charlottesville, Virginia 22904, USA
[13]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[14]Technical University Munich, Arcisstraße 21 80333 München, Germany
[15]Computational Science Initiative Brookhaven National Laboratory Upton, New York 11973, USA
[16]Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854
[17]National Center for Supercomputing Applications, University of Illinois, Urbana-Champaign, Urbana, Illinois 61801, USA
[18]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[19]Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[20]School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[21]San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, USA
[22]Helsinki Institute of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
[23]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[24]Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki 57001, Greece
*elihu@anl.gov, elihu@uchicago.edu

arXiv:2210.08973v1 [cs.CY] 30 Sep 2022

DOI: 10.1038/s41597-022-01710-x

scientific data

Explore content ~ About the journal ~ Publish with us ~

nature › scientific data › articles › article

Article | Open Access | Published: 14 October 2022

Introducing the FAIR Principles for research software

Michelle Barker, Neil P. Chue Hong, Daniel S. Katz, Anna-Lena Lamprecht, Carlos Martinez-Ortiz, Fotis Psomopoulos, Jennifer Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez & Tom Honeyman

Universal FeynRules Output (UFO) Models

- UFO models are used for simulating Beyond Standard Model Physics with Monte Carlo generators
- Custom Python libraries that pack necessary physics content as modules
- Designed to be **Interoperable** across multiple generator
- Heavily used in ATLAS and CMS analyses

Model-independent files	Model-dependent files

`__init__.py`
`object_library.py`
`function_library.py`
`write_param_card.py`
	`particles.py`
	`coupling_orders.py`
	`parameters.py`
	`vertices.py`
	`couplings.py`
	`lorentz.py`

- No uniform convention for management and citation practices for these models
- Most models are preserved at the FeynRules Model Database w/o dedicated metadata preservation or version controlling

Why FAIRify UFOs?
The Case for UFO Citations

- In many cases, only the theory paper describing the physics model is cited
- The actual digital object i.e. the implementation of the model itself is not always cited
- When cited, no uniform convention is followed
- Often digital repositories are cited, that may be lost/changed when the hosting service or the hosting account becomes inactive

[220] DM forum repository, *Higgs_scalar UFO model webpage*, https://svnweb.cern.ch/cern/wsvn/LHCDMF/trunk/models/Higgs_scalar_UFO/.

[221] DM forum repository, *Zp2HDM_UFO UFO model webpage*, https://svnweb.cern.ch/cern/wsvn/LHCDMF/trunk/models/EW_Higgs_2HDM/.

[222] DM forum repository, *DMS_tloop UFO model webpage*, https://svnweb.cern.ch/cern/wsvn/LHCDMF/trunk/models/Monojet_DMS_tloop/.

[223] DM forum repository, *DMScalarMed_loop UFO model webpage*, https://svnweb.cern.ch/cern/wsvn/LHCDMF/trunk/models/HF_S+PS/.

[224] DM forum repository, *dmS_T UFO model webpage*, https://svnweb.cern.ch/cern/wsvn/LHCDMF/trunk/models/Monojet_tChannel/contributed_by_Amelia_Brennan/.

From the references in JHEP 05 (2019) 142

Making UFOs FAIR

- Making UFOs FAIR requires consistent standards for
  - preserving the models with persistent identifiers like DOIs (F, A)
  - storing domain-specific enriched metadata (I, R)
  - allowing search and download models (F, R)
- FAIRification should be independent of developing the UFO, compatible with existing models (hence, compatible with both Python 2 and 3)

UFOManager

Dedicated software developed to

- Validate models
- Create enriched metadata
- Store said metadata in dedicated repo
- Publish model in Zenodo with DOI
- Facilitate search and download of models
- Allows version controlling

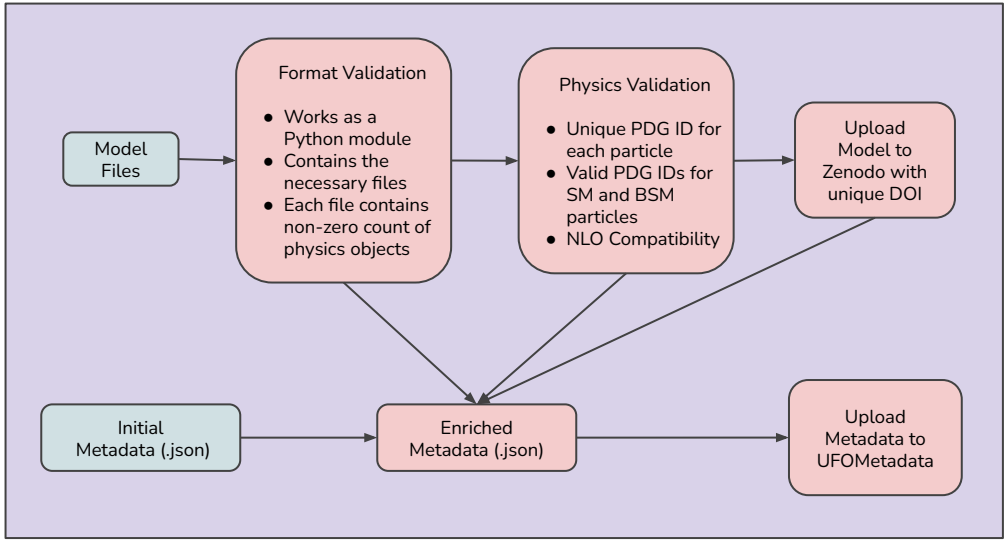
UFOMetadata

Dedicated repository developed to

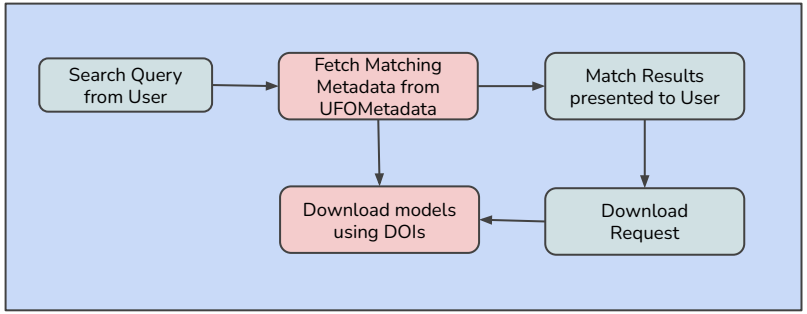
- Store metadata of published UFO models
- Validate metadata format with continuous integration

UFOManager

Upload Workflow



Search and Download Workflow



The Enriched Metadata


```
{
   "Author":  [
                     {"name" :  "Luca Panizzi",
                     "affiliation":  "Uppsala University",
                     "contact":  "luca.panizzi@physics.uu.se"},
                     {"name" :  "Benjamin Fuks",
                     "affiliation":  "Sorbonne University",
                     "contact":  "fuks@lpthe.jussieu.fr"}
   ],
   "Paper_id":          {"doi":  "10.1140/epjc/s10052-017-4686-z",
                        "arXiv" :  "1610.04622"},
   "Description":       "Vector-like Quark UFO Model at NLO QCD
                        with four flavour scheme",
   "Model Homepage" :  "https://feynrules.irmp.ucl.ac.be/wiki/NLOModels"
}
```

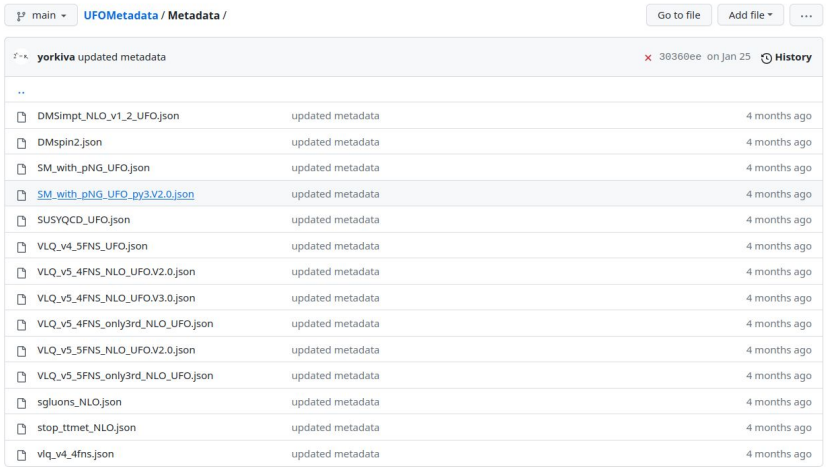

Initial Metadata

Enriched Metadata


```
{
   "Model name":                          "UFO model for Vector-like Quarks at NLO QCD with four flavor scheme",
   "Model Doi":                           ""10.5281/zenodo.6977663"",
   "Author":  [
                                          {"name" :  "Luca Panizzi",
                                          "affiliation":  "Uppsala University",
                                          "contact":  "luca.panizzi@physics.uu.se"},
                                          {"name" :  "Benjamin Fuks",
                                          "affiliation":  "Sorbonne University",
                                          "contact":  "fuks@lpthe.jussieu.fr"}
   ],
   "Paper_id":                            {"doi":  "10.1140/epjc/s10052-017-4686-z",
                                          "arXiv" :  "1610.04622"},
   "Description":                         "Vector-like Quark UFO Model at NLO QCD
                                          with four flavour scheme",
   "Model Homepage" :                     "https://feynrules.irmp.ucl.ac.be/wiki/NLOModels",
   "Number of decays":                    10,
   "Number of coupling orders":           3,
   "Number of coupling tensors":          97,
   "Number of lorentz tensors":           39,
   "Number of parameters":                98,
   "Number of vertices":                  125,
   "Number of propagators":               4,
   "Model Python Version":                2,
   "Model Version":                       1.0,
   "Allows NLO calculations":             true,
   "All Particles":                       {"ve":  12, "ghG~":  -82, "y~":  -6000008,
                                           "u~":  -2, "vm~":  -14, "vt":  16, "s":  3,
                                           "c~":  -4, "t~":  -6, "G0":  250, "u":  2,
                                           "ve~":  -12, "tp":  6000006,
                                           "x~":  -6000005, "G-":  -251,
                                           "ta+":  -15, "G+":  251, "vt~":  -16, "e-":  11,
                                           "y":  6000008, "e+":  -11, "H":  25, "t":  6,
                                           "vm":  14, "d~":  -1, "s~":  -3,
                                           "mu-":  13, "ghG":  82, "bp":  6000007, "a":  22,
                                           "x":  6000005, "ta-":  15,
                                           "b~":  -5, "Z":  23,
                                           "d":  1, "g":  21,
                                           "W-":  -24, "W+":  24,
                                           "tp~":  -6000006, "mu+":  -13,
                                           "bp~":  -6000007, "c":  4, "b":  5, },
   "SM Particles":                        {"ve":  12, "u~":  -2, "vm~":  -14, "vt":  16, "s":  3,
                                           "c~":  -4, "t~":  -6, "u":  2,"ve~":  -12,
                                           "ta+":  -15, "vt~":  -16, "e-":  11,
                                           "e+":  -11, "H":  25, "t":  6,
                                           "vm":  14, "d~":  -1, "s~":  -3,
                                           "mu-":  13, "a":  22, "ta-":  15,
                                           "b~":  -5, "Z":  23,
                                           "d":  1, "g":  21,
                                           "W-":  -24, "W+":  24,
                                           "tp~":  -6000006, "mu+":  -13,
                                           "bp~":  -6000007, "c":  4, "b":  5, },
   "New elementary particles":            {"G0":  250, "y~":  -6000008, "tp":  6000006,
                                           "bp":  6000007, "x~":  -6000005, "x":  6000005,
                                          "y":  6000008, "tp~":  -6000006, "G-":  -251,
                                           "G+":  251, "bp~":  -6000007 }
   "BSM particles with standard PDG codes":  {"G-":  -251, "G+":  251,
                                          "tp":  6000006, "tp~":  -6000006,
   "Particles with PDG-like IDs":         {"G0":  {"charge":  0.0, "spin":  1, "id":  250},
                                          "bp":  {"charge":  -1/3, "spin":  2, "id":  6000007},
                                          "x ":  {"charge":  -5/3,"spin":  2,"id":  -6000005 },
                                          "y ":  {"charge":  4/3, "spin":  2, "id":  -6000008 },
                                          "y":  {"charge":  -4/3, "spin":  2, "id":  6000008 },
                                          "x":  {"charge":  5/3, "spin":  2, "id":  6000005 },
                                          "bp ":  {"charge":  1/3, "spin":  2, "id":  -6000007}}
}
```

UFOMetadata

- Stores Metadata for FAIRified UFO models
- Interfaces with the Upload script of UFOManager to enable the storage of enriched metadata, and with the Download script to enable search for the right UFO models
- Has a built-in Continuous Integration Workflow to automate the integration of new UFO models



Summary and Outlook

- The work presented here is a demonstrative example of the kind of tools and approaches needed to FAIRify custom-HEP DOs
- Room for improvement:
  - Seamless integration with existing model repositories and databases
  - Work to be done to include compatibilities with UFO v2.0
  - Making it a part of a more sustainably preserved cyberinfrastructure for development, preservation, dispersion, and citation of UFO models
- Ultimately, we hope to initiate and engage in a larger community-wide discussion on adaptation of FAIR principles for DOs (UFOs and beyond) in HEP

Mark Neubauer

Zijun Wang

References

1. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
2. Chue Hong, Neil P., Katz, Daniel S., Barker, Michelle et al. RDA FAIR4RS WG. (2022). FAIR Principles for Research Software (FAIR4RS Principles) (1.0). https://doi.org/10.15497/RDA00068
3. Richardson, R. A., et al. "User-friendly Composition of FAIR Workflows in a Notebook Environment." Proceedings of the 11th on Knowledge Capture Conference. 2021. https://doi.org/10.1145/3460210.3493546
4. Katz, D. S., Psomopoulos, F. E., and Castro, L. J. "Working towards understanding the role of FAIR for machine learning." DaMaLOS@ ISWC (2021): 1-7. https://doi.org/10.4126/FRL01-006429415