IRIS-HEP Fellowship Proposal:
Exploring the FAIR principles for preservation of UFO models

Zijun Wang[1]

[1]University of Illinois at Urbana-Champaign

April 2022

1 Project Description

The Standard Model describes three fundamental interactions and encompasses elementary particles that have been discovered so far, which is of great significance to fundamental physics. However, there are still many unresolved mysteries, such as gravitational interactions or the nature of dark matter, that cannot be explained by the Standard Model. Therefore, physicists have also been working to explore physics beyond the Standard Model (BSM). The rich spectrum of BSM theory allows us to explore fundamental particle interactions at the Large Hadron Collider and look for statistical or kinematic footprints of physics from new particles.

To search for new physics beyond the Standard Model, mathematical description of the BSM theory, the Lagrangian, needs to be stored as a serialized object so that they can be used by Monte Carlo event generator softwares like MadGraph [1] to simulate collider physics in the context of the BSM theories. A number of customs have been historically followed by theorists and phenomenologists to store physics information of particle systems. Currently, the Universal `FEYNRULES` Output (UFO) [2] model is one of the most popular ways to store BSM physics details. Unlike other models which store BSM physics information in a collection of text files, UFO models store different information of the particle model, in a generator-independent way, into different `PYTHON` files. Such a storage method allows event generators working directly with UFO models without any modification or further interfacing. Thus, UFO models allow an easier way for accessions to physics model information on different platforms and attributions to existing objects.

However, like any other digital content, UFO models have software and platform dependencies, require version controlling, and can benefit from a unified way of preserving and distributing these resources. Many of these models are now stored as simple archived objects hosted by some webpage [3] with little or no information about physics validation or digital verification. In this project, we aim to create a set of tools, guided by the FAIR (Findable, Accessible, Interoperable, and Reusable) Principles [4], as a bridge among the developers and users of UFO models. Originally envisioned for scientific data management, the FAIR guideline is being explored in the context of other digital objects like scientific software and notebooks across a number of disciplines. Thus, in this project, we will explore the specific interpretation of FAIR for UFO models and implement an interactive platform to optimize the preservation and investigation of UFO models to help high-energy physicists in their research.

For developers, a central, public repository will be published to handle registration and verification of the UFO models. Equipped with CI/CD tools available via GitHub, the repository will allow UFO developers to submit and register their model via the git CLI. The CI/CD tool will perform automated checks on the files being committed and will provide necessary feedback to the developers when necessary. It will identify each model with a set of keywords, including the arXiv identifier and/or the DOI for the associated publication. Each registered model and its different versions will themselves be assigned DOIs which can be used to identify and cite the exact model a research project is going to use. As a result, the work done by the theorists and phenomenologists will receive better visibility within the community and they will be able to document and revise their models better.

Additionally, this project will aim towards developing an API built for users who can search for UFO models using keywords. The API will search the Git repository and offer the user with a choice of potential matches that the user can choose from and eventually use the API to download and use them. In this way, users can find existing models more conveniently and quickly to meet the needs of research.

This project will be carried out under the supervision and mentorship of Dr. Avik Roy (University of Illinois at Urbana-Champaign) and Dr. Matthew Feickert (University of Illinois at Urbana-Champaign) and Prof. Mark Neubauer (University of Illinois at Urbana-Champaign).

2 Proposed Timeline

- **Week 1 to 3:**
  Learn basic properties of the Standard Model. Familiarize with necessary software and tools, such as Git, GitHub, and MadGraph. Learn a few benchmarks of UFO models and determine how different information is stored as `PYTHON` scripts. Familiarize with the FAIR Principles. Create a public GitHub repository for the project.
- **Week 4 and 5:**
  Come up with data processing techniques, and FAIR criterion for UFO models. Design a first set of tests and use them to test the integrity and the format of such models.
- **Week 6 and 7:**
  Incorporate those mentioned tests into GitHub's CI/CD platform. Keep working on improving the tests.
- **Week 8 and 9:**
  Register models with DOI. Test the availability of registering new UFO models and modifying existing models. Start working on API building.
- **Week 10 and 11:**
  Finalize the initial version of the API, test the availability of accurate accession to models needed, and make it available via Python Package Interface (PyPI).
- **Week 12:**
  Depending on the completion of the main project, try to import more existing BSM models into UFO format to the software. Compile the final result and software in the form of a report and presentations.

References

[1] Johan Alwall et al. "MadGraph 5: going beyond". In: *J. High Energy Phys.* 2011.6 (2011), p. 128. DOI: `https://doi.org/10.1007/JHEP06(2011)128`.

[2] Céline Degrande et al. "UFO – The Universal FeynRules Output". In: *Computer Physics Communications* 183.6 (2012), pp. 1201–1214. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/j.cpc.2012.01.022`.

[3] *FeynRules models to be used for NLO calculations with aMC@NLO*. URL: `https://feynrules.irmp.ucl.ac.be/wiki/NLOModels`.

[4] Mark D Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1 (2016), pp. 1–9. DOI: `https://doi.org/10.1038/sdata.2016.18`.