
Plan Overview

A Data Management Plan created using DMPonline

Title: Higher-order electroweak corrections to the Standard Model effective field theory

Creator: Hesham El Faham

Principal Investigator: Eleni Vryonidou

Data Manager: HESHAM EL FAHAM

Affiliation: University of Manchester

Funder: European Commission

Template: Horizon Europe Template

ID: 132023

Start date: 01-10-2024

End date: 01-10-2026

Last modified: 16-08-2023

Grant number / URL: <https://ec.europa.eu/>

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Higher-order electroweak corrections to the Standard Model effective field theory

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- None of the above

2. Is The University of Manchester collaborating with other institutions on this project?

- No - only institution involved

3. What data will you use in this project (please select all that apply)?

- Re-use existing data (please list below)
- Acquire new data

Reuse of existing open-source public software tools for relevant computations. The generation of the so-called Monte Carlo (MC) data from such software tools is foreseen. MC data is used for physics analysis purposes. Experimental high-energy physics data from various experimental collaborations can be used. The usage of these data is restricted to public ones available on hepdata.net, the corresponding repository.

4. Where will the data be stored and backed-up during the project lifetime?

- Other storage system (please list below)

GitHub or Dropbox.

5. If you will be using Research Data Storage, how much storage will you require?

- Not applicable

6. Are you going to be receiving data from, or sharing data with an external third party?

- Yes

Data can be shared with my collaborators.

7. How long do you intend to keep your data for after the end of your project (in years)?

- 5 - 10 years

Guidance for questions 8 to 13

Highly restricted information defined in the [Information security classification, ownership and secure information](#)

[handling SOP](#) is information that requires enhanced security as unauthorised disclosure could cause significant harm to individuals or to the University and its ambitions in respect of its purpose, vision and values. This could be: information that is subject to export controls; valuable intellectual property; security sensitive material or research in key industrial fields at particular risk of being targeted by foreign states. See more [examples of highly restricted information](#).

If you are using 'Very Sensitive' information as defined by the [Information Security Classification, Ownerships and Secure Information Handling SOP](#), please consult the [Information Governance Office](#) for guidance.

Personal information, also known as personal data, relates to identifiable living individuals. Personal data is classed as special category personal data if it includes any of the following types of information about an identifiable living individual: racial or ethnic origin; political opinions; religious or similar philosophical beliefs; trade union membership; genetic data; biometric data; health data; sexual life; sexual orientation.

Please note that in line with [data protection law](#) (the UK General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de-identified) as soon as practically possible. You must obtain the appropriate [ethical approval](#) in order to use identifiable personal data.

8. What type of information will you be processing (please select all that apply)?

- No confidential or personal data

9. How do you plan to store, protect and ensure confidentiality of any highly restricted data or personal data (please select all that apply)?

- Not applicable

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- Not applicable

11. Will the participants' information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- Not applicable

12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?

- Not applicable

13. Are you planning to use the personal information for future purposes such as research?

- No

No personal information is involved.

14. Will this project use innovative technologies to collect or process data?

- Yes, and innovative technologies will not collect or process personal data (please list the innovative technologies below)

Innovative technology is in the form of public tools used for particle physics computations. These technologies will be used and will

not collect or process any personal data.

15. Who will act as the data custodian for this study, and so be responsible for the information involved?

Hesham El Faham

16. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2023-08-16

Data Summary

Will you re-use any existing data and what will you re-use it for?

Existing open-source public software tools will be used. Existing Monte Carlo (MC) data; data files containing numerical and textual information can also be used. MC data are the output of open-source public software tools used for particle physics computations. MC existing data can be used, or new ones can be generated throughout the project. Experimental high-energy physics public data from various experimental collaborations can be used. These data (existing software, MC or experimental data) will facilitate computations, cross-check new ones made throughout the project, and be used in physics analyses.

What types and formats of data will the project generate or re-use?

Software tools used are written in Python, C and Fortran. Any code will be stored on GitHub. Monte Carlo data are files with numerical and textual information and will be stored on Dropbox. Experimental data can be in a data file or JSON format.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Existing software tools, Monte Carlo, and experimental data will facilitate computations, cross-check new ones made throughout the project, and be used in physics analyses.

What is the expected size of the data that you intend to generate or re-use?

Software tools are on the orders of MBs. Monte Carlo and experimental data can be on the orders of GBs.

What is the origin/provenance of the data, either generated or re-used?

The software tools used are open-source and public. Monte Carlo data are generated using open-source public software tools. Experimental public data is available on a public repository.

To whom might your data be useful ('data utility'), outside your project?

User who want to perform similar computations.

FAIR data

2.1. Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?

Any software releases as an outcome of this project will be identified through their release on GitHub. Information on using existing public open-source software tools to regenerate Monte Carlo data used in the project will also be available on the repository.

2.1. Making data findable, including provisions for metadata: Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Any release of data will be accompanied by the appropriate metadata in the same repository.

2.1. Making data findable, including provisions for metadata: Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Search keywords will be provided in the metadata.

2.1. Making data findable, including provisions for metadata: Will metadata be offered in such a way that it can be harvested and indexed?

Yes.

2.2. Making data accessible - Repository: Will the data be deposited in a trusted repository?

Yes.

2.2. Making data accessible - Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?

No arrangements beyond having an account on GitHub and Dropbox with the nominal storage.

2.2. Making data accessible - Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes.

2.2. Making data accessible - Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

After the software outcome has been tested, it will be publicly released with no restricted access. Relevant Monte Carlo data or information on generating it will also be publicly released.

2.2. Making data accessible - Data:

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

N/A

2.2. Making data accessible - Data:

Will the data be accessible through a free and standardized access protocol?

Yes.

2.2. Making data accessible - Data:

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

N/A

2.2. Making data accessible - Data:

How will the identity of the person accessing the data be ascertained?

N/A

2.2. Making data accessible - Data:

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

No.

2.2. Making data accessible - Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Yes.

2.2. Making data accessible - Metadata:

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

As long as the repository remains available.

2.2. Making data accessible - Metadata:

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Software tools needed to access or read the data are all public tools the interested user can access. Information on where to find these tools will be provided.

2.3. Making data interoperable:

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Data are in the format of code written in a coding language, e.g. Python, C, Fortran, or text files. The metadata is in textual form. Users with knowledge of code and text will be able to reuse the data.

2.3. Making data interoperable:

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Nothing special will be needed.

2.3. Making data interoperable:

Will your data include qualified references^[1] to other data (e.g. other data from your project, or datasets from previous research)?

[1] A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

No.

2.4. Increase data re-use:

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

README files will be provided to explain the source of the data and how to reproduce it. Academic papers will also be written to show the importance of the outputs.

2.4. Increase data re-use:

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes.

2.4. Increase data re-use:

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes.

2.4. Increase data re-use:

Will the provenance of the data be thoroughly documented using the appropriate standards?

Yes.

2.4. Increase data re-use:

Describe all relevant data quality assurance processes.

Software will be tested and cross-checked.

2.4. Increase data re-use:

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

OK.

Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g.

software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Modification of existing software and Monte Carlo data will be generated, all mentioned above.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

All have been addressed in the previous questions.

Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

No costs.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

N/A

Who will be responsible for data management in your project?

Hesham El Faham

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

As long as the repository exists.

Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Data on GitHub repositories are backed up by default.

Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

No.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

No personal data is involved.

Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No.