TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

- **Instructions and footnotes in blue must not appear in the text.**
- **For options *[in square brackets]*: the option that applies must be chosen.**
- **For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): enter the appropriate data.**

Introduction

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. You should develop a single DMP for your project to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell this out.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

FAIR data management

In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept.

More information about FAIR:

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

Structure of the template

The template is a set of questions that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.

Project[1] Number: [101094300]

Project Acronym: [MuCol]

Project title: [A Design Study for a Muon Collider complex at 10 TeV centre of mass]

DATA MANAGEMENT PLAN

[1] The term 'project' used in this template equates to an 'action' in certain other Horizon 2020 documentation

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

The consortium will produce predominantly textual reports for internal use or for external dissemination, as well as presentations in standard formats, together with data sets from simulation results.

What types and formats of data will the project generate/collect?

The main data that the project will collect and produce is a table of parameters, describing from start to end the characteristics of an entire chain of accelerators that will produce intense beams of charged muons and collide them at different energies (the baseline is 3 TeV and 10Tev, an option at 14 TeV is being considered).

This table will be stored in the online tool Overleaf (www.overleaf.com) as a working document accessible to the members of the project for editing and proof-reading. At specified moments in time (M6, M18 and M30), the table will be frozen and registered in a pdf document, containing also narrative explanations of the parameters and the reasons for the choices, that will be registered in Zenodo as proof of achievement of the corresponding Milestone. These documents will also be used as basis for further publications, yet to be determined.

Each WorkPackage will produce simulation data with Softwares typical of the domain of the workpackage. Data will be produced according to the formats that are defined by each of these software. Their storage will depend on the format, on the Institute that will produce the data and from the use of the data, that will be defined by each workpackage. In general, rather than keeping the data, which are in general large size files which would be difficult to store, scientists will store input files (in the proprietary format of the software being used) and will record the version of the software used. Several simulation packages available as open source or anyway in free release, will make use of the CERN GitHub (https://github.com/CERN). Access rights will be modulated depending on the status of the data (work in progress or published). A publication policy is being agreed among all the institutes participating to MuCol and to the International Muon Collider Collaboration. As sson as agreed by all institutes and published it will be added as a reference to this document. The details of data to be produced/treated by each WorkPackage are the following:

Workpackage 1: Coordination and Integration

WP1 will coordinate the editing of reports to be delivered as proof of achievement for milestones and deliverables. Will therefore store all reports in Zenodo (https://zenodo.org/communities/mucol/). Several reports will also be produced even if not linked directly to Milestones or Deliverables, in this case they will be stored either under Zenodo, or in the CERN CDS (http://cdsweb.cern.ch/) . Documents submitted to CDS receive a DOI and can therefore be easily found and referred to. The details of the publication policy will be stored in the document already quoted before.

Workpackage 2: Physics and Detectors

As part of WP2 (Physics and Detectors Requirements), simulated data, the beam-induced background in the detector at different centre-of-mass energies, will be produced. Primary muons interactions will be generated with common HEP Monte Carlo programs. Input files used for the generation will be made available on git. Beam-induce background data will be produced in collaboration with WP5, High energy complex.

Additional simulated data will be used and produced in the consortium as part of WP3 (Proton Complex), WP4 (Muon Production & Cooling), WP5 (High Energy Complex). For accelerator simulations, the data formats will be the ones produced by the simulation softwares, which are typically freely available (for example, as with Xsuite https://xsuite.web.cern.ch/).0

WorkPackage 3: Proton Complex

WP3 aims at producing a preliminary design for the linac and accumulator ring for the proton complex, as well a description of the R&D needed for the future. The WP will then develop lattices and simulations scripts using softwares such as Mad-X (https://mad.web.cern.ch/mad/), Orbit (https://web.ornl.gov/~holmesja1/JHolmes/ORBIT.html) and pyOrbit (https://github.com/PyORBIT-Collaboration/py-orbit). All those simulation codes are open source and freely available. For the linac side, the primary simulation code used is TraceWin (https://www.dacm-logiciels.fr/tracewin), which is a commercial software, however the output files can be exported as flat files with particle distributions, or binaries that can be read using freely available software; the lattice can also be made available as text file. The WP will also produce particle distribution files at interfaces to facilitate the work between tasks and WPs (mainly WP2). All the data will be in the format of the respective simulation program used and made available to the public either via CDS or gitHub.

Workpackage 4: Muon Production and Cooling

WP4 aims to deliver muon target and cooling models. The WP has a deliverable to amend the BDSIM software to enable simulation (http://www.pp.rhul.ac.uk/twiki/bin/view/JAI/BdSim). BDSIM is published under GPLv3 open source software license, and hence is fully open. The WP will also develop lattices and simulations using software such as FLUKA, RFTrack, G4Beamline and ICOOL. G4Beamline and ICOOL binaries and source code are freely available. RFTrack and FLUKA have restricted availability owing to Export Control Regulations that are outside of the control of the development teams, but our lattice files will be published as outlined above.

Workpackage 5: High Energy Complex

The aim of this WP is to simulate and optimise the high-energy acceleration and collider. This WP will produce optics files and survey files. The accelerator softwares used for the accelerator design are freely available: Mad-X (https://mad.web.cern.ch/mad/) and Xsuite (https://xsuite.web.cern.ch/). The multi-turn tracking of the particles will be performed with free softwares like Xsuite, BLonD (https://blond.web.cern.ch/), or PyHEADTAIL (https://github.com/PyCOMPLETE/PyHEADTAIL). The study of the matter interaction will be performed with the free code FLUKA (https://fluka.cern/). This WP will exchange input with WP2 (Physics and Detectors), with WP6 (RadioFrequency), and WP7(Magnets). The main interaction with WP2 is the machine-detector interface; the configuration of the detector will be integrated into the accelerator simulation codes. The exchange format with WP6 is mainly for the impedance models and can be given in ASCII tables. The exchange format with WP7 will be field maps provided by WP7.

Workpackage 6: RadioFrequency

The objective of this work package is to assess crucial feasibility issues and technological challenges of the RF systems. The study will concentrate on the two most challenging sections, the Muon Cooling Complex (MCC), and the muon acceleration stage of the High Energy Complex (HEC), for which a baseline concept of most critical RF components will be outlined.

The focus of the task related to the MCC is to lay out a conceptual design of the RF systems for the MCC, based on a consistent set of parameters for all RF cavities and associated systems to be integrated into the cooling cells of the MCC obtained from inputs given by WP4 and WP8. For the muon cooling section, one challenge already pointed out in the preliminary MAP study, is to achieve gradients of at least 30 MV/m in RF cavities that will be placed in magnetic fields of up to 13 T.

The task related to the muon acceleration stage aims to provide a preliminary design concept for the SRF cavities for acceleration in the Rapid Cycling Synchrotrons (RCS) of the HEC of the muon collider. For the acceleration stage of the HEC, the short muon lifetime requires the highest possible acceleration rate to reach energy gains on the order of 10 GeV per turn. This is foreseen to be provided with very high voltage SRF cavities. A suitable cavity technology, including the accelerating cavity type and shape, the cavity material, and the main RF frequency, will be determined for this system.

The conceptual radiofrequency designs of both the warm cavities for the MCC as well as the SRF cavities for the HEC, including related power couplers, will be performed with commercial software suites such as ANSYS HFSS (https://www.ansys.com/products/electronics/ansys-hfss), COMSOL (https://www.comsol.com/) and CST Studio (https://www.3ds.com/products-services/simulia/products/cst-studio-suite/). The results will be in form of summary tables of cavities parameters and performances, as well as 3D models in STEP format that will be used for preliminary mechanical design (in particular for WP8) using commercial suites like Dassault CATIA 3D, and stored in the CERN EDMS (https://edms.cern.ch).

Breakdown simulations for high gradient MCC cavities placed in high magnetic fields, combining electron tracking in EM fields, field emission and multipactor effects, will be based on a mix of commercial codes like CST Studio, free software suites like ASTRA (https://www.desy.de/~mpyflo/),
SuperFish (https://poisson-superfish.software.informer.com/7.1/),
RFtrack (https://gitlab.cern.ch/alatina/rf-track-2.0)
and XOOPIC (https://github.com/rinku-mishra/xoopic), as well as custom software using platforms like MatLab (https://www.mathworks.com/products/matlab.html).

At last, this work package also includes a task dedicated to conceptual designs of efficient RF power sources for MCC and HEC. These designs will be performed using a combination of a specific code developed at CERN, KlyC (https://cds.cern.ch/record/2812568) for quickly optimizing the bunching in the klystron interaction line, and CST Studio (as Particle In Cell solver) to compute the precise klystron performances.

Workpackage 7: Magnets

The aim of this WP is to produce conceptual design of the magnetic system of the muon collider, evaluate technology readiness, and identify required R&D to achieve the magnet performance need to reach beam specifications. This WP will produce conceptual magnet designs that will be documented in the CERN Engineering Document Management System (https://edms.cern.ch/). The conceptual designs will make use of commercial software such as ROXIE (https://roxie.docs.cern.ch/) and OPERA (https://www.3ds.com/products-services/simulia/products/opera/) for magnetic analysis, ANSYS (https://www.ansys.com/) and COMSOL (https://www.comsol.com/) for electro-mechanical analysis, as well as custom software developed for the specific application using platforms such as MatLab (https://www.mathworks.com/products/matlab.html). The results will be in the form of summary tables of magnet characteristics and performance, field maps, stress and strain maps. In a few instances, pre-engineering and integration studies will be produced. The 3D model will be designed with Dassault CATIA 3D, stored in the CERN Drawing Directory (https://edms-service.web.cern.ch/cdd/).

Workpackage 8: Cooling Cell

This WP will receive input from WP 4, WP 6 and WP7 to produce a full 3D model of a representative cooling cell. It will mainly use geometrical models coming from the simulation tools described in each WP. These data will be re-elaborated in order to perform the mechanical design of an integrated cell, made up of a Superconducting Solenoid, a radiofrequency cavity, one or two Absorbers and some diagnostic devices yet to be defined. The 3D model will be designed preferably with CATIA 3D or with Inventor (the decision will be taken in autumn). Files will be exchanged among different institutes wil 3D models in STEP format that will be stored either in the CERN Engineering Document Management System (https://edms.cern.ch/), or in the CERN PLM. Again, the decision will be finalised once the tool to be used will be agreed upon, and the DMP will be updated accordingly.

Will you re-use any existing data and how?

Yes, we received some input files from the US Muon Accelerator Program (see e.g. https://iopscience.iop.org/journal/1748-0221/page/extraproc46) as initial starting points to compare and optimise.

What is the origin of the data?

Coming from the different simulation codes used to study all the necessary mechanisms (see above: beam-induced background in the detector, accelerator physics, etc.).

What is the expected size of the data?

The expected size of these datasets is in the range of a few terabytes and they will be made available as open data at CERN.

To whom might it be useful ('data utility')?

WP6 (RF), WP7 (Magnets) and WP8 (Cooling Cell) will produce Engineering simulations and mechanical drawings that will be stored in Engineering databases with open access to the muon collider community, and to the general public when relevant. In particular we will publish publicly data necessary to referee or review scientific publications, annual reports etc…

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Public reports will be published via established document repositories, such as the CDS, arXiv, Zenodo and other institutional repositories of consortium partners if needed. Both CDS and Zenodo use persistent identifiers (DOIs) to identify the reports and are considered trusted repositories. Whenever possible, the research outputs will be linked e.g. to an article, further documentation, auxiliary measurements/datasets etc.

Beam-induced background data set produced for each center of mass energy will be published on Zenodo.

What naming conventions do you follow?

The naming conventions will depend on the nature of the data, and it is impossible to list in this document all the details. Typically a file describing the naming convention for the dataset will accompany each set of data for scientists and general public to be able to use them properly. It is to be underline that most of the data (apart from reports) need specialised competences and knowledge to be usefully exploited.

Where possible, the collaboration will make use of standard naming conventions used at CERN (as host institution). For instance the naming convention for accelerator components in an accelerator complex at CERN (see https://quality.web.cern.ch/quality/)

(Beam-induced background data will have in the name the center of mass energy, the version of the interaction region lattice and the nozzle identifier with which they are produced.) *I would not detail here for each workpackage….)*

Will search keywords be provided that optimize possibilities for re-use?

Yes.

Do you provide clear version numbers?

Yes. The Collaboration will generally make use of the the standards used by each simulation programme. Where different a document explaining the difference will be produced and referenced to in the DMP.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Metadata will be entirely managed by the repositories to be used (Zenodo, CERN CDS and its creation and storage will be entirely transparent to the community.

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Public reports and presentations will be made available through the repositories and the web sites of the related workshops, Conferences etc…, and will inherit the publication policy of the conference. Where legally possible, reports and papers will be made available also through Zenodo and/or CERN CDS (e.g. the Beam-induced background models will be available through Zenodo). The Consortium will privilege events and journals adopting the open access policy.

The Consortium will publish publicly only data and reports having satisfied its internal quality assurance procedures (references to those procedures will be added accordingly) . All datasets used for a given publication will be made available through the most convenient platform/repository so that readers can check by themselves the results.

Publications of certain datasets might be limited where their publication requires an overwhelming effort of documentation and polishing. This applies for instance to the detailed models of Beam-Induced background scenarios. It is expected that several iterations will be necessary between WP5 and WP2 to get to a satisfactory situation, and it will be just impossible to publish all the intermediate steps. So the Consortium will only publish one or maximum two versions per year as deemed useful. The same might apply to other datasets.

All articles will be submitted to open access journals. Well justified exceptions may be agreed by the Consortium Governing Board in one of its meetings..

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

The Consortium is starting its studies from the configurations that were kindly provided by the members of the US Muon Accelerator Programme (https://iopscience.iop.org/journal/1748-0221/page/extraproc46). Input files for simulations have their own publication rules and therefore we will respect those rules. Original output generated directly by the Consortium will be published in Open Access or in Open Source.

For instance we plan to create repositories (e.g. as here for the lattices of the different machines https://acc-models.web.cern.ch/acc-models/mc/ and here for the impedances of the different machines https://muc-impedance.docs.cern.ch/).

As another example Beam-induced background files produced by the Consortium will be uploaded in a compressed format on Zenodo.

What methods or software tools are needed to access the data?

All software tools and methods to be used have been presented in par. 1.

Is documentation about the software needed to access the data included?

All the data used and produced can be used by scientists having a decent knowledge of the fields involved (accelerator physics, Magnet technology etc…). A limited documentation intended for experts of such domain will always be produced. In some cases the consortium will provide training material, for instance on BIB.

Is it possible to include the relevant software (e.g. in open source code)?

The consortium will always reference the software being used for every activity, and will provide links to download the software from the original producer. If the Consortium will modify or produce new software, such software will be made available through the Consortium website (https://mucol.web.cern.ch/)..

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

The platform to publish data, reports and presentations will be many, the main ones being:

Zenodo, CERN CDS, CERN EDMS, CERN GitHub etc… (see par. 1)

Is there a need for a data access committee?

When necessary, this function will be performed by the Consortium Governing board, advised by the Scientific Advisory Committee. We do not expect however any need for such a committee.

Are there well described conditions for access (i.e. a machine readable licence)?

Yes, we will publish through standard Open access and Open source licences

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

The Consortium will not keep track of the identity of people accessing the data, apart from very specific cases. For instance the Beam Induced Background model will be published through a ZENODO group registration on a flat file including the name and purpose of the downloading, this to ensure the right of students who have written such models to publish their work before those sources become public.

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

The repositories currently identified for publishing the research outputs, use standardized metadata schemas (e.g. Datacite Metadata Standard) that enable easy discovery. Whenever possible and applicable, community standards will be used, e.g. to submit data to the HEPData repository. Also for accelerators data, the formats are typically easily exchangeable and reusable. For 3D models, the STEP format will be privileged for exchange. For many of the technologies used exist field specific standards to exchange data.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

This will be matter of negotiations between the different teams where a standard does not exist already. For instance for beam-induced background data produced in this project will have the same metadata vocabulary and will be produced by using the same methodology already applied by the previous US project MAP. Therefore, everything will be interoperable inside the HEP community.

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Yes.

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?

We do not anticipate to be in the need of developing new ontologies or vocabularies.

2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

The consortium will preserve its assets at CERN, by using the standard and trusted storage facilities and software tools already developed for the LHC machine and experiments. To further the reusability of research outputs, the consortium will aim at linking its research outputs to provide more context to the individual assets, e.g. datasets on HEPdata and articles are linked, software components are linked to datasets or articles as well.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

As soon as possible.

Are the data produced and/or used in the project usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Yes for further studies/projects on accelerators and colliders.

How long is it intended that the data remains re-usable?

At least one decade (still to be defined).

Are data quality assurance processes described?

Still to be defined.

Beam-induced background data will follow an internal review that is under preparation by WP2 and WP5. The steps of the data quality assurance will be published in the MuCol web page.

Further to the FAIR principles, DMPs should also address:

To be defined/checked if other aspects are needed.

3. Allocation of resources

What are the costs for making data FAIR in your project?

To be defined.

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

To be defined.

Who will be responsible for data management in your project?

Elias Métral.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

CERN has in place facilities to preserve data on the long term and the same facilities will be used for this project.

4. Data security

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

CERN has in place facilities to preserve data on the long term and the same facilities will be used for this project.

Is the data safely stored in certified repositories for long term preservation and curation?

Probably yes, to be checked with the CERN storage service.

Yes for Zenodo.

5. Ethical aspects

Are there any ethical 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).

We don’t think so.

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

There is no personal data.

6. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

So far, not.

What does it mean? Out of the project?

7. Further support in developing your DMP

The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.

The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

Registry of Research Data Repositories

Some repositories (like Zenodo, an OpenAIRE and CERN collaboration) allow researchers to deposit both publications and data, while providing tools to link them.

Other useful tools include DMP online and platforms for making individual scientific observations available such as ScienceMatters.

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identif persistent and unique identifiers such as Digital Object Identifier • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no type of metadata will be created and how

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

HISTORY OF CHANGES

Version	Publication date	Change
1.0	13.10.2016	▪ Initial version