DATA MANAGEMENT PLAN

PROJECT	
Project number:	101094300
Project acronym:	MuCol
Project name:	A Design Study for a Muon Collider complex at 10 TeV centre of mass

DATA MANAGEMENT PLAN	
DOI	10.5281/zenodo.10054352
CERN EDMS number:	2974112
Date:	27/10/2023
Version:	1.0

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

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

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 10 Tev, 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, the MuCol website or CERNBox as proof of achievement of the corresponding Milestone. These documents will also be used as basis for further publications, yet to be determined.

Each Work Package (WP) will produce simulation data with software typical of the domain of the WP. 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 WP. 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 soon 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 WP are the following:

WP 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. Other working documents which are not considered reports may be stored in CERNBox (https://cernbox.cern.ch/).

WP 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 software, which are typically freely available (for example, as with Xsuite https://xsuite.web.cern.ch/).

WP 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 software 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.

WP 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.

WP 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 software 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 software 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.

WP 6: RadioFrequency

The objective of this WP 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 conceptual RF 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 Multiphysics® (https://www.comsol.com/) and CST Studio Suite® (https://www.3ds.com/products-services/simulia/products/cst-studio-suite/). However, for the initial optimization of the cavity shapes in the different accelerator rings, the free FEM-code SuperLANS (https://www.classe.cornell.edu/rsrc/Home/Research/SRF/SuperLANS/SRFD960814-09.pdf) will be used. During this optimization phase, the analysis of wakefields in the cavities is conducted by using the free 2D ABCI code (https://abci.kek.jp/abci.htm). The optimization of couplers in the RCS accelerators will be performed in the free circuit simulator QucsStudio (http://qucsstudio.de/) and CST Studio Suite. Longitudinal beam stability studies are implemented with the macroparticle tracking code BLonD (https://gitlab.cern.ch/blond/BLonD). The results will be in form of summary tables of cavity 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 Engineering Document Management System (EDMS: https://edms.cern.ch). The CST Studio Suite files will also be made available through CERN EDMS. The results of the optimization of the RCS cavities will be shared with WP5 and WP7.

Breakdown simulations for high-gradient MCC cavities placed in high magnetic fields, combining electron tracking in electromagnetic fields, field emission and multifactor effects, will be based on a mix of commercial codes like CST Studio Suite, 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), Multipac (https://accelconf.web.cern.ch/e08/papers/mopp137.pdf) and XOOPIC (https://github.com/rinku-mishra/xoopic), as well as custom software using platforms like MATLAB (https://www.mathworks.com/products/matlab.html) and Python (https://www.python.org/).

At last, this WP 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 Suite (as Particle-In-Cell solver) to compute the precise klystron performances.

WP 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 EDMS (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/).

WP 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 with 3D models in STEP format that will be stored either in the CERN EDMS (https://edms.cern.ch/), or in the CERN Product Lifecycle Manager (PLM). Again, the decision will be finalised once the tool to be used will be agreed upon, and the Data management plan (DMP) will be updated accordingly.

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

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.

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

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

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

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

Will data be identified by a persistent identifier?

Both CDS and Zenodo use persistent identifiers (DOIs) to identify the reports

The naming conventions will depend on the nature of the data. 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/)

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.

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. For what concerns input files to simulation programs, the format is determined by the code.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Yes.

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?

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

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Yes

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

CDS and Zenodo both attribute a DOI to every document stored.

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.

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.

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.

At present we do not foresee to apply an embargo on any part of the project. Should the need arise, we will apply it for the time needed to go through the administrative procedure to request a patent.

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

Yes, for what concerns publications we will publish through standard Open access and Open source licences.

For what concerns data (input files, drawings, etc.) the naming conventions and the access protocol will depend on the nature of the data. 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 underlined 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/)

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

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.

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/).

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.

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

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.

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

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

All publications in CDS and ZENODO will be preserved indefinitely. This applies as well to data such as input files (either on Zenodo, github or other similar services) drawing (on CERN EDMS), etc.

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)?

All the data used and produced can be used by scientists having 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 Beam Induced Background.

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/).

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?

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.

The vocabularies, standards, formats, or methodologies to follow to make our data interoperable 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.

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?

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

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

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). Reference to papers pubished by the authors of the studies of those data.

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.)?

We will use standard tools and methodologies of the fields concerned (detector physics, Radiofrequency Engineering, etc.). Data will be published with reference to the code used to perform simulations and its version, and any other information necessary to reproduce or further develop the models.

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?

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.

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

Yes, for further studies/projects on accelerators and colliders, and for as long as CERN will maintain its documentation repositories.

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

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

Describe all relevant data quality assurance processes.

MuCol together with the International Muon Collider Collaboration has set up a Scientific Advisory Committee (SAC) and a Publication and Speakers Committee (PSC), the second with the specific task to provide quality control of any publication in peer reviewed journals.

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.

The research object of MuCol will produce no sensitive data and therefore does not require specific security or ethical provisions, other than what normally used in scientific studies, and that are implemented through using standard tools and repositories, and submitting results to the SAC and the PSC.

3. 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.).

Intangible assets produced by MuCol will be published in Open Access or Open source. Eventual hardware - prototypes, if useful, will remain in the Institute that produced it and its re-use for scientific purposes, if relevant, will have to be agreed by the governing board of MuCol, or by the IMCC Collaboration Board after the end of MuCol.

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.

4. 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.)?

The cost will be transparent to MuCol as we will make use of standard tools and repositories.

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)

No costs are foreseen to be covered by MuCol.

Who will be responsible for data management in your project?

Elias Métral (CERN) in his function of Communication and Dissemination Officer.

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)?

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

5. 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)?

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

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes.

6. 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).

We are not aware of any potential or existing legal issues.

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

There is no personal data.

7. 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)?

So far, no.

8. Acknowledgement

Funded by the European Union (EU). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or European Research Executive Agency (REA). Neither the EU nor the REA can be held responsible for them.

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	27.10.2023	Initial version