

# CC-IN2P3 (Centre de Calcul - Institut national de physique nucléaire et de physique des particules du CNRS): CC-IN2P3 - DMP template (english) - General

## 1.1. General - Topic

What is the main research question of the project?

Please give some keywords describing the research question.

## 1.2. General - Research field

Which research field(s) does this project belong to?

- Natural Sciences / Condensed Matter Physics
- Natural Sciences / Particles, Nuclei and Fields
- Natural Sciences / Astrophysics and Astronomy
- Natural Sciences / Mathematics
- Natural Sciences / Geophysics and Geodesy
- Natural Sciences / Geochemistry, Mineralogy and Crystallography
- Natural Sciences / Water Research
- Engineering Sciences / Production Technology
- Engineering Sciences / Mechanics and Constructive Mechanical Engineering
- Engineering Sciences / Process Engineering, Technical Chemistry
- Engineering Sciences / Materials Engineering
- Engineering Sciences / Materials Science
- Engineering Sciences / Computer Science
- Engineering Sciences / Construction Engineering and Architecture
- Life Sciences / Medicine
- Life Sciences / Neurosciences
- Life Sciences / Agriculture, Forestry and Veterinary Medicine
- Natural Sciences / Molecular Chemistry
- Natural Sciences / Physical and Theoretical Chemistry
- Natural Sciences / Analytical Chemistry, Method Development (Chemistry)
- Life Sciences / Plant Sciences
- Humanities and Social Sciences / Fine Arts, Theatre and Media Studies
- Humanities and Social Sciences / Linguistics
- Humanities and Social Sciences / Literary Studies
- Humanities and Social Sciences / Theology
- Humanities and Social Sciences / Philosophy
- Humanities and Social Sciences / Educational Research
- Humanities and Social Sciences / Social Sciences
- Humanities and Social Sciences / Economics
- Natural Sciences / Biological Chemistry and Food Chemistry
- Natural Sciences / Polymer Research
- Natural Sciences / Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas
- Natural Sciences / Atmospheric Science, Oceanography and Climate Research
- Humanities and Social Sciences / Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies
- Humanities and Social Sciences / Psychology
- Humanities and Social Sciences / Jurisprudence
- Life Sciences / Basic Research in Biology and Medicine
- Life Sciences / Zoology
- Life Sciences / Microbiology, Virology and Immunology
- Natural Sciences / Chemical Solid State and Surface Research
- Natural Sciences / Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics
- Natural Sciences / Geology and Palaeontology
- Natural Sciences / Geography
- Engineering Sciences / Heat Energy Technology, Thermal Machines, Fluid Mechanics
- Engineering Sciences / Electrical Engineering and Information Technology
- Humanities and Social Sciences / History
- Engineering Sciences / Systems Engineering
- Humanities and Social Sciences / Ancient cultures

### Recommendations:

The list of disciplines follows the [subject classification of the DFG \(German Research Foundation\)](#).

Which persons or institutions are responsible for the project coordination?

## 1.3. General - Project schedule

When does the project start?

When does the project end?

## 1.4. General - Project coordination

Which persons or institutions are responsible for the project coordination?

## 1.5. General - Project partner(s)

Project partner(s)

Does your institution have rules or guidelines for the handling of research data? If yes, please briefly outline them and refer to more detailed sources of information if necessary. Please also indicate, if the rules / guidelines are mandatory or optional.

*Recommendations:*

More and more universities and scientific institutions adopt research data management policies. These contain, among other things, recommendations and / or demands concerning the handling of research data by researchers of the institution.

Who is/are the contact person(s) for data management questions?

*Recommendations:*

Please give the name and an email address.

## 1.6. General - Funding

Who is funding the project?

Is the project within a special funding programme?

Does the funder have rules or recommendations for data management? If yes, please briefly outline them and refer to more detailed sources of information if necessary. Please also indicate, if the rules / guidelines are mandatory or optional.

*Recommendations:*

Funders of research also increasingly specify requirements regarding the management of research data in funded projects.

## 1.7. General - Others requirements I

Are there requirements regarding the data management from other parties (e.g. the scholarly/scientific community)?

- No
- Yes
- To be clarified (specify your answer in additional information)

*Recommendations:*

Examples of discipline-specific requirements

## 1.8. General - Others requirements II

Which are these additional requirements regarding data management?

*Recommendations:*

Please briefly outline them and refer to more detailed sources of information if necessary. Please also indicate, if the rules/guidelines are mandatory or optional.

## **CC-IN2P3 (Centre de Calcul - Institut national de physique nucléaire et de physique des particules du CNRS): CC-IN2P3 - DMP template (english) - Datasets**

### **2.1. Content classification - Dataset(s)**

What kind of dataset is it?

*Recommendations:*

Please briefly describe the data type and/or the method used to create or collect the data, for example:

- quantitative online survey
- 3D model / digital reconstruction of a stone age settlement
- software developed within the project.

### **2.2. Content classification - Data origin**

Is the dataset being created or re-used?

- Reused
- Created

If re-used, who created the dataset?

If re-used, under which address, PID or URL can the dataset be found?

### **2.3. Content classification - Reuse**

Which individuals, groups or institutions could be interested in re-using this dataset? What are possible scenarios?

### **2.4. Content classification - Reproducibility**

Is the dataset reproducible in the sense that it could be created/collected anew in case it got lost?

- No or only with disproportionate cost or effort
- Yes, with moderate, but reasonable effort
- Yes, with little effort
- No, the data is not reproducible per se

*Recommendations:*

Some data can, technically, be created anew at any time, as is the case with scientific experiments or digitised versions of analog objects (as long as the originals are still there and in good shape). However, this can consume a considerable amount of time and cost. With respect to long-term preservation, the effort of re-creation has to be weighed up against the effort of long-term preservation. Other data cannot be collected or created anew.

### **3.1. Technical classification - Data collection**

When does data collection or creation start?

When does data collection or creation end?

When does data cleansing/data preparation start?

When does data cleansing/data preparation end?

When does data analysis start?

When does data analysis end?

### **3.2. Technical classification - Data size**

What is the actual or expected size of the dataset?

- More than 100 TB
- Not yet defined
- 1 GB to 1 TB
- 1 TB to 100 TB
- Less than 1 GB
- Exact size (specify your answer in additional information)

How much data is produced per year? *Optional. This is only of concern if the data production rate reaches TB scale.*

### **3.3. Technical classification - Formats**

Which file formats are used?

*Recommendations:*

When choosing a data format, one should consider the consequences for collaborative use, long-term preservation as well as re-use. It is advisable to prefer formats that are standardised, open, non-proprietary and well-established in the respective scholarly community.

### 3.4. Technical classification - Tools

Which tools, software, technologies or processes are used to generate or collect the data?

*Recommendations:*

This information is necessary to be able to reconstruct the process by which the data was generated. It is also a prerequisite to judge the objectivity, reliability and validity of the dataset. For reproducible data, it is also required to re-generate the data if need be.

Which software, processes or technologies are necessary to use the data?

*Recommendations:*

To be able to re-use data (e.g. to replicate studies, for meta analysis or to solve new research questions), along with the data the software, equipment and knowledge about special methods to use the data are required. Just as with the formats, the recommendation is: the more standardised, open and established, the better for re-use.

Is documentation about relevant software needed to use the data?

### 3.5. Technical classification - Versioning

Are different versions of the dataset created?

Which versioning strategy is applied for this dataset?

*Recommendations:*

Please briefly describe the project-internal regulations for the versioning of data sets (e.g.: What kind of changes require a new version? How are changes documented? What are the naming rules for different versions?)

Which technology or tool is used for versioning?

- Not yet decided
- Simple copying
- Version control system
- Others (specify your answer in additional information)

### 4.1. Data usage - Usage scenarios

How / for what purpose will this dataset be used during the project?

How often will this dataset be used?

To what extent will infrastructure resources be required (e.g. CPU hours, bandwidth, storage space... etc.).

- The following infrastructure resources are needed (specify your answer in additional information)
- Infrastructure resources of the usual workplace equipment are sufficient

Are there actual or potential usage scenarios that could benefit from support by a data management or IT expert, or that even require such support?

- No
- Yes (specify your answer in additional information)

### 4.2. Data usage - Data organisation

Where is the dataset stored during the project?

Under which URL can the dataset be accessed during the project?

Are there internal project guidelines for a consistent organisation of the data? If so, where they are documented?

- Not yet
- Yes
- No

Is there a internal project guideline for naming the data? If so, please briefly outline the naming conventions and, if necessary, link to the documentation.

- Not yet
- No
- Yes

### 4.3. Data usage - Data storage and security

Who is allowed to access the dataset?

How and how often will backups of the data be created?

*Recommendations:*

This question refers to backups while the data is being worked with. Questions of long-term preservation will be addressed in the respective section.

Who is responsible for the backups?

*Recommendations:*

This question refers to backups while the data is being worked with. Questions of long-term preservation will be addressed in the respective section.

Which measures or provisions are in place to ensure data security (e.g. protection against unauthorized access, data recovery, transfer of sensitive data)?

#### 4.4. Data usage - Interoperability

Is this dataset interoperable, i.e. allowing data exchange and re-use between researchers, institutions, organisations, countries etc.?

- The dataset adheres to standardised formats
- The dataset can easily be re-combined with different datasets from different origins
- Other aspects in terms of interoperability
- The dataset is usable with available (open) software applications or software applications that are established and widely used in the respective community

#### 4.5. Data usage - Data sharing and re-use

Will this dataset be published or shared?

- Yes, externally for everyone
- Yes, internally with everyone, as long as they don't pass on the data
- No
- Yes, externally limited with individual approval

If no, please explain why not. Please differentiate between legal and contractual reasons and voluntary restrictions.

If yes, under which terms of use or license will the dataset be published or shared?

- No derivative work (ND)
- Other (specify your answer in additional information)
- Non-commercial (NC)
- Attribution (BY)
- Public domain (CC0)
- Share-alike (SA)

*Recommendations:*

The options refer to the licenses of the creativecommons.org

If there are any restrictions on the re-use of this dataset, please explain why.

When will the data be published (if they are)?

#### 4.6. Data usage - Collaboratif work

Will the data be collaboratively used?

- Yes, by several persons at various institutions
- Yes, by multiple persons of the same workgroup at the same institution
- No

Which platform / tools is / are used for collaboratively working on data and publications?

How is the collaborative work on the same files organised?

#### 4.7. Data usage - Quality assurance

Which measures of quality assurance are taken for this dataset?

#### 4.8. Data usage - Data integration

Is the integration between the re-used and newly created data ensured? If yes, by which means?

#### 4.9. Data usage - Costs

What are the personnel costs for data management associated with the creation or acquisition of data in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs for data management associated with the creation or acquisition of data in the project?

*Recommendations:*

Please estimate the effort in **Euro**.

What are the personnel costs for data management associated with the usage of data in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs for data management associated with the usage of data in the project?

*Recommendations:*

Please estimate the costs in **Euro**.

What are the personnel costs associated with data storage and data security in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel costs associated with the storage of the data sets during the project?

*Recommendations:*

Please estimate the costs in **Euro**.

## 5.1. Metadata and Referencing - Metadata

Which information is necessary for other parties to understand the data (that is, to understand their collection or creation, analysis, and research results obtained on its basis) and to re-use it?

- Creation process
- Agents
- Sources
- Time
- Technology
- Content
- Identifiers
- Location
- Documentation of the software necessary to use the data
- Other (specify your answer in additional information)
- Methodology

Which standards, ontologies, classifications etc. are used to describe the data and context information?

- No fixed system for the description is used
- Other
- Discipline-specific standards, classifications etc. are used
- A custom description system is used (please briefly outline and, if necessary, indicate where it is documented in more detail)

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?

*Recommendations:*

This information is needed for a Horizon 2020 data management plan.

Which metadata are collected automatically?

Which metadata are collected semi-automatically?

Which metadata are collected manually?

Are metadata and context information being checked for correctness and completeness?

- Automatic check for completeness
- Manual check for correctness
- Other
- Manual check for completeness

Who is responsible for documenting the metadata and context information and for checking if they are correct and complete?

## 5.2. Metadata and Referencing - Metadata costs

What are the personnel costs associated with the the creation of metadata and context information in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs associated with the creation of metadata and context information in the project?

*Recommendations:*

Please estimate the costs in **\*\*Euro\*\***.

### **5.3. Metadata and Referencing - Structure, granularity and referencing**

What is the structure of the data? How are the individual components of the dataset related to each other? How is the dataset related to other datasets used in the project?

### **5.4. Metadata and Referencing - Persistent identifiers(s)**

Will persistent identifiers (PIDs) be used for this data set?

Which system of persistent identifiers shall be used?

- PURL
- Handle/DOI
- ISLRN
- ARK
- URN
- Other (specify your answer in additional information)

Which (sub-) entities/sub units should be referenced using identifiers? Which of those identifiers should be persistent and citable?

Who is responsible for the maintenance of the PIDs and the object maintenance (i.e. who is responsible notifying the PID-Service about object relocation and the new address)?

*Recommendations:*

A prerequisite for PIDs to work as promised is that they - as well as the objects they refer to - are maintained in a continuous and reliable way. This means, for example, that if the object location changes, this information is updated. When the data are stored in a data centre or repository, these tasks are usually taken care of by the data centre / repository. However, to be sure, the responsibilities should be checked beforehand.

### **5.5. Metadata and Referencing - PIDs Costs**

What are the personnel costs associated with of persistent identifiers in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs associated with persistent identifiers in the project?

*Recommendations:*

Please estimate the effort in Euro.

## **CC-IN2P3 (Centre de Calcul - Institut national de physique nucléaire et de physique des particules du CNRS): CC-IN2P3 - DMP template (english) - Legal & Ethics**

### **6.1. Legal and ethics - General legal issues**

Does the legal situation of different countries have to be considered?

*Recommendations:*

If you answer this question with "Yes", please get in touch with the legal department or a respective contact person at your institution to clarify if this has consequences for the project and its data management and if yes, what consequences these are.

### **6.2 Legal and ethics - Personal data**

Does this dataset contain personal data?

*Recommendations:*

European Data Protection Law may apply.

### **6.3. Legal and ethics - Data protection**

Which law applies with respect to the aspects of data protection in the project?

- European Union
- French Republic
- Others (specify your answer in additional information)

*Recommendations:*

It depends on the kind of institution which law applies.

### **6.4. Legal and ethics - Sensitive data**

Does the dataset contain information on racial and ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sex life?

*Recommendations:*

These kinds of data are considered particularly sensitive and require even more extensive safeguards. If you answer this question with "Yes", please get in touch with the data protection officer of your institution to check which additional protection measures are necessary.

Will the data be anonymised or pseudonymised?

- Yes, after the data analysis/before publication
- Yes, during the collection
- No
- Yes, before/at the beginning of the data analysis

To what extent is the "informed consent" obtained from the persons concerned?

- The "informed consent" is not obtained
- Only for analysis / use of the data within the project
- For analysis / use of the data within the project as well as for re-use

*Recommendations:*

Basically, the collection, processing, archiving and publication of personal data is only admissible, when the "informed consent" of the person in question has been obtained.

If no "informed consent" is obtained, please give the reasons for not doing so.

Where and how is the "informed consent" documented?

By when will the (un anonymised or unpseudonymised) original data be safely deleted?

### **6.5. Legal and ethics - Other sensitive data**

Does this dataset contain sensitive data other than personal data?

*Recommendations:*

Examples are data that contain trade or business secrets or geoinformation on endangered species.

If yes, please describe the non-personal sensitive data used in the project.

### **6.6. Legal and ethics - Sensitive data costs**

What are the personnel costs associated with the anonymization of sensitive data in the project?

*Recommendations:*



Please estimate the effort in person months.

What is the amount of non-personnel-costs associated with the anonymization of sensitive data in the project?

*Recommendations:*

Please estimate the costs in **\*\*Euro\*\***.

What are the personnel costs associated with other (non-technical) security measures for sensitive data in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs for other (non-technical) security measures for sensitive data for the project?

*Recommendations:*

Please estimate the costs in **\*\*Euro\*\***.

## 6.7. Legal and ethics - Official approval

Has the project been approved by a research ethics committee? *Specify your answer in additional information.*

- Not yet, it will be handed in for review by
- Yes, approved under obligations which will be complied in the following way
- Yes, reviewed and approved by the following committee
- No, a review is not necessary, because
- Not yet, but it is already in the review process

Is a statutory approval/permit needed for the research? *Specify your answer in additional information*

- Yes. The permit will be applied for by
- No
- Yes. The permit has been applied for on
- Yes. The permit has been received.

If yes, which permit?

If yes, which is the responsible agency?

Is a data access committee needed to handle access requests to the published data of the project?

## 6.8. Legal and ethics - Intellectual property rights I

Does the project use and/or produce data that is protected by intellectual or industrial property rights?

*Recommendations:*

Data or software can be subject to intellectual or industrial property rights. Applicable laws differ broadly even within EU.

## 6.9. Legal and ethics - Intellectual property rights II

Does copyright law apply to this dataset?

- Work of literature, scholarship or the arts
- Collected edition or database work
- Translation or other edition of a work
- Other (specify your answer in additional informational)
- No

Do other intellectual property rights apply to this dataset?

- Patent
- Utility model
- Trademark
- Integrated circuit layout design protection
- Geographical indication
- Other (specify your answer in additional informational)
- No
- Plant variety rights protection
- Registered design

Was investigated who the rights owner is?

- Yes
- No

## 6.10. Legal and ethics - Intellectual property rights costs

What are the personnel costs associated with intellectual property rights in the project?

*Recommendations:*

Please estimate the effort in person months.

What is the amount of non-personnel-costs regarding intellectual property rights in the project?

*Recommendations:*

Please estimate the costs in **\*\*Euro\*\***.

## **CC-IN2P3 (Centre de Calcul - Institut national de physique nucléaire et de physique des particules du CNRS): CC-IN2P3 - DMP template (english) - Storage and long term preservation**

### **7.1. Storage and long term preservation - Selection**

What are the criteria/rules for the selection of the data to be archived (after the end of the project)?

Who selects the data to be archived?

### **7.2. Storage and long term preservation - Long term preservation**

Does this dataset have to be preserved for the long-term?

What are the reasons this dataset has to be preserved for the long-term?

- Documentation, because it is relevant to society
- Self-commitment
- Other (specify your answer in additional information)
- Used in a publication / proof of good scientific practice
- Re-use in subsequent projects or by others
- Legal obligations

How long will the data be stored?

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

Where will the data (including metadata, documentation and, if applicable, relevant code) be stored or archived after the end of the project?

- Has not yet been decided
- Discipline specific data center
- Own institution
- Other (specify your answer in additional information)
- Generic data center
- IN2P3 Computing Centre

Is the repository or data centre chosen certified (e.g. Data Seal of Approval, nestor Seal or ISO 16363)? (If the dataset is archived at several places, you may answer this question with yes, if this applies to at least one of these.)

Have you explored appropriate arrangements with the identified repository?

*Recommendations:*

(original question from Horizon 2020 FAIR Data Management Plan)

Shall there be an embargo period before the data are made available?

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

By when will the data be archived?

### **7.3. Storage and long term preservation - Long term preservation costs**

What are the personnel costs associated with long-term preservation for the project?

*Recommendations:*

Please estimate the effort in person months.

How will the data management costs of the project be covered?

What is the amount of non-personnel-costs regarding long-term preservation for the project?

*Recommendations:*

Please estimate the costs in \*\*Euro\*\*.