

### ATTENTION

This is **NOT** the application form to access the RES archival resources. This document has been created only as a guide showing the information needed to fill the application form online. All the applications must be presented through the online form available at <https://www.bsc.es/res-intranet/>.

## Data Management Plan Template

Data projects are evaluated for their scientific relevance and technical feasibility. Be as precise as possible. You should provide enough information to allow the evaluation of the project. Please write only the information you have permission to disclose. If you do not have the answers yet for an element, write at which point you will have a response or write N/A if it does not apply to you.

### Project title

The Data Project title you choose will be used to refer to this application in all the communications to the Applicant.

### Applicant Information

Name, affiliation and contact email of the person that will be contacted and notified for all that refers to the Data Project.

### Project general information

The desired project start date can go from April, if data are already available, or a later date, up to one year, if the activity of data generation has been already planned, or ongoing, but not finalized. In this case, please put a reference to the approved project/activity in the technical description of the data project.

### Total storage (in TB)

Give the total(maximum) amount of data you are requesting to store during the project, including the backup space required. This value will be checked with the availability of the resources in the nodes. It will permit the correct allocation of resources. In the call terms, the minimum size of storage to require to pass the evaluation is indicated.

### Resources needs per year

Give an estimation of storage, and eventually of computing, required for each of the first three years of project. Separate the needs required in a file storage, in an object storage and the backup volume required (0 if not needed). Nodes offer a limited computing

capacity to data projects for data exploitation. Please check the availabilities in the preferred node and indicate the HPC (in CPU/h) required, or number of VMs (middle size flavor). Put 0 if no computing is needed.

### Host nodes

Indicate your preferred host node (no binding). Choose the host name from the dropdown list, considering the storage type offered by each node. In case there is no preferred node, select “Any”. The host node for each project will be selected by the Access Committee, considering the selection and motivation given.

### Research Team PI

Name, current affiliation and short CV of the data project Principal Investigator (PI).

In the Publications box, a maximum of 5 publications references (including DOIs) is admitted, to support and confirm the PI team experience and scientific excellence of the data project.

### Funded projects related to the proposed data project

Reference to up to 5 projects, related to the activity proposed for the data project (data preparation aspects, data production, exploitations activities not included in the data project, ...). Please explain for each project which is the funded aspect that is related to the data project. This information will be considered to avoid double funding and to justify the data project in the context of other funded activities related to it.

### Scientific Project Description

Short description of the project scientific value and relevance of the data sets and of the activity of data exploitation for the scientific community.

Add any project technical details in the “Technical Project Description”

### Motivation

Short motivation for asking resources to the RES, with clear references to the related funded projects, and expected impact.

### Technical Project Description

Short description of the project technical value. Add any project scientific details in the “Scientific Project Description”

Please include the following elements

- Describe the technical infrastructure you will need based on your project requirements. Include details such as amount of storage, CPUs, VM and network you may need to perform data your processing and analysis tasks. Cross-check the feasibility of your project with the description of the node you are requesting as host.

- Briefly describe the data collection stage including the data sources and privacy related aspects of data cleaning (like anonymization). List the software and services for the next stages of your data exploitation, processing and analysis (for example, Hadoop or Spark for data processing tasks).
- Describe the output of your data processing cycle. Additionally, please describe how do you expect to share this data with the scientific community (if applies). For example, which data sharing tools you will use, access levels, etc.
- Describe how your project fits to the FAIR (Findable, Accessible, Interoperable, Reusable) principles. Mention if and how your data fulfil any of these principles, and the technical aspects of the FAIRfication process. For example, use of extra software or additional stages in the data processing cycle for reuse of data.

### **Number of datasets**

Give the number of datasets included in the project and described in the Dataset description section. Each dataset is a group of files, of the same or different formats, with homogeneous or heterogeneous content, with the same technical needs in terms of storage, permissions, services, etc.

### **Datasets description**

It is requested to identify and describe the main characteristics and technical needs of the datasets (at least 1, and up to 3) of the data project.

#### **Dataset Basic information**

Name and ID (externally given, or chosen in the context of this project) of the dataset. It will be used to refer to it. The Author can be a person, a project, an institution, or whatever indicates the data provenance.

#### **Dataset volume**

Indicate the total dataset size. If the total size of the dataset is not known at the moment of the application, give your best guess.

#### **Data format**

Give information about the data format. It is encouraged to use standard (or at least well documented) formats, to allow researchers to reuse the project data.

#### **Storage Type main Policy**

Indicates your preferred choice:

- File Tape, or HMS (Suitable for very large datasets/slower access)
- File HDD (Fast access)
- Object Storage Disk (Redundant, high availability storage).

RESTRICTION TO ONE CHOICE PER DATASET. Note: Each node varies in the resources offered. Please consult the Nodes Resources page before making a Storage Type selection to ensure that the node you want to select, matches your requirements.

### **Backup policy**

Indicate if backup for data/subsets is needed throughout the project's length. The volume of the backup of the data in the same node is considered included in the total data volume request. If you need to backup your data or part of your data, you can request it as an additional service. Note as well that some nodes offer storage for projects backup, and their capacity varies. Please consult their characteristics on the Nodes Resources page.

### **Legal/ethical restrictions**

Indicate if the data is subject to any legal or ethical restriction (for example, in case of personal data, clinical data, genomic information, ...). Describe the actions taken for your dataset not to have any ethical restrictions and how did it comply with such legally.