# MLOps platform prototype user manual

Marcelo Peñafiel

---

## 1 Introduction

This document is a brief description on how to use the prototype. Since this is a prototype some bugs are expected, any feedback is appreciated.

## 2 Prerequisites

Make sure to execute the *install.sh* script. If you are not on a ubuntu base distribution use the script as a guide to install the dependencies.

### 2.1 GraphDB preparation

The platform uses web semantic technologies as data model. Make sure to install GraphBD and listening in the default port.

If you haven't done it yet, follow the next instructions to import the ontology to GraphDB:

1. Create a new repository. To do this go to Setup ¿ Repositories in the sidebar and create a new repository with name *mashitesis*. Make sure that the repository is running while using the platform.
2. Go to Import in the side bar and click the option *Upload RDF files* and select the file *DMProcess.owl* located at the root of the project.
3. Click the import button and fill the settings with:
   - **Base IRI:** http://www.semanticweb.org/DM/ontologies/MLOpsExp
   - Select the **Named graph** option
   - **Target graphs:** http://mlops/data
4. Confirm the import by clicking the import button

Now everything is ready to start using the platform

## 3 Dataset Management

The system allows the creation and management of different datasets. To do this, a dataset needs to be created first. This *dataset manager* acts as a way to group data into different versions, so that if the same process needs to be tested with different variations of the data, the changes can be made immediately without losing previous information.

To create a dataset, follow the steps illustrated in Fig. 1:

1. Select the dataset manager from the main screen. This will open a list of the different datasets that have been imported into the system.
2. Click on the *New Dataset* button. This will open a panel asking for the name of the new dataset to be created.
3. Once the dataset name is entered, click on the confirm button, and a new dataset will be created, from which different versions can be created.

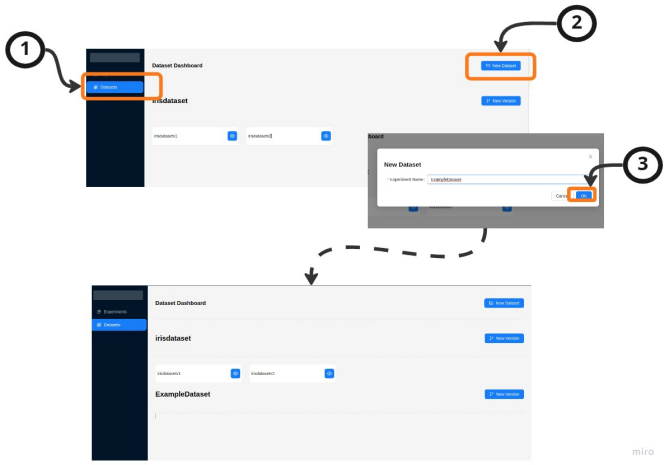


**Figure 1: Dataset creation process**

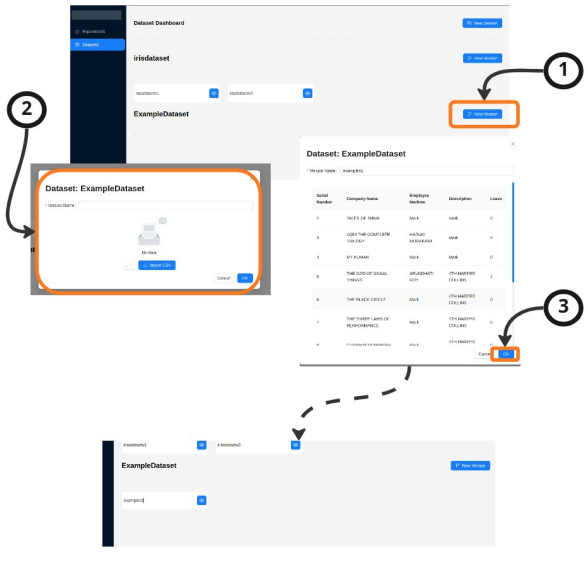


**Figure 2: Dataset version creation process**

## 3.1 Dataset Version Creation

Once a dataset is created (refer to section 3 for more information), a dataset version can be created, which represents a specific set of data.

To create a dataset version, follow the steps illustrated in Fig. 2:

1. On the dataset management screen, click on the *New version* button for the corresponding dataset. This will open a screen where the necessary information for a new dataset version is requested.

2. Fill in the required information, such as the version name, a CSV file containing the data, and the corresponding separator for the file. Once the file is selected, a preview of the data will be displayed.

3. After entering all the information, click on the confirm button. This imports the data from the CSV file into the system and exposes it to other components. This process may take a few seconds, upon completion, a new entry is generated in the manager, from which you can get a preview of the data.

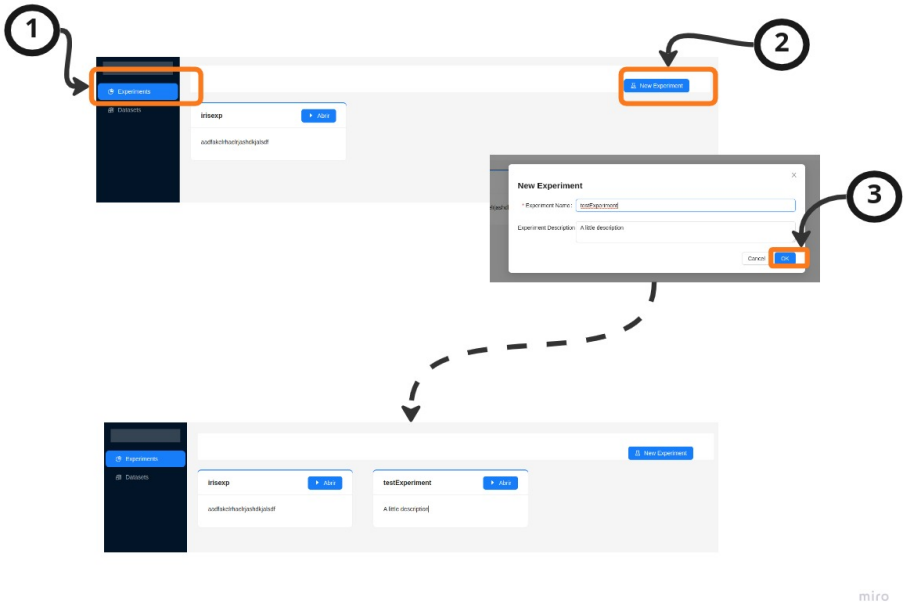


**Figure 3: Experiment creation process**

# 4 Experiment Management

The system allows the creation and management of different experiments. To do this, you need to create an *Experiment*. This experiment serves as a way to group processes into different versions, making it easy to maintain multiple versions of the same experiment.

To create an experiment, follow the steps illustrated in Fig. 3:

1. Select the experiment manager on the main screen. This will open a list of the different experiments that have been created in the system.
2. Click on the *New Experiment* button. This will open a window where you can enter the name and description of the new experiment.
3. Once you fill all the information, click on the confirm button. This will create a new entry in the manager.

## 4.1 Experiment Version Creation

Once an experiment is created (refer to section 4 for more information), you can create a version of that experiment, which represents a specific pipeline.

To create an experiment version, follow the steps illustrated in Fig. 4:

1. On the experiment management screen, click on the *Open* button for the corresponding experiment. This will open the editor from which you can create different versions of the experiment.
2. On the left side of the editor, there is a sidebar that displays a list of the different versions of the experiment. Click on the *New Version* button.
3. Enter the version name and click the confirm button. This will create a new entry in the version list.

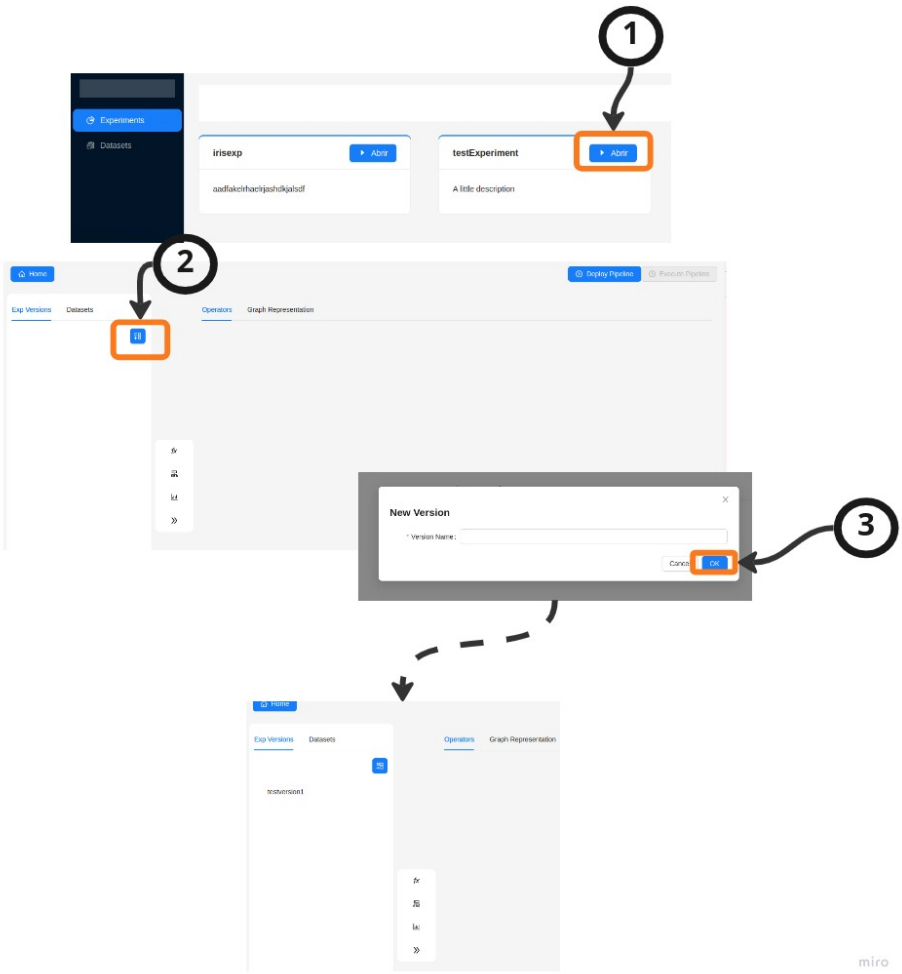


**Figure 4: Experiment version creation process**

# 5 Experiment Description and Execution

If you have followed all the steps described, you are ready to describe and execute a pipeline, as indicated in the following steps and in Fig. 5:

1. Select the version of the experiment you want to work with.
2. From the operator bar, you can select which operator to add to the pipeline. When selecting an operator, a panel will open where you can enter the parameters for the corresponding operator. The example shown illustrates the *Default Reader* operator, which has only one input and one output. **NOTE -** The *Default Reader* operator is the only operator that can read a previously imported dataset. These datasets can be check by changing the view in the sidebar.
3. Once you have entered the operator information, click the confirm button, which will generate a new entry in the operator list where you can see all the information previously entered.
4. If necessary, you can edit the parameters of an operator by clicking the *edit* button for that operator's entry.
5. Add operators until you have described each step of the process. **NOTE -** You can switch to a graph view from the list view, which provides a better visualization of the generated process.
6. Once you have all the desired operators, you can deploy the pipeline by clicking the *Deploy Pipeline* button, which starts the provisioning process. This may take a few minutes the first time it is executed.
7. When the provisioning process is complete, the *Start Pipeline* button will be enabled, which will execute the selected version of the experiment at that time.

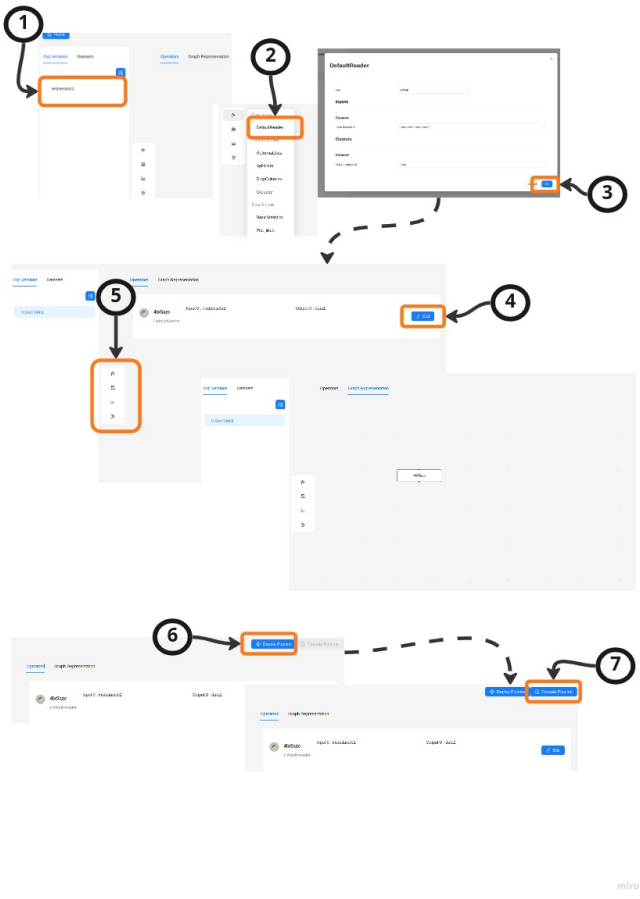


**Figure 5: Pipeline definition and execution process**