# OSCARS
Open Science Clusters' Action for Research & Society

# Funded Project

# Streamlining open data policies in Rucio

Presenter: Giovanni Guerrieri, CERN, ORCID: 0000-0002-3403-1177
On behalf of Project PI: HUGO GONZALEZ LABRADOR, CERN, ORCID: 0000-0001-5670-1437

Implemented by

CERN

Funded by the European Union

OSCARS

## What problem(s) did you plan to solve?

The project has now entered the  the 2nd phase (2026-2027).

**Open data is usually copied** from existing data storages to open systems, incurring extra costs because of duplicated data..

This project proposes a **paradigm shift: linking data at the source versus copying it.**

## What have you done to solve the problem?

In the 1st phase of the project, Open data support to tag datasets as "open data" was added to the Rucio upstream project, available to all the communities using Rucio since major version 39.

We developed a new functionality in the core of Rucio, to tag, search and navigate through open data sets. We have also added the option to link these data sets to DOIs.

One of the major challenges was to understand how different communities handle open data and what fields are required. The implementation we have done is very flexible and allows arbitrary metadata stored as json payloads alongside the data.

**What are the key results achieved to date and how have you made them available to the broader community?**

Rucio since major version 39 contains now the functionality to tag and serach for open data sets.

This functionality is reachable through the rucio-cli and the REST API.

As this feature is part of the upstream project, is automatically available when upgrading the Rucio software package.

**How will make your results sustainable over time - How will the scientific community/-ies further exploit them?**

The results of this project are not an independent project or not-merged branch, they are a core part of the upstream project.

We are currently working on a solution to download this data through Rucio, enforcing embargo policies for the data sets with access tokens.

OSCARS

**Who has been doing it?**

Hugo Gonzalez Labrador (CERN): ORCID: 0000-0001-5670-1437
Martin Barisist (CERN): ORCID: 0000-0003-0253-106X
Luis Obis Aparicio (CERN): ORCID: 0000-0002-7990-2060
Giovanni Guerrieri (CERN): ORCID: 0000-0002-3403-1177