



HAL
open science

Towards a semantic annotation of microbial information using the open-source software Dataverse

Marco Beccuti, Jean-Luc Legras, Gerard Wade, Jonathan Mineau-Cesari,
Simone Donetti, Sandro G Contaldo, Rossana Damiano, Paolo Romano

► To cite this version:

Marco Beccuti, Jean-Luc Legras, Gerard Wade, Jonathan Mineau-Cesari, Simone Donetti, et al..
Towards a semantic annotation of microbial information using the open-source software Dataverse.
ECCO XL 2022: New Horizons in Accessing Microbial Diversity, Sep 2022, Braunschweig, Germany.
hal-04132207

HAL Id: hal-04132207

<https://hal.science/hal-04132207>

Submitted on 18 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Towards a semantic annotation of microbial information using the open-source software Dataverse



Marco Beccuti¹, Antonio d'Acierno², Jean-Luc Legras³, Gerad Wade³, Jonathan Mineau-Cesari³, Simone Donetti¹, Sandro G. Contaldo¹, Rossana Damiano¹ and Paolo Romano⁴

¹Department of Computer Science, University of Turin, Italy

²National Research Council/Institute of Food Sciences, Avellino, Italy,

³SPO, University Montpellier, INRAE, Institut Agro Montpellier (CIRM - Levures), Montpellier, France,

⁴IRCCS Ospedale Policlinico San Martino, Genoa, Italy



OSPEDALE POLICLINICO SAN MARTINO

The Italian National Research Infrastructure MIRRI-IT and new challenges to face

The Joint Research Unit (JRU) MIRRI-IT is mainly aimed at the development of the Italian network of collections of microbial resources supporting the implementation of the Microbial Resource Research Infrastructure MIRRI-ERIC.

It was recognized by the Italian government as a **National Research Infrastructure** in 2021. Recently, the project "**Strengthening the MIRRI Italian Research Infrastructure for Sustainable Bioscience and Bioeconomy**" (**SUS-MIRRI.IT**) has been funded to support its empowerment.

The main objectives of this project are:

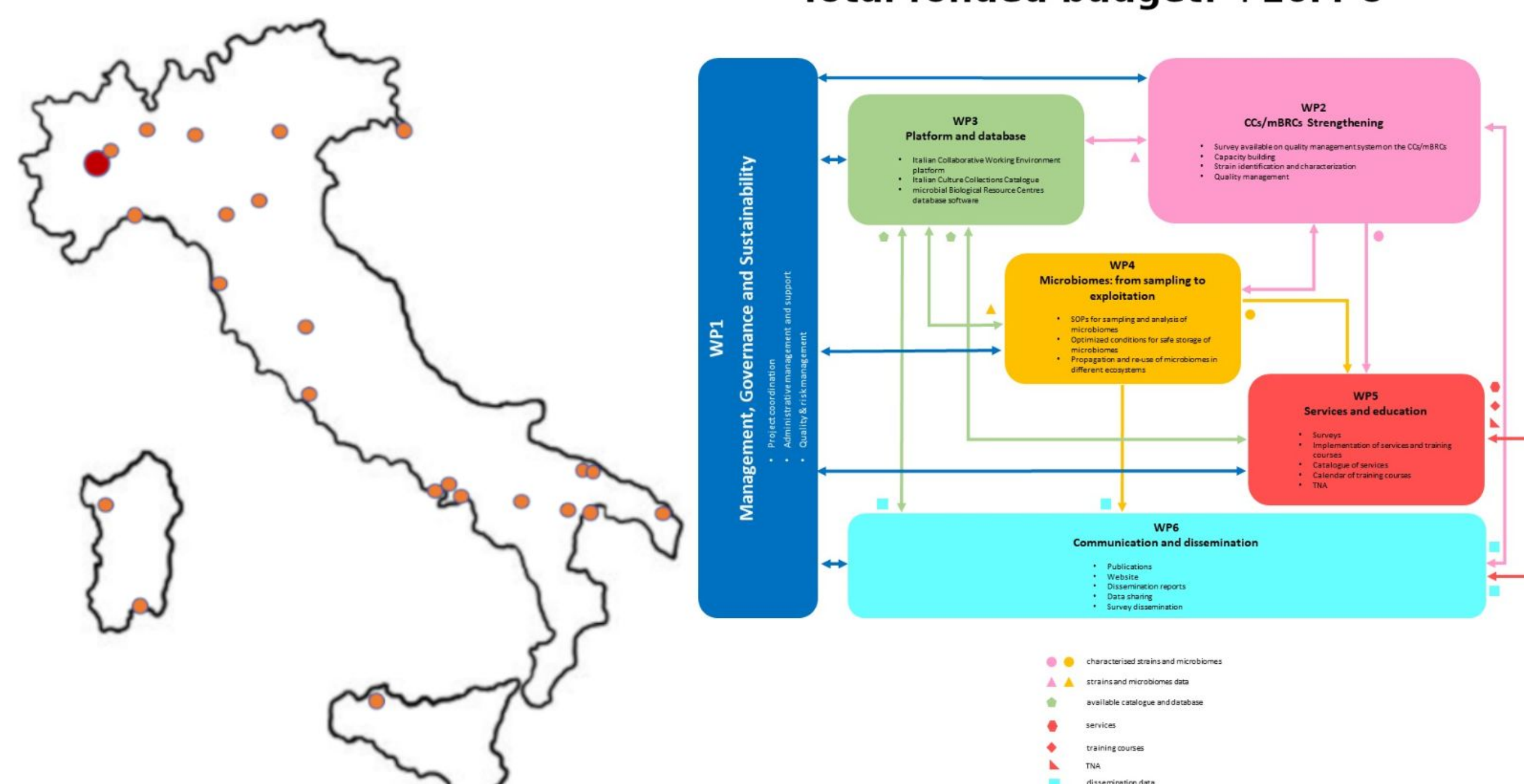
- to implement MIRRI-IT's organisation and set up its operative procedures and quality standards;
- to improve the Italian microbial Biological Resource Centers database quality and conceive MIRRI-IT's services based on the partners' expertise and genetic resources;
- to set up a single entry point platform to promote MIRRI-IT's resources in terms of data, services, cutting-edge technologies and expertise.

Members:

U001 - TUCC - UNITO
 U002 - CNR - ISA AVELLINO
 U003 - CNR - IPSP BARI
 U004 - CNR - ISPA BARI
 U005 - CNR - IBBA PISA
 U006 - CNR - ICB POZZUOLI
 U007 - CNR - IPSP TORINO
 U008 - CNR - IRSA VERBANIA
 U009 - ENEA - C.R. BRINDISI
 U010 - ENEA - C.R. CASACCIA
 U011 - ENEA - C.R. PORTICI
 U012 - ENEA - C.R. TRISAIA
 U013 - COSMI - OGS
 U014 - UNIBAS - SAFE
 U015 - MBDS-UNICA
 U016 - UNIGE - DISTAV
 U017 - UNIMIB - MICROMIB
 U018 - UNIMORE - UMCC
 U019 - UNINA - DIP. BIO.
 U020 - UNIPA - SAAF
 U021 - UNIPG - DSA3
 U022 - UNIPR - ALIFAR
 U023 - MBDS-UNISSCC
 U024 - UNIVR - DIP. BIOTECH.

SUS-MIRRI.IT project

Total funded budget: +16M €

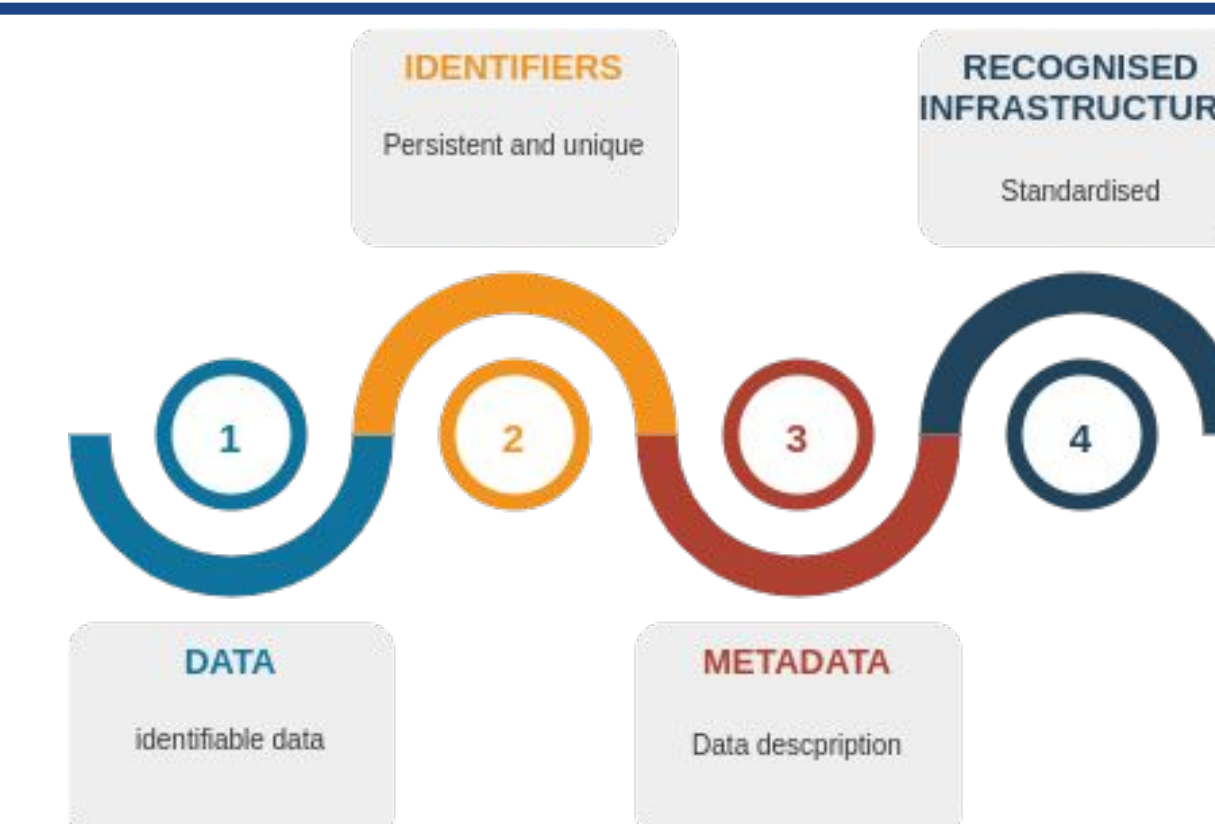


How to enhance the FAIRness of MIRRI-IT data

In the context of this project, the new needs for data **integration** and system **interoperability**, stressed by the **FAIR approach** to data sharing, have become evident and urgent.

We propose a pilot project to investigate how the semantic annotation of microbial information could be suitably stored using the **open-source repository software Dataverse Harvard**, one of the most common tools for improving data FAIRness.

It will be carried out thanks to a strong collaboration between **MIRRI-IT** and the **MIRRI task force for the FAIRification of MIRRI Information System (MIRRI-IS)**, potentially leading to new best practices for semantic annotation of microbial information.



Dataverse MIRRI infrastructure

This pilot project will be based on a **Dataverse repository instance** made up of five main components: the application, the database, the search engine, the file storage and the Digital Object Identifier (DOI) service (Fig. 1). To ensure a high level of performance, security and scalability for the whole project an open-source cloud Platform-as-a-Service (PaaS), namely **OpenDeep (OD)** will be specifically developed (Fig. 2). OD platform will be a template to deploy an **"HPC secure tenant"**. Technically, OD deploys a virtual compute cluster on encrypted storage volume pre-configured with software components including an authentication service (Keycloak), docker registry (Harbor), micro-service orchestrator (Kubernetes), a multi-tenancy manager (capsule), a firewall+VPN (Pfsense). Fig. 3 reports the common exploitation of the developed infrastructure.

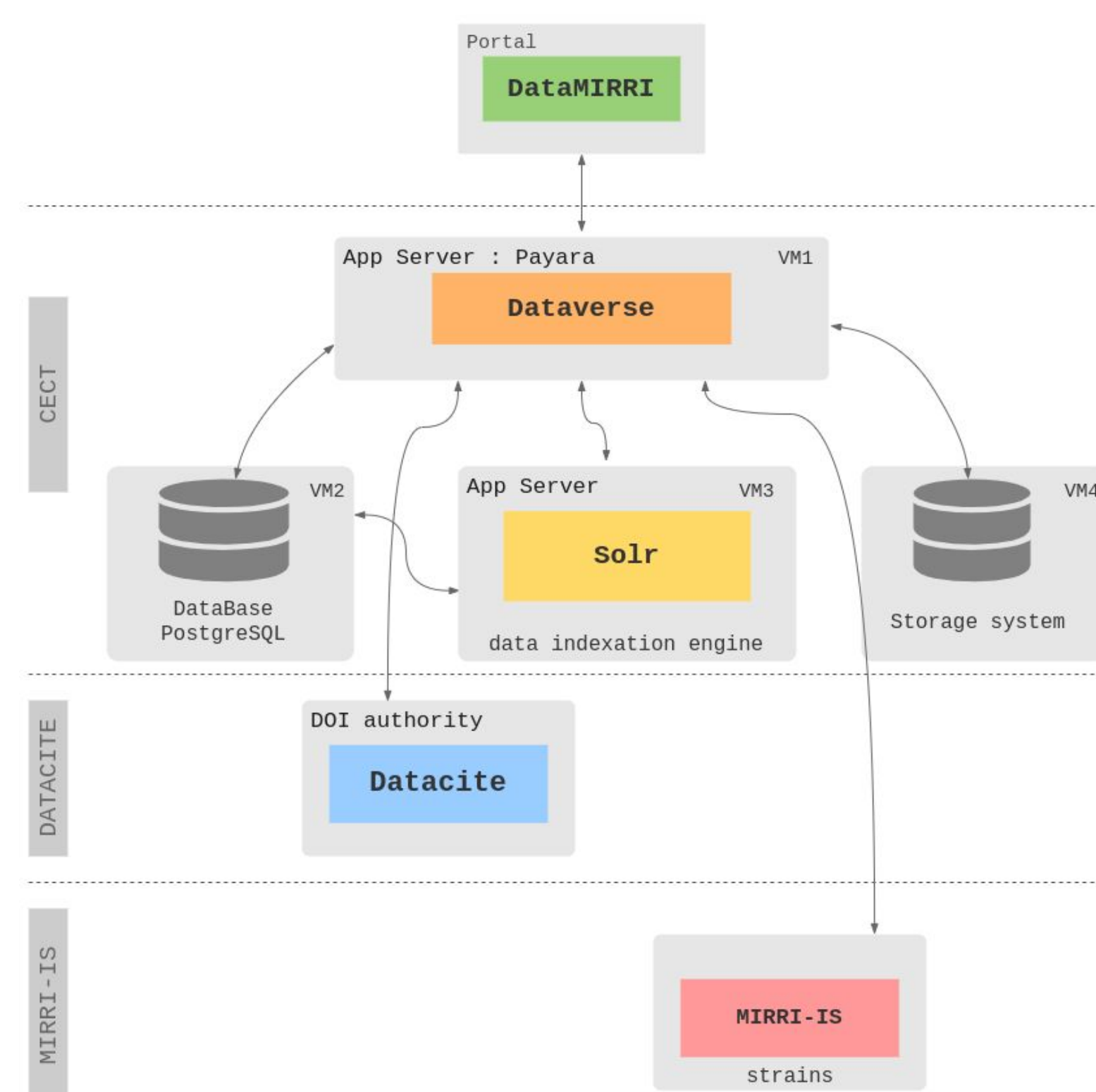


Fig. 1: **MIRRI-IT Dataverse** instance in which each component is a Docker Container deployed on **OD**.

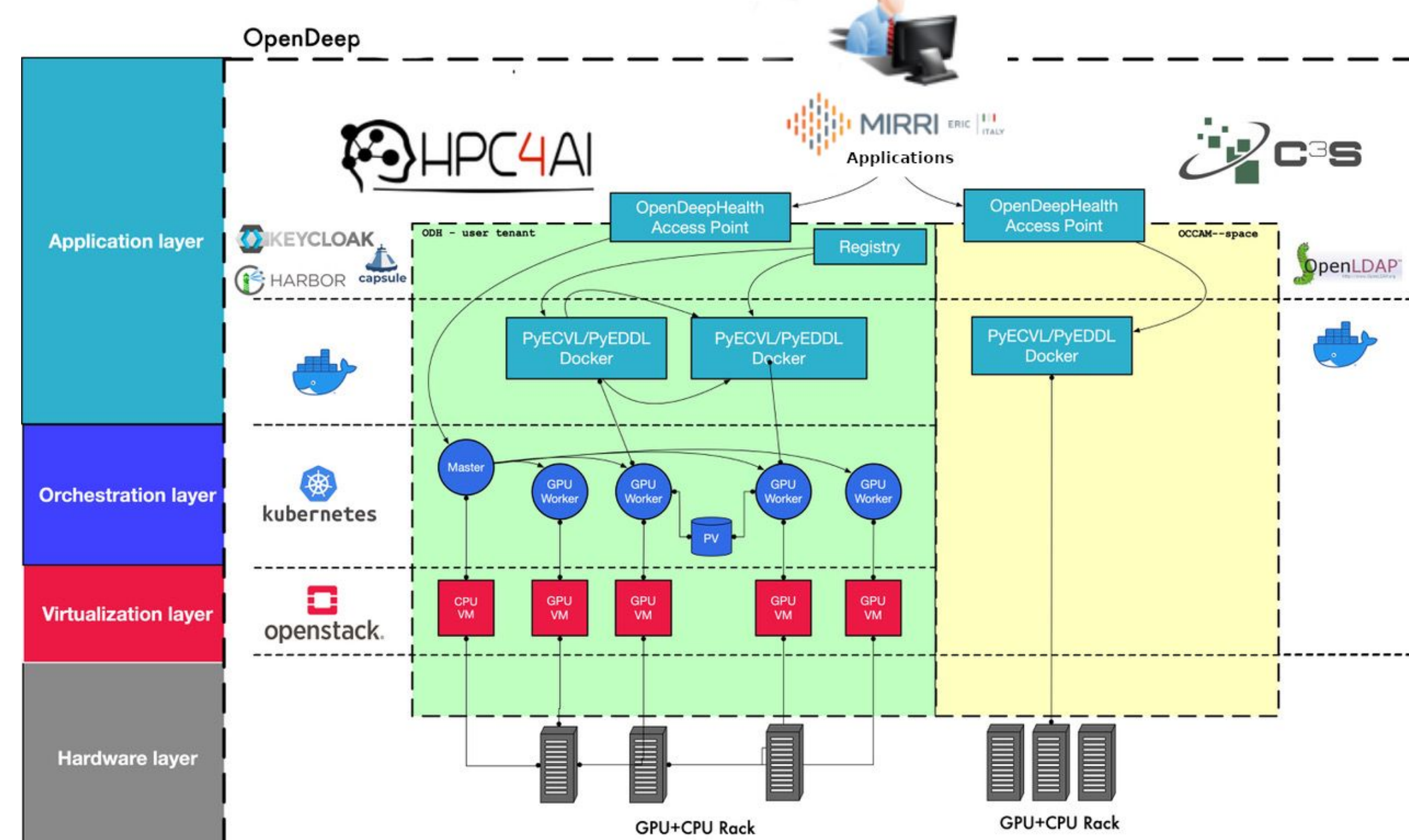


Fig. 2: Schema of the open-source cloud PaaS **OpenDeep** with respect to **SUS-MIRRI.IT** applications where MIRRI Application box is detailed in Fig.1

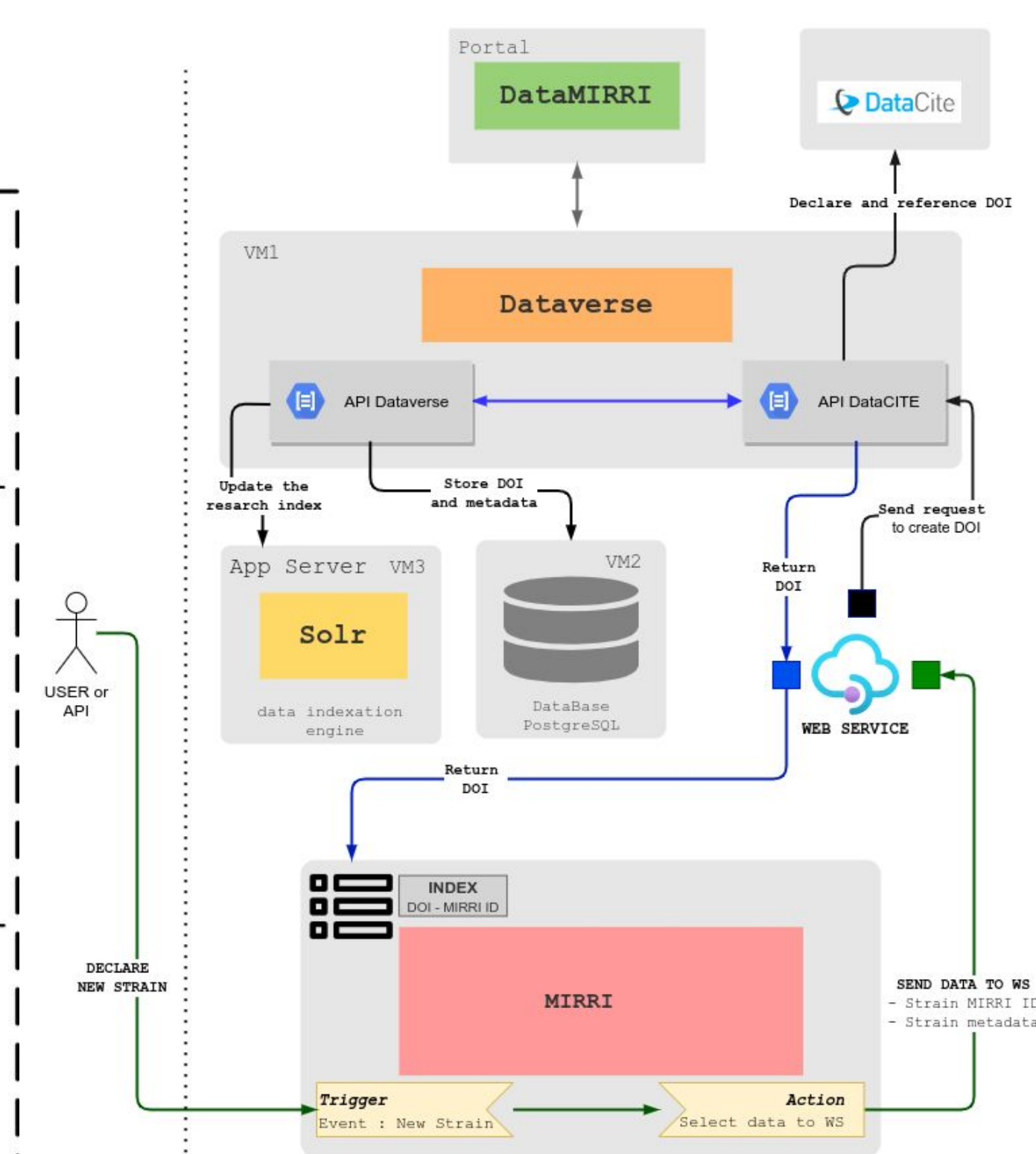


Fig. 3: Task flow diagram for **MIRRI-IT Dataverse**.

- References
1. MIRRI-IT: <http://www.mirri-it.it/>
 2. MIRRI-IS: <https://www.mirri.org/>
 3. Dataverse Harvard: <https://dataverse.org/about>

Contacts
marco.beccuti@unito.it
antonio.dacerno@isa.cnr.it
paolo.romano@hsanmartino.it