Grid y Computación de Altas Prestaciones GRyCAP



Infrastructure Manager

Automated Onboarding and Deployment for the Green Deal Data Space

Miguel Caballer[1], Estíbaliz Parcero[1], Ignacio Lamata Martinez[2], Mateusz Swiercz[3], Raúl Palma[3], Germán Moltó[1]

SAGE

[1] Universitat Politècnica de València
[2] EGI
[3] Poznan Supercomputing and Networking Center

UPV

Introduction

Data Space definition:

> *"Interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants"*

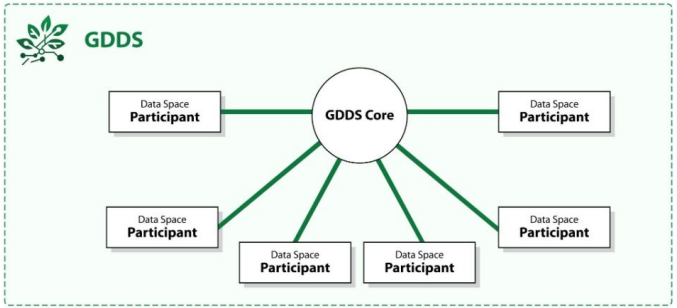
- The main constituents of a data space are:
  - **Data Space Participants**, which can act as data providers or data consumers.
  - Data Space Governance Authority, which is the entity in charge of the overall operation of the Data Space.
  - Service Providers, which are parties providing services to the general Data Space and their Participants.

Introduction

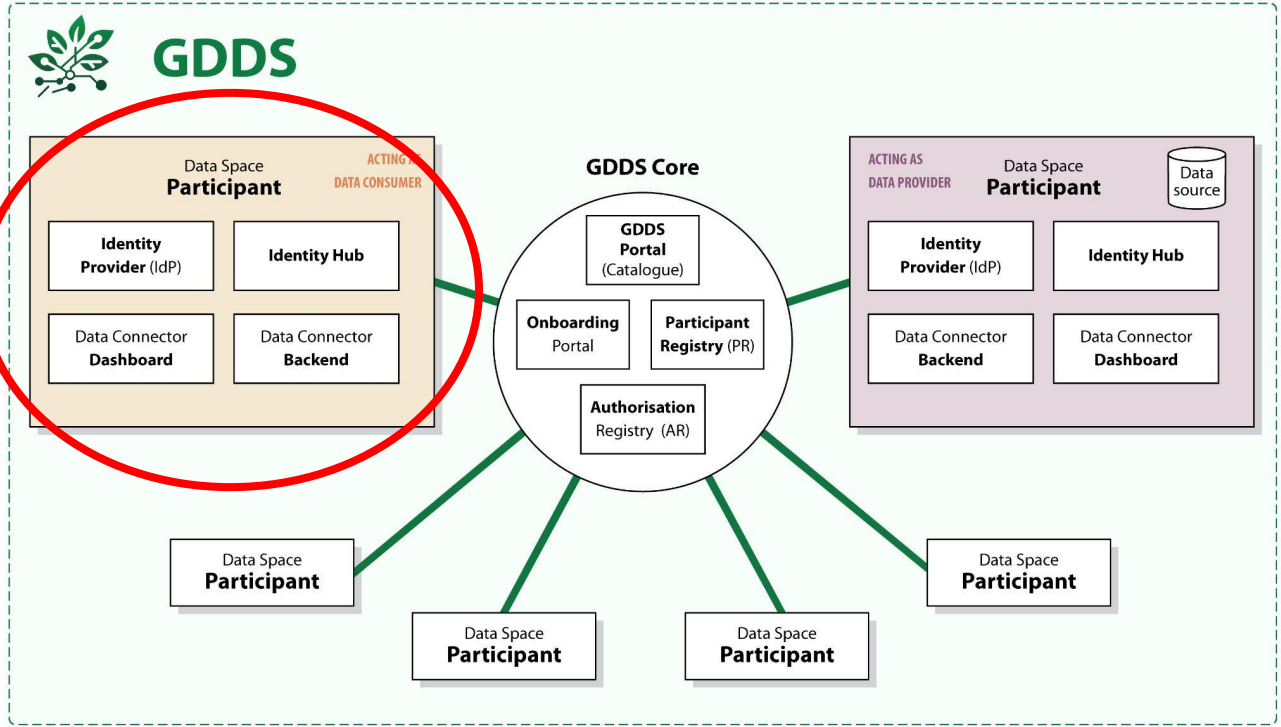
- The **Green Deal Data Space** (GDDS).
  - Aims to enable secure, reliable, and interoperable data sharing across environmental and sustainability domains.
  - Implemented by the **SAGE** project

SAGE
The Data Space for a Sustainable Green Europe

- Participants in the GDDS must **deploy** and manage a set of components that ensure interoperability within the data space
  - **IM** supports the installation of the SAGE Data Connector stack

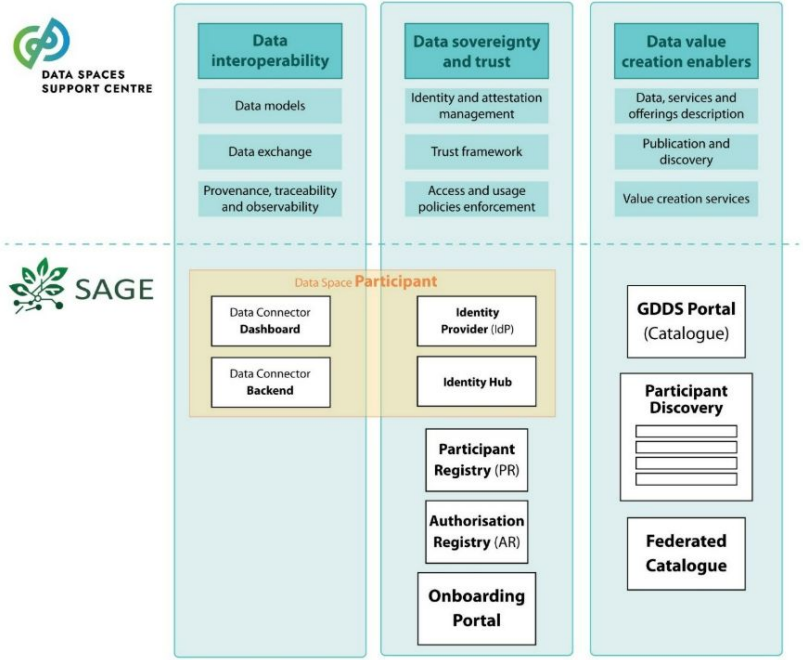


Introduction



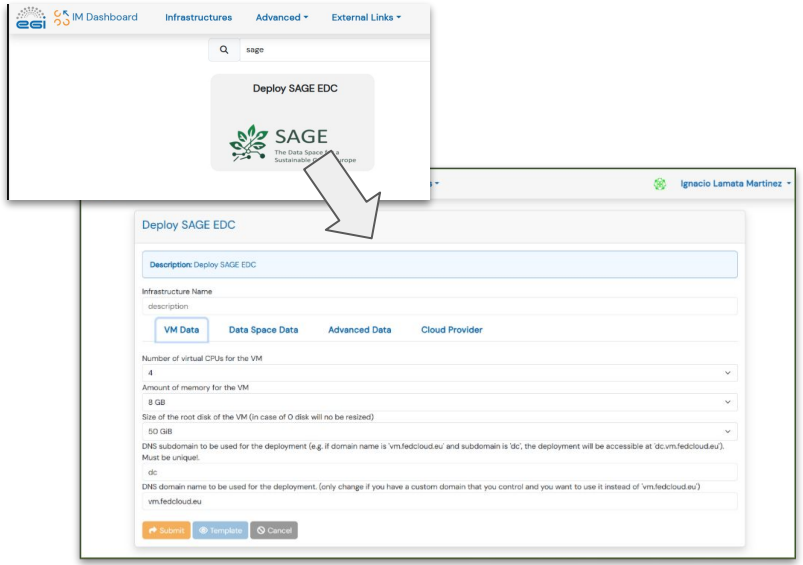
Data Connector components

- **Data Connector Dashboard:** GUI of the Data Connector.
  - Add data assets, data policies and contracts.
  - Access protected with the support of the Identity Provider.
  - It is implemented with Angular.
- **Data Connector backend**: operates the data transfers with the Data Connector backends of other Participants of the data space
  - It is implemented with EDC and custom extensions written in Java.
- **Identity Hub**: stores and manages digital identities.
  - Also part of EDC supported by PostgreSQL and Hashicorp Vault.
- **Identity Provider**: to authenticates users and systems.
  - Keycloak distribution based on RCIAM, which has been adapted for iSHARE compatibility.



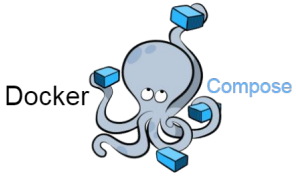
SAGE EDC Deployment

- Deployment of the Cloud resources needed for the Participant node
- Management of Participant credentials in a secure manner
- Automatic creation of DNS entries
  - EGI DyDNS service
- Installation of the Data Space Participant stack of components to operate in the GDDS **MVP**

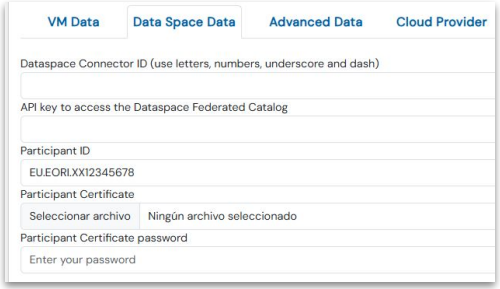


Single Tenant Deployment

- Single VM with docker compose
  - PSNC created base docker compose file, scripts and instructions (link)
- Services:
  - Traefik
  - Connector
    - Connector DB (PostgesSQL)
  - Consumer Back-end
  - Dashboard
  - Keycloak: RCIAM version
  - Identity-Hub
    - Identity-Hub DB (PostgesSQL)
    - Vault
  - Storage (RustFS)



TOSCA template



- Two group of input values:
  - Mandatory:
    - Minimum set of values.
  - Advanced
    - Can use default values in most of the cases.
- The PKCS12 certificate must be managed securely:
  - IM-Dashboard enables the usage of special type of input “secret”
    - It is stored it in an internal vault
    - Sets an unique use url in the TOSCA template
      - The certificate is not set as input in the TOSCA template.
    - The Ansible role downloads the certificate
      - It is automatically deleted from the Vault.



https://github.com/grycap/tosca/blob/main/templates/vc_sage.yaml

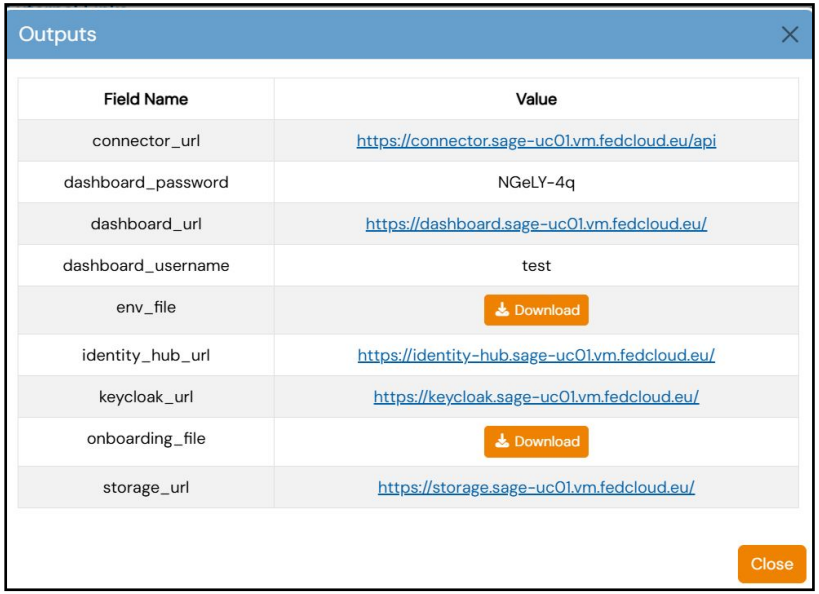
Ansible Role



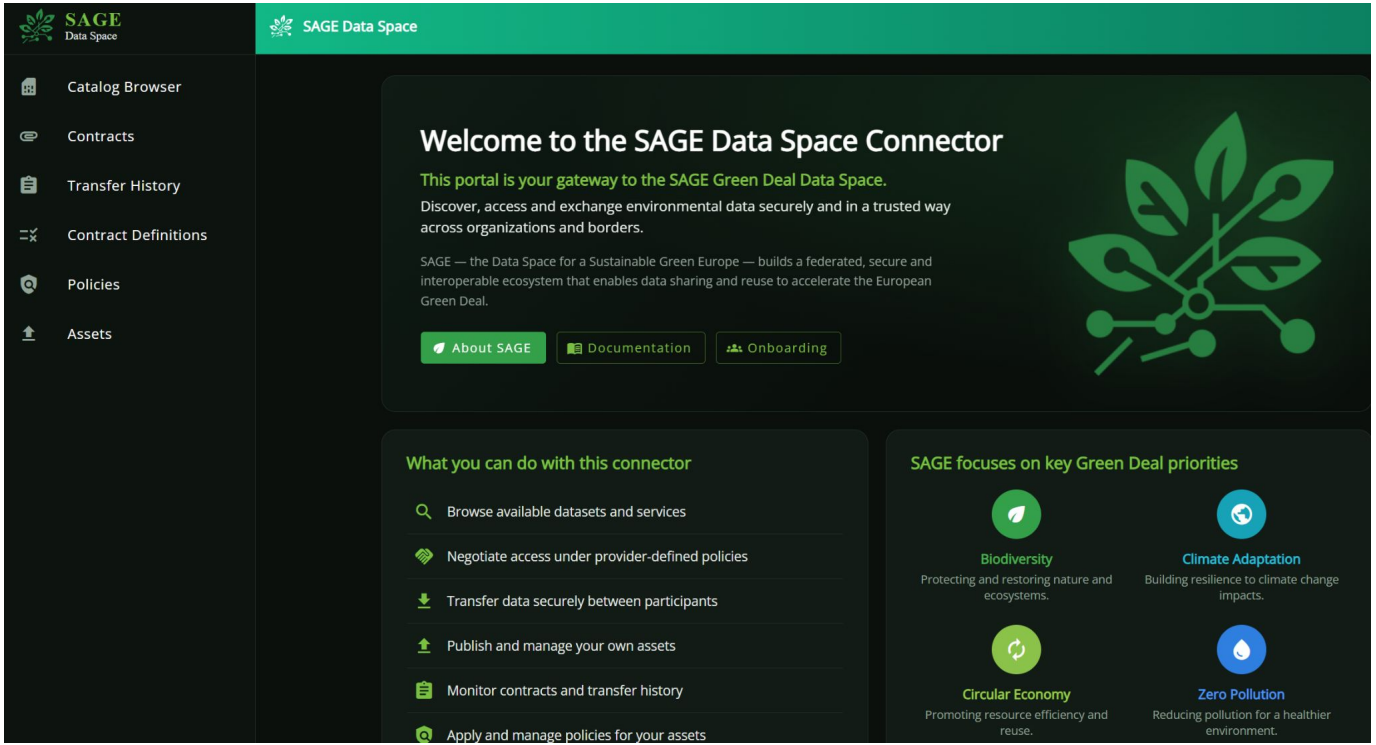
- Ansible Role
  - https://github.com/grycap/ansible-role-sage_edc
- Steps:
  - Generate passwords
  - Get PKCS certificate alias
  - Replace “vanilla” Keycloak with RCIAM one.
  - Configure SAGE realm in Keycloak
  - Init Vault instance.
  - Launch services.
  - Split certificate and upload it to Vault.

Deployment

- Multiple clouds used in the project:
  - EGI Cloud Compute
  - Azure
  - UpCloud
- TOSCA Outputs:
  - Endpoints to all deployed components
  - Env file with all administrative passwords.
  - Information for the onboarding of the participant.
    - Register the data connector to the data space



Deployment


SAGE
Data Space
SAGE Data Space
Catalog Browser
Contracts
Transfer History
Contract Definitions
Policies
Assets
Welcome to the SAGE Data Space Connector
This portal is your gateway to the SAGE Green Deal Data Space.
Discover, access and exchange environmental data securely and in a trusted way across organizations and borders.
SAGE — the Data Space for a Sustainable Green Europe — builds a federated, secure and interoperable ecosystem that enables data sharing and reuse to accelerate the European Green Deal.
About SAGE
Documentation
Onboarding
What you can do with this connector
Browse available datasets and services
Negotiate access under provider-defined policies
Transfer data securely between participants
Publish and manage your own assets
Monitor contracts and transfer history
Apply and manage policies for your assets
SAGE focuses on key Green Deal priorities
Biodiversity
Protecting and restoring nature and ecosystems.
Climate Adaptation
Building resilience to climate change impacts.
Circular Economy
Promoting resource efficiency and reuse.
Zero Pollution
Reducing pollution for a healthier environment.

More Information



IM Info Web:

https://grycap.upv.es/im/

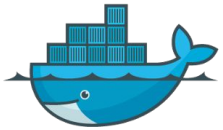
SAGE project Web:

https://www.greendealdata.eu/

Video demos in YouTube:

https://youtube.com/playlist?list=PLgPH186Qwh_37AMhEruhVKZSfoYpHkrUp







Contact

Miguel Caballer
micafer1@upv.es



Ignacio Lamata Martinez
ignacio.lamata@egi.eu

Co-funded by the European Union

This project has received co-funding from the Digital Europe Programme (DIGITAL) under Grant Agreement No. 101195471.