

Federated Assets: Datasets and Services

Andrea Manzi(EGI Foundation)

The ecosystems of HPC and Big Data/AI: what can they learn from each other?
15/03/2022



The EUHUBS4DATA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951771



This project is part of **BDV PPP**

Outline

- The Federated catalogue
 - First version, improvements and details
- Expanding the offer
 - new DIHs and community building activities
 - Open calls and experiments
- Data Privacy and protection

Federated Offer Instantiated at Regional level

DATASETS

Access to datasets provided by EUHubs4Data members and open data sources, ready to be used.

Showing 1 to 20 of 160 entries

FILTERS	NAME	DOMAIN	LICENSE TYPE	COVERAGE	DESCRIPTION
DOMAIN	2025 AI/ML IMPROVEMENT		Creative Commons	United Kingdom	This dataset consists of the research results of the...

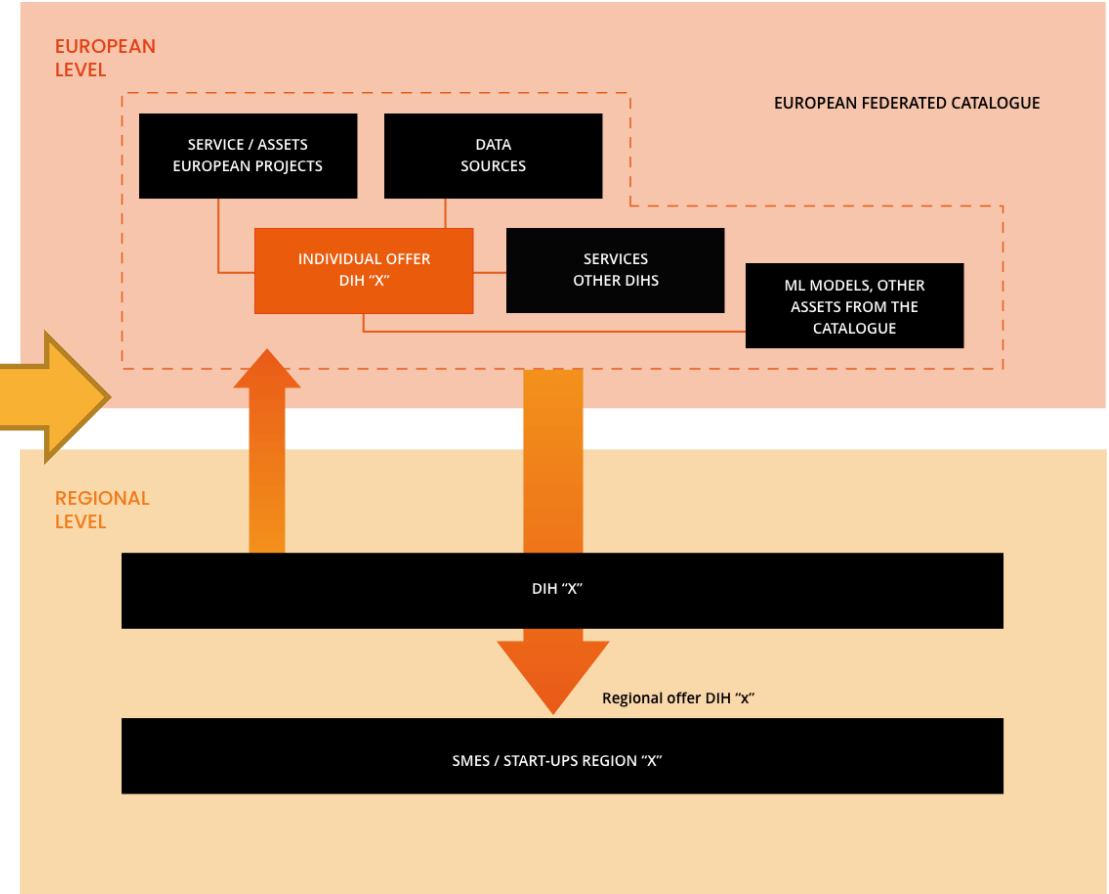
SERVICES

Discover all the services the EUHubs4Data members can offer to your organization.

ALL SERVICES OPEN CALL SERVICES

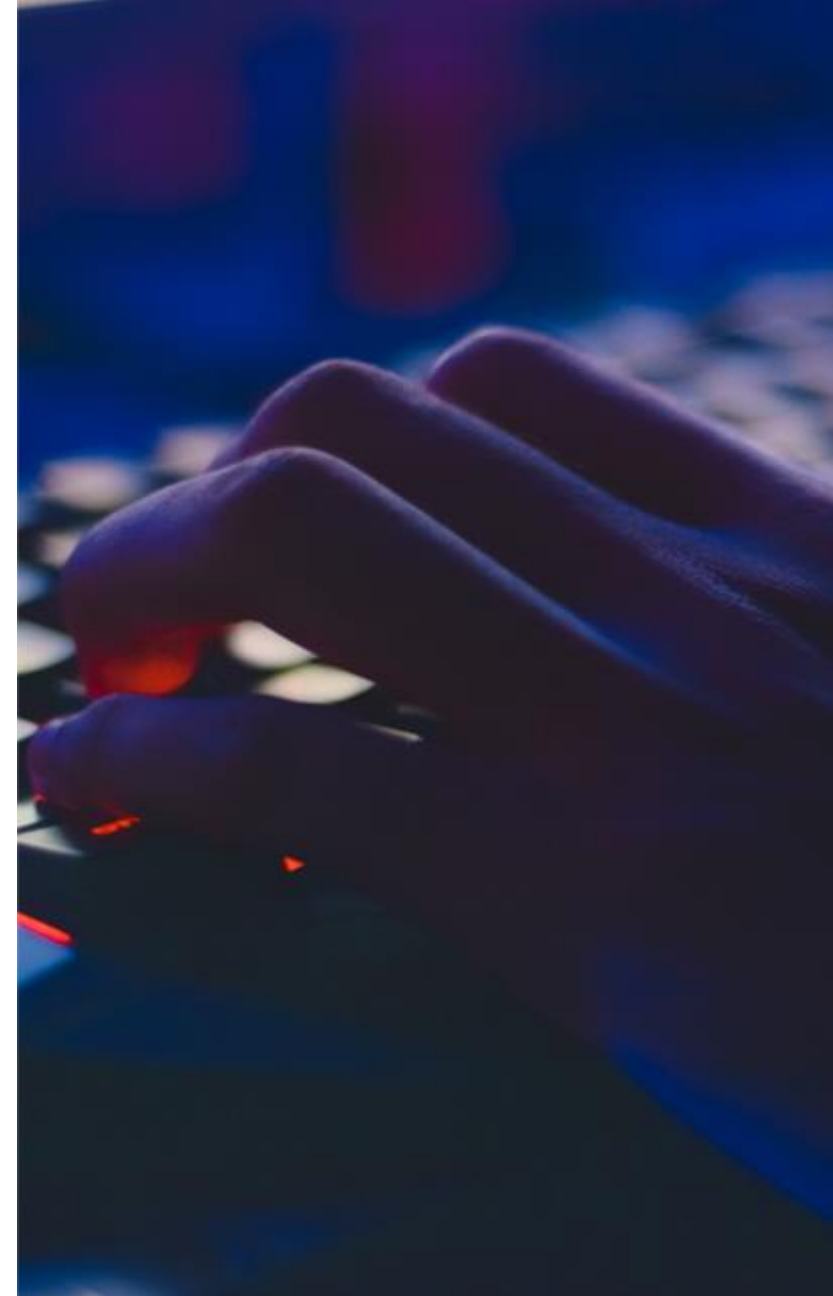
Showing 1 to 20 of 161 entries

FILTERS	SERVICE NAME	CATEGORY	DESCRIPTION	KEYWORDS	MEMBER
TYPE OF SERVICE	Access and support to Big Data-AI stack environment	Big Data Platforms (PaaS)	Access and support to a ready-to-use Big Data and AI experimental...	analytics, artificial intelligence, HDFS, notebooks, Python libra...	The Data Cycle Hub
Business Development & Ecosystem Building					
Data, AI & Technology	Access to Big Data	Big Data			



Building the catalogue

- 2 WPs in the project dealing with catalogue: Services (WP4) and Datasets (WP5)
- The first version of the catalogue has been released after 3 months of the project
 - Ready for the first project Open Call in Dec 2020
- Initial definition of the catalogue entries for both services and datasets
 - Preparation of service and datasets templates, categorization
 - Identification of the offer from the DIHs networks (services and datasets) and high value datasets coming from EU repositories



Second version of the Catalogue

- Based on the Feedback from First open call applicants, in 2021 we started improving the catalogue in parallel to building the frameworks for catalogue synchronization, data sharing and interoperability (See talk from Guillermo Amat)
 - Datasets
 - License type detailed information; categorization according to data privacy and protection; European Data Portal controlled vocabulary for domains; new disclaimer for datasets provided by third parties
 - Services
 - Improved categorization; service granularity and descriptions; added Success stories and detailed DIH security info

Catalogue details

- 160 Services

TYPE OF SERVICE

- ☐ Business Development & Ecosystem Building
- ☐ Data, AI & Technology
- ☐ Infrastructure Access
- ☐ Technology Consultancy and Training



- ☒ Data, AI & Technology
- ☐ Dataset Access
- ☐ Product Development
- ☐ Proof-of-Concept (PoC) Research & Development
- ☐ Prototype Research & Development
- ☒ Infrastructure Access
- ☐ Big Data Platforms (PaaS)
- ☐ Computing and Storage Infrastructure (IaaS)
- ☐ Data Management

Private/proprietary datasets included under “Data Accesses” entries and selected for Open Calls to be consumed by Data Driven services.

- 150 Open Data sources (90 single datasets and 60 repositories)

New DIHs and community building

- The project DIHs increased by 9 in 2021 and 9 new DIHs will also join in 2022
 - DIHs from 2021 published already services and private/proprietary datasets
 - Both new DIHs from 2021 and 2022 will also publish open datasets to enrich the catalogue
- Dedicated WP in the project (WP2) for community building, with the aim of extending the catalogue
 - recent WS with H2020 ICT-13 Data Platforms projects

Open Calls and experiments

- The catalogue has been used by SMEs to select the services/datasets for the 1st and 2nd project Open calls
 - 1 more open call is planned during the project
 - Hard requirement to use services from at least 2 DIHs
- Selected SMEs from the second open call will either use datasets from the catalogue or brought the datasets to be published
- Experiments produce datasets that might be integrated in the catalogue
 - 1st open call experiments produced ~20 datasets available for publications and sharing

Data privacy and protection

- Provide general guidelines for Data Management for use in DIHs and project participants
 - Categorization of datasets from the privacy and personal data protection perspective
 - Definition of the roles of the project participants under the GDPR
 - Protection of personal data under GDPR
 - Guidelines for the processing of the personal data

Guidelines and checklist



- First version of the Guidelines released in Feb 21, and updated in Dec 21
- First Checklist implemented, intend to provide the members of the federation with a short recap on the privacy-related aspects that should be taken into account
 - Particular focus on Datasets providers when including new datasets in the catalogue
 - Under extension to cover recommendations for experiments participants

Thank you!



www.euhubs4data.eu



The EUHUBS4DATA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951771



BIG DATA VALUE
PUBLIC-PRIVATE PARTNERSHIP

This project is part of **BDV PPP**