IT4LIA AI FACTORY

GenOA Week 2026

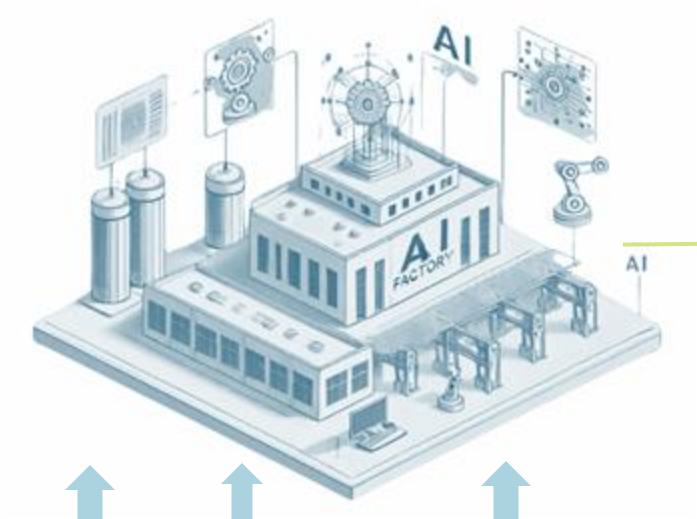
Open Science, dati e Intelligenza Artificiale

Un dataset FAIR è davvero pronto per l'AI? Governance, stewardship e AI-Readiness nell'esperienza di IT4LIA AI Factory

Donatella Sforzini

d.sforzini@cineca.it

Che cos’è un AI Factory



AI for SCIENCE

AI for INNOVATION

AI for PUBLIC GOOD

DATA

STRATEGIC DOMAINS

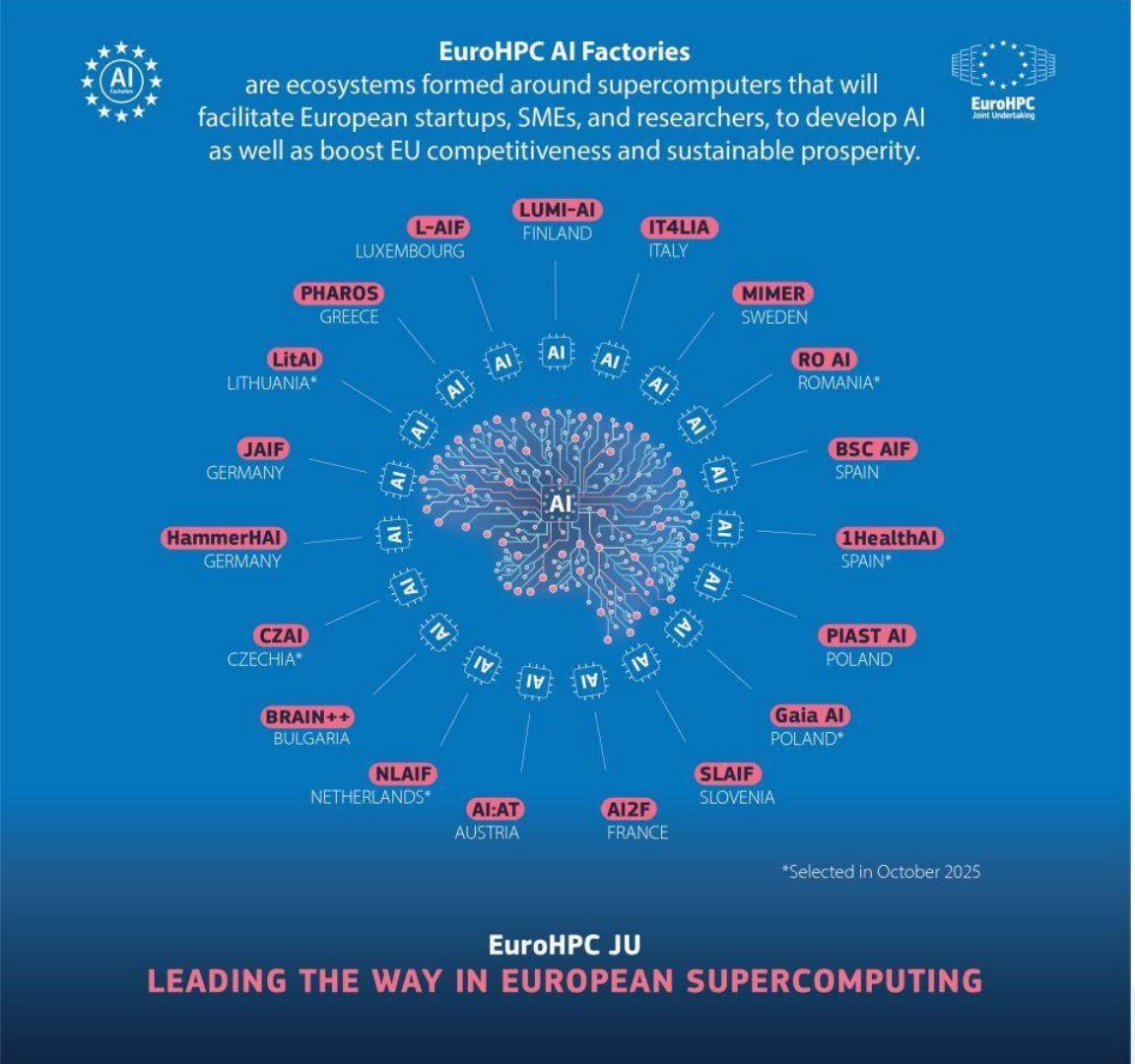
AI OPTIMIZED SUPERCOMPUTER

EXPERT SUPPORT

AI-FRIENDLY HPC

AI ECOSYSTEM

AI SKILLS



Rete europea

IT4LIA fa parte della rete di "AI Factories" co-finanziate dalla Commissione Europea.

Supercomputer EuroHPC

Le AI Factories sfruttano la potenza dei supercomputer EuroHPC.

Innovazione settoriale

Promuovono sviluppo in sanità, manifattura, clima, finanza e spazio.

IT4LIA: un'opportunità per chi innova

Promoters

Hosting Entity

CINECA

Other Participating Member States

Partners

Chi può accedere?

IT4LIA si rivolge a **imprese**, **startup, enti pubblici e istituzioni accademiche** che stanno sviluppando soluzioni basate sull'Intelligenza Artificiale.

Cosa offre?

Offre accesso a **potenza di calcolo ad alte prestazioni**, **competenze tecniche** specializzate e **supporto** lungo il percorso di sviluppo, senza bisogno di investire in infrastrutture proprie.

A quale costo?

È **gratuito** per **startup**, **PMI**, **pubbliche amministrazioni** e **istituzioni accademiche** idonee, oltre a offerte personalizzate per altri attori dell'innovazione.

E per chi non fosse ancora pronto?

Non hai ancora un progetto AI definito? IT4LIA offre anche **percorsi formativi e servizi di consulenza introduttiva**, tra cui la valutazione del potenziale AI della tua organizzazione.

Cosa offre IT4LIA AI Factory

Risorse, **competenze** e **strumenti** per accelerare l'adozione dell'Intelligenza Artificiale

Ore calcolo HPC

Supercomputer EuroHPC LEONARDO, LISA, cloud GAIA, MEGARIDE, IT4LIA.
Migliaia di GPU disponibili.

Supporto di esperti

Consulenza tecnica e strategica end-to-end da parte di esperti AI e HPC

Soluzioni ad accesso libero

Repository di servizi, dataset e modelli AI aperti, sviluppati nei domini verticali di IT4LIA

Repository aperto →

Formazione

Programma completo con formazione dedicata, corsi a catalogo hackathon e tirocini.

Calendario formativo →

Tirocini →

Matchmaking

Connessione tra imprese, ricercatori, startup e istituzioni per partnership AI.

SUK →

Link UP→

Accesso tramite bando
Accesso libero

Portfolio dei Servizi IT4LIA AI Factory

Tutto ciò che serve per passare dall'idea all'implementazione



01 Servizi di Consulenza

AI Assessment · Definizione del Progetto AI, dall'idea alla pianificazione tecnica.



02 Servizi relativi ai dati

Discovery · Creation · Management · Preparation



03 Servizi Orizzontali

AI Development · Setup · Test · Trust



04 Servizi Verticali

Agrifood · Earth · Manufacturing · Cybersecurity



05 Sviluppo delle Competenze

Formazione HPC & AI

Alla base c'è l'infrastruttura HPC di IT4LIA: la potenza di calcolo che rende possibile sviluppare, testare e scalare soluzioni di IA.

Portfolio dei Servizi IT4LIA AI Factory

Servizi di Consulenza

- AI Assessment
- Definizione del Progetto AI

Servizi relativi ai Dati

Data Discovery
- Esplorazione e selezione dati

Data Creation
- Creazione di dati sintetici

Data Management
- High Performance Data Lake (HPDL) per analisi dei dati su HPC
- Simulation Data Lake (SDL) per la gestione dei dati di simulazione
- Trasferimento massivo di dati

Data Preparation
- Preprocessing di dati su HPC
- Anonimizzazione dati strutturati

Servizi Verticali

Supporto e Advisory su soluzioni di dominio
- Individuazione e selezione di dataset AI-ready di dominio
- Preparazione e standardizzazione dei dati per modelli AI
- Selezione e caratterizzazione di modelli AI

Supporto allo sviluppo di prototipi e applicazioni di dominio
- Dominio Agrifood
- Dominio Earth
- Dominio Manufacturing
- Dominio Cybersecurity

Servizi Orizzontali

AI Development
- Setup di workflow RAG (pipeline custom)
- Sistemi Agentic AI
- Addestramento federato
- Pre-training e fine-tuning su HPC di modelli AI large scale
- Sviluppo di modelli AI custom

AI Setup
- Configurazione rapida di framework e workflow su HPC
- MLOPs su HPC

AI Test
- Ottimizzazione, validazione, scalabilità dell'AI

AI Trust
- Tecniche di controllo e incertezza dell'AI
- Compliance legale dell'AI
- Valutazione del rischio dell'AI
- Valutazione tecnica dell'affidabilità e Sandbox

Sviluppo delle Competenze

- Formazione per l'accesso alla potenza di calcolo

I Data Services nel ciclo di vita dei dati



Data Discovery

Selezione di dataset pertinenti e valutazione di **qualità**, idoneità, **condizioni di accesso** e vincoli di licenza.



Data Preparation

Trasformazione dei dati in formati **AI-ready**.

I servizi includono pulizia, anonimizzazione e standardizzazione dei dati.



Data Creation

Generazione di dati sintetici realistici a supporto di attività di analisi, addestramento e validazione quando i dati disponibili sono limitati.



Data Management

Workflow efficienti per processi AI.

Ricerca, **archiviazione sicura**, **trasferimento** di **grandi volumi** di dati e integrazione fluida tra **diverse fonti e sistemi** di storage.

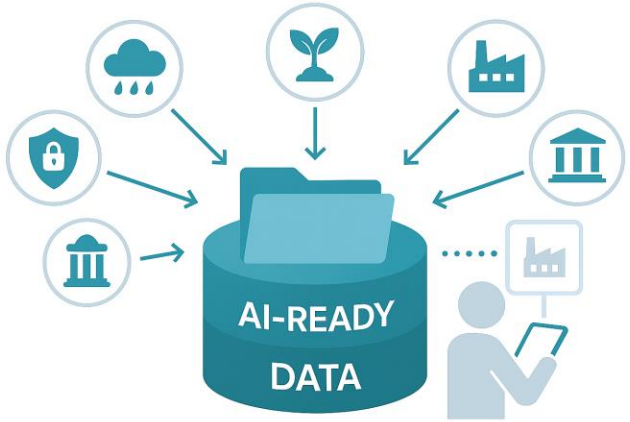
Il National Data Registry

Il National Data Registry è una **piattaforma federata** sviluppata per **catalogare** e **governare** assets, rendendoli **navigabili** grazie all'integrazione di **layer** di **metadati**.

- Supporta la **scoperta** e la **ricerca** delle risorse disponibili;
- Migliora il rispetto di **criteri FAIR** e di **AI-readiness**
- **Condivide** i dati in maniera **sicura e controllata**
- Ingloba una solida **data stewardship**

valorizzazione del **dato**

sviluppo di **applicazioni AI affidabili**



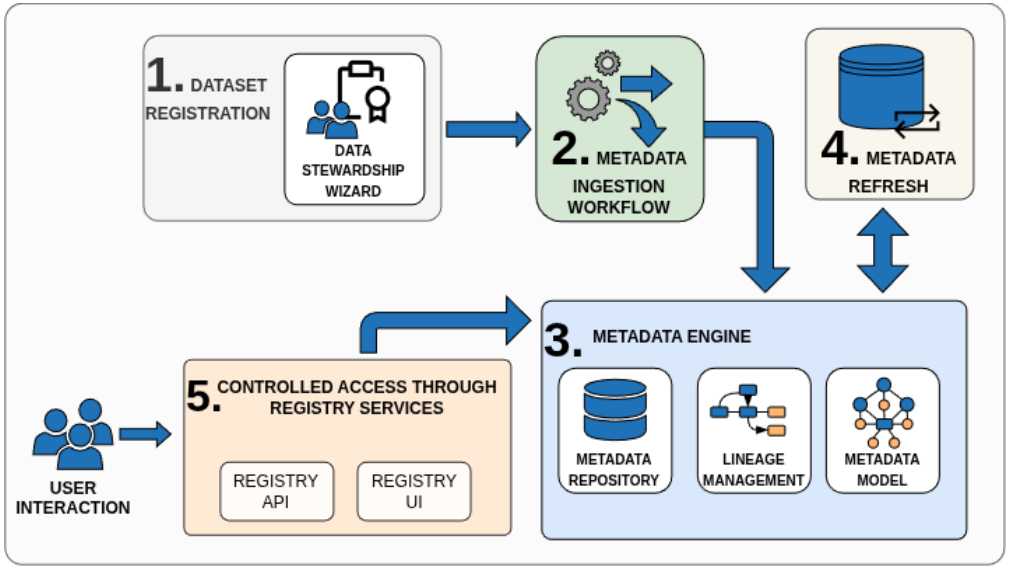
L’architettura

Data Stewardship Wizard (DSW) per la raccolta e validazione dei metadati

Metadata Ingestion Workflows per la sincronizzazione automatica

Metadata Engine (OpenMetadata) come catalogo centrale

Registry API e Registry UI come punto di accesso per utenti e applicazioni.
Integrazione con sistemi IAM per autenticazione e autorizzazione



Funzionamento e Governance



1. RACCOLTA DATI

Il data provider compila un **questionario DSW** per la classificazione e la valutazione della qualità dei dati.



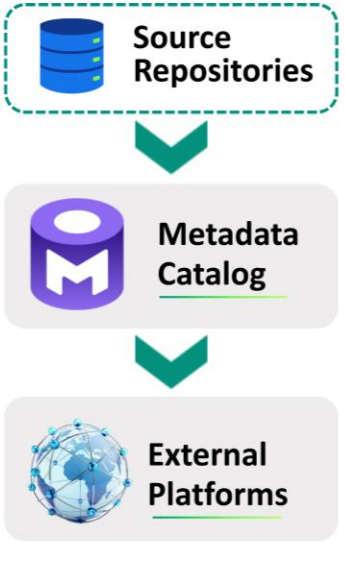
2. DATA STEWARSHIP

Il **Chief Librarian** e i **Data Steward** supportano e supervisionano il processo di onboarding, garantendo la conformità ai requisiti di qualità IT4LIA.



3. APPROVAZIONE

Quando i dataset soddisfano i requisiti di **maturità, accessibilità e sicurezza**, vengono **approvati** e possono essere inseriti nel catalogo.



La Data Stewardship…la supervisione umana sull'AI

Cosa ha implementato IT4LIA

- **Chief Librarian**
- **Rete dei Data Steward**
- **Procedure di Onboarding**
- **Procedure di Governance prima che i dati entrino nell'ecosistema**

Chief Librarian e Data Steward di IT4LIA…

- Esaminano i datasets
- Valutano gli aspetti di governance
- Verificano la documentazione
- Valutano l'AI readiness
- Monitorano la conformità alle normative e alle policy
- Formano i Data Steward dei partner…

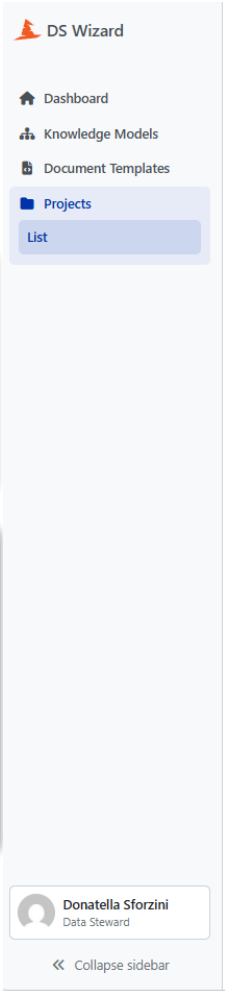
La Data Stewardship concretizza la supervisione umana sul patrimonio informativo, garantendo qualità, conformità e adeguatezza dei dati prima dell'impiego nei modelli di Intelligenza Artificiale.

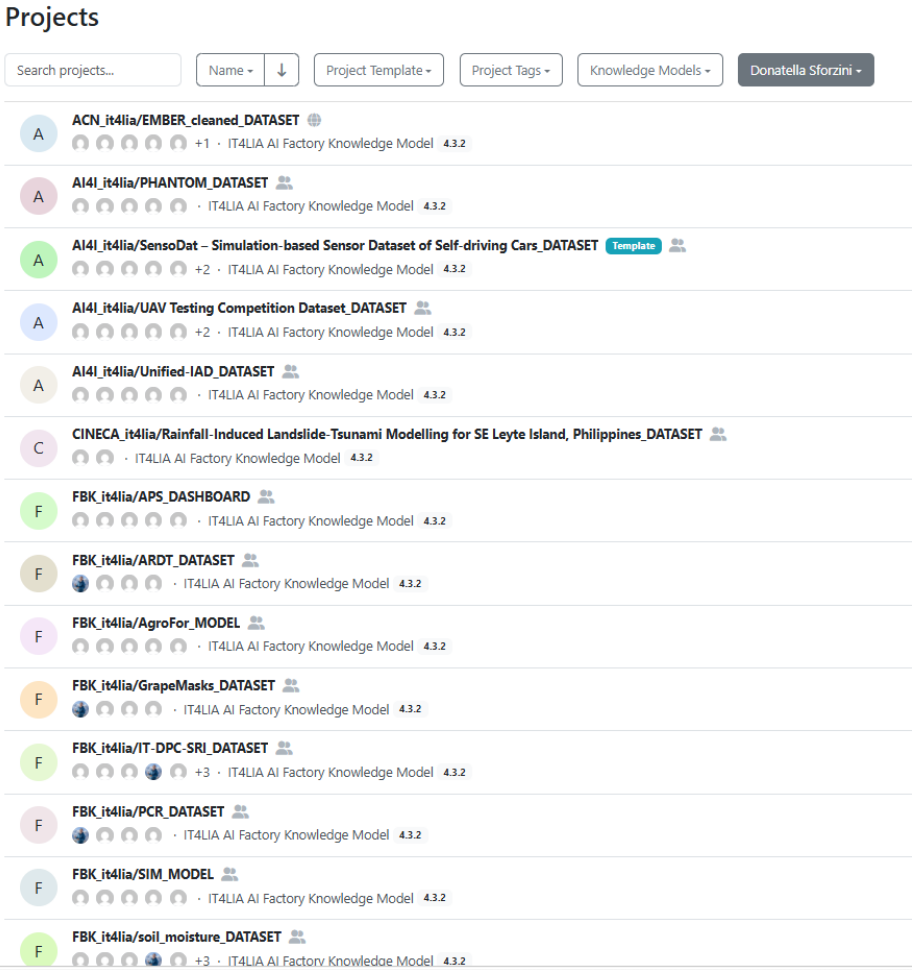
DS Wizard

Data Stewardship Wizard è uno strumento che rende più efficiente il processo di onboarding di dati.

Ai Data Steward e Chief Librarian di IT4LIA si affiancano uno o più Data Steward **per ogni partner**, formati ad hoc per l'utilizzo coerente della piattaforma.

DSW
DATA STEWARDSHIP WIZARD



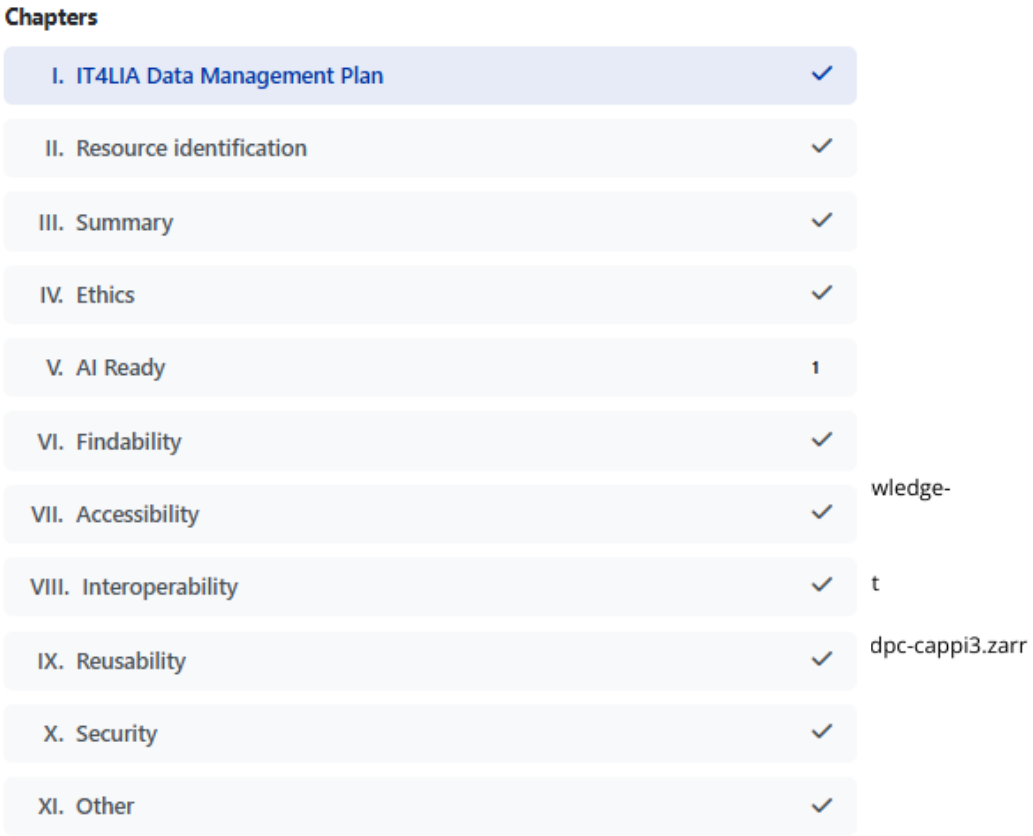


DS Wizard

Data Stewardship Wizard è uno strumento che rende più efficiente il processo di onboarding di dati.

Ai Data Steward e Chief Librarian di IT4LIA si affiancano uno o più Data Steward **per ogni partner**, formati ad hoc per l'utilizzo coerente della piattaforma.

DSW
DATA STEWARDSHIP WIZARD



DS Wizard

FBK_it4lia/IT-DPC-CAPPI3_DATASET

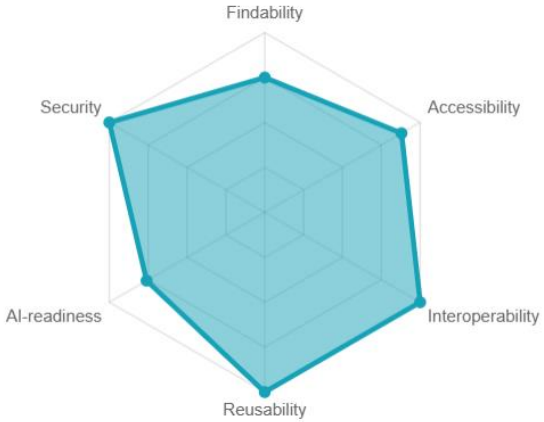
Questionnaire | Metrics | Preview | Documents | Settings

Summary Report

Answered (current phase): 77/79

Answered: 161/178

Metrics	Measure
Findability	0.75
Accessibility	0.88
Interoperability	1.00
Reusability	1.00
AI-readiness	0.76
Security	1.00



s

is

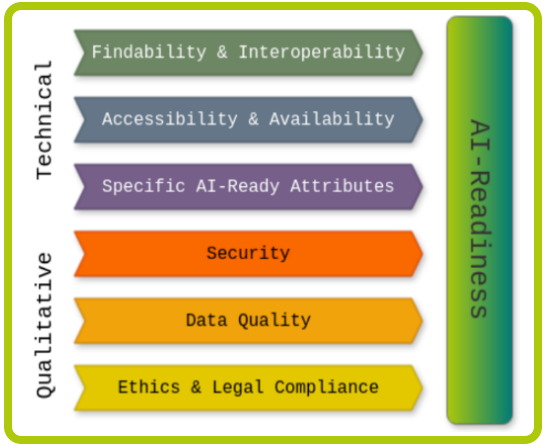
nt, training, validation, and deployment of AI models. AI readiness also takes into
ystem risk as defined in the AI Act.

DSW
DATA STEWARDSHIP WIZARD

La AI-Readiness

Il **Data Management Plan** di IT4LIA non si limita ad allinearsi ai principi FAIR.

AI-Readiness come **estensione** delle pratiche FAIR e dei Data Management Plan tradizionali.

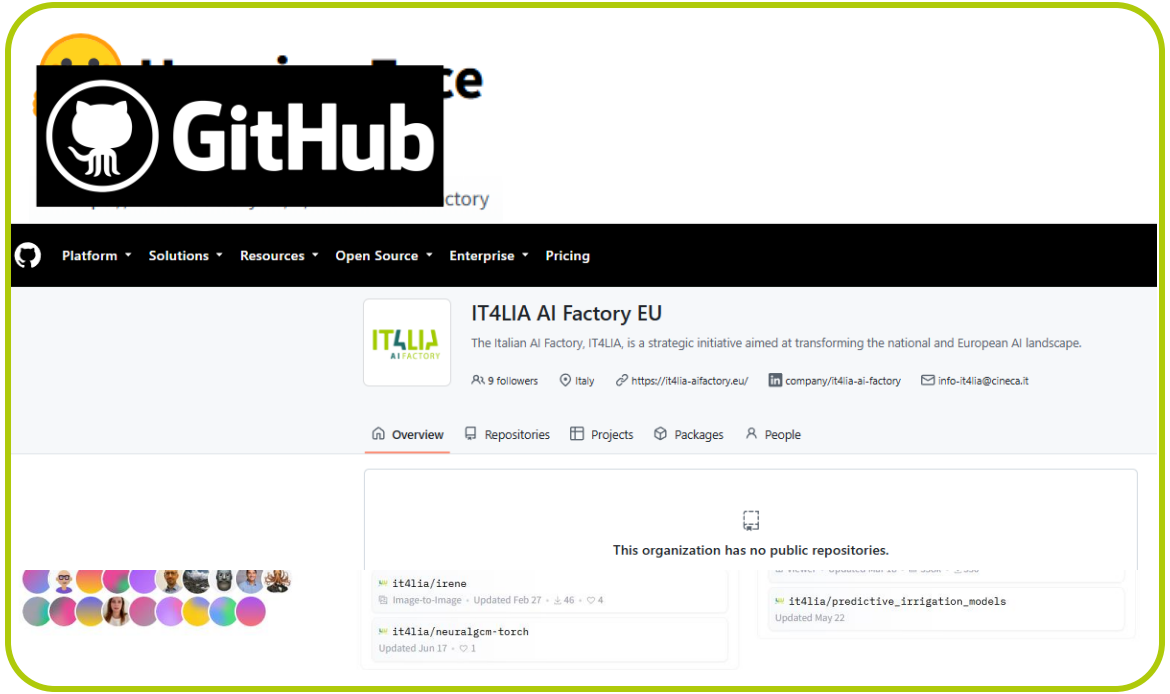


V. AI Ready	1	Is the language of the asset or its annotations specified? (y/n/NA language-indipendent asset)
	1.a.1	Specify the language(s) of the data or annotations (yes)
	2	What is the availability of the asset? (online / offline / not available)
	3	Is the asset structured into subsets (e.g. training, validation, test)? (y/n)
	3.a.1	Specify the data splits within the asset (yes)
	3.a.1.a.1	Subset (e.g. training, validation or test) (yes)
	3.a.1.a.2	Percentage (%) (yes)
	4	Is the provenance (lineage) of the data or components of the asset documented? (y/n)
	4.a.1	Describe the transformations, processing steps, or workflows applied to the data (including those resulting from IT4LIA services)
	5	Is the asset versioned? (y/n)
	5.a.1	Specify the asset version identifier (yes)
	6	Is information on the asset creation and updates available? (y/n)
	6.a.1	Provide the asset temporal information: (yes)
	6.a.1.a.1	Date created (yes)
	6.a.1.a.2	Last updated (yes)
	7	Are annotations available for the data or components of the asset? (y/n/Not applicable (annotations are not required for this type of asset)
	7.a.1	Specify the type(s) of annotations (sentiment labels, bounding boxes etc) (yes)
	7.a.2	Are annotation tools used for generating them? (yes) (y/n)
	7.a.2.a.1	Specify the annotation tools used: (yes)
	7.a.2.a.1.a.1	Tool name (yes)
	7.a.2.a.1.a.2	Description (yes)
	8	Are target AI tasks specified for the asset?
	8.a.1	Specify the target AI tasks for the asset
	8.a.1.a.1	AI Task (classification, NER…)
	9	Is ground truth data available for the asset? (y/n)
	9.a.1	Describe the available ground truth data (yes)
	10	Are quality assessment results available for the asset? (y/n)
	10.a.1	Describe the quality assessment results and performance metrics
	11	Are bias mitigation strategies applied to the asset? (y/n/Not applicable (no particular bias mitigation strategies are required for this type of asset))
	11.a.1	Describe the bias mitigation strategies applied (e.g. data balancing techniques…) (yes)
	12	Are AI model cards available for the asset or its associated models?
	12.a.1	Provide references or links to the available AI model cards
	13	type of asset))
	13.a.1	Describe how concept drift is monitored and managed over time (yes)
	14	What is the structure of the features or components of the asset? (raw/ vectorised/ both available/ NA)
	14.b.(c).1	Specify the type(s) of vectorisation applied (e.g. tf-idf, embeddings…) (vectorised/both)
	15	Are notebooks available for data management and processing? (y/n)
	15.a.1	Provide the available notebooks for data management (yes)
	15.a.1.a.1	Name or description of the notebook (yes)
	15.a.1.a.2	Reference (URL or repository link) (yes)
	16	Is the AI risk context of the asset defined?
	16.a.1	across multiple risk categories)
	16.a.2	Describe how the asset relates to the identified AI risk context(s) (e.g. potential high-risk use cases, etc)

Repositories IT4LIA

IT4LIA AI Factory promuove l'**open innovation**, fornendo accesso a repository in cui ricercatori, imprese e pubblica amministrazione possono **scoprire, riutilizzare e contribuire a fornire risorse per l'AI**.

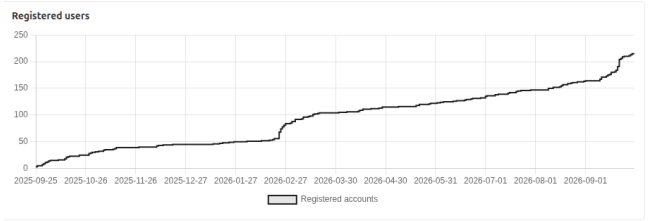
Attraverso **GitHub**, **Hugging Face** e ArcoDataHub, IT4LIA supporta la **diffusione di soluzioni scalabili, sicure e interoperabili**.

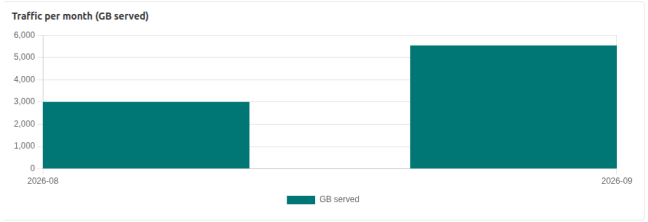


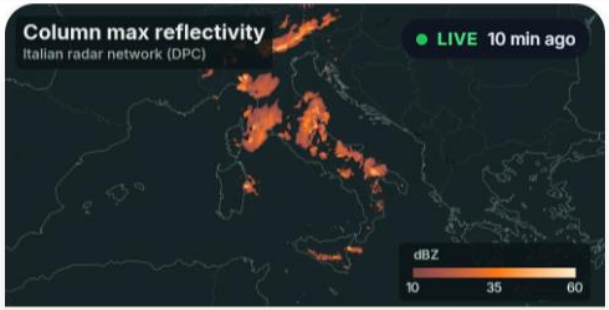
Le Repositories IT4LIA: ArcoDataHub

Un servizio cloud per accedere a **dataset aggiornati in tempo reale**, disponibili in formato **Analysis-Ready** and **Cloud-Optimized** (**ARCO**). Gli stessi dataset sono accessibili sull'infrastruttura HPC di IT4LIA e sulle Cloud VM per attività di analisi, addestramento di modelli ed esplorazione dei dati.

ArcoDataHub







italian-radar-dpc-vmi.zarr

Italian Weather Radar (DPC)

Live-updated Weather Radar Vertical Maximum Intensity (VMI) Composite reflectivity product (dBZ), covering the Italian peninsula from 2018 until now.

216 REGISTERED USERS	52 NEW THIS MONTH	14 ACTIVE ON 30 SEP (DAU)
36 ACTIVE 7 DAYS (WAU)	82 ACTIVE 30 DAYS (MAU)	5541.1 GB SERVED THIS MONTH

Comunità **continua espansione**.

Infrastrutture dati dinamiche, con aggiornamenti e nuovi caricamenti continui.

Contatti

Sito web

www.it4lia-aifactory.eu

Social

IT4LIA AI Factory

IT4LIA AI Factory

Scannerizza il QR Code

Compila il form

Ricevi la brochure e rimani aggiornato su IT4LIA

Donatella Sforzini d.sforzini@cineca.it



Italy for Artificial Intelligence

IT4LIA AI FACTORY is funded by the Ministry of University and Research and the National Cybersecurity Agency. The initiative is co-funded by the European Commission through the European High Performance Joint Undertaking under Grant Agreement No. 101234224. Additional contributions come from Austria, Slovenia, as well as from Cineca, the Emilia-Romagna Region, the Italian National Institute for Nuclear Physics, the Italian Institute of Artificial Intelligence for Industry, Fondazione Bruno Kessler, ad the ItaliaMeteo Agency.