

FAIRDATA repository @ INFN-LNGS: sfide, criticità ed opportunità

Francesca Marchegiani*, Stefano Stalio, Sandra Parlati

Laboratori Nazionali del Gran Sasso (LNGS), Istituto Nazionale di Fisica Nucleare ROR: <https://ror.org/02s8k0k61>

*ORCID: 0000-0002-2712-560X, email: francesca.marchegiani@lngs.infn.it

DOI: 10.15161/oar.it/9h7xj-s0683

Abstract

Nell'ambito del **progetto PNRR LNGS-FUTURE** (IR0000024) e, in particolare, delle attività previste nel Work Package 3.2, è stata sviluppata presso i Laboratori Nazionali del Gran Sasso (LNGS) una piattaforma software pilota per la gestione, conservazione, condivisione e riutilizzo dei **dati della ricerca**, in linea con i **principi FAIR**.

Questa piattaforma ha l'obiettivo di supportare i ricercatori nella gestione dei dati durante l'intero ciclo di vita del progetto, anche dopo la sua conclusione. I dati della ricerca, intesi come tutte le informazioni necessarie a sostenere, convalidare o riprodurre i risultati di un progetto scientifico, comprese le informazioni contestuali e i dati secondari, devono infatti essere conservati e, ove possibile, resi accessibili, al fine di promuovere una scienza aperta, riproducibile, trasparente e affidabile.

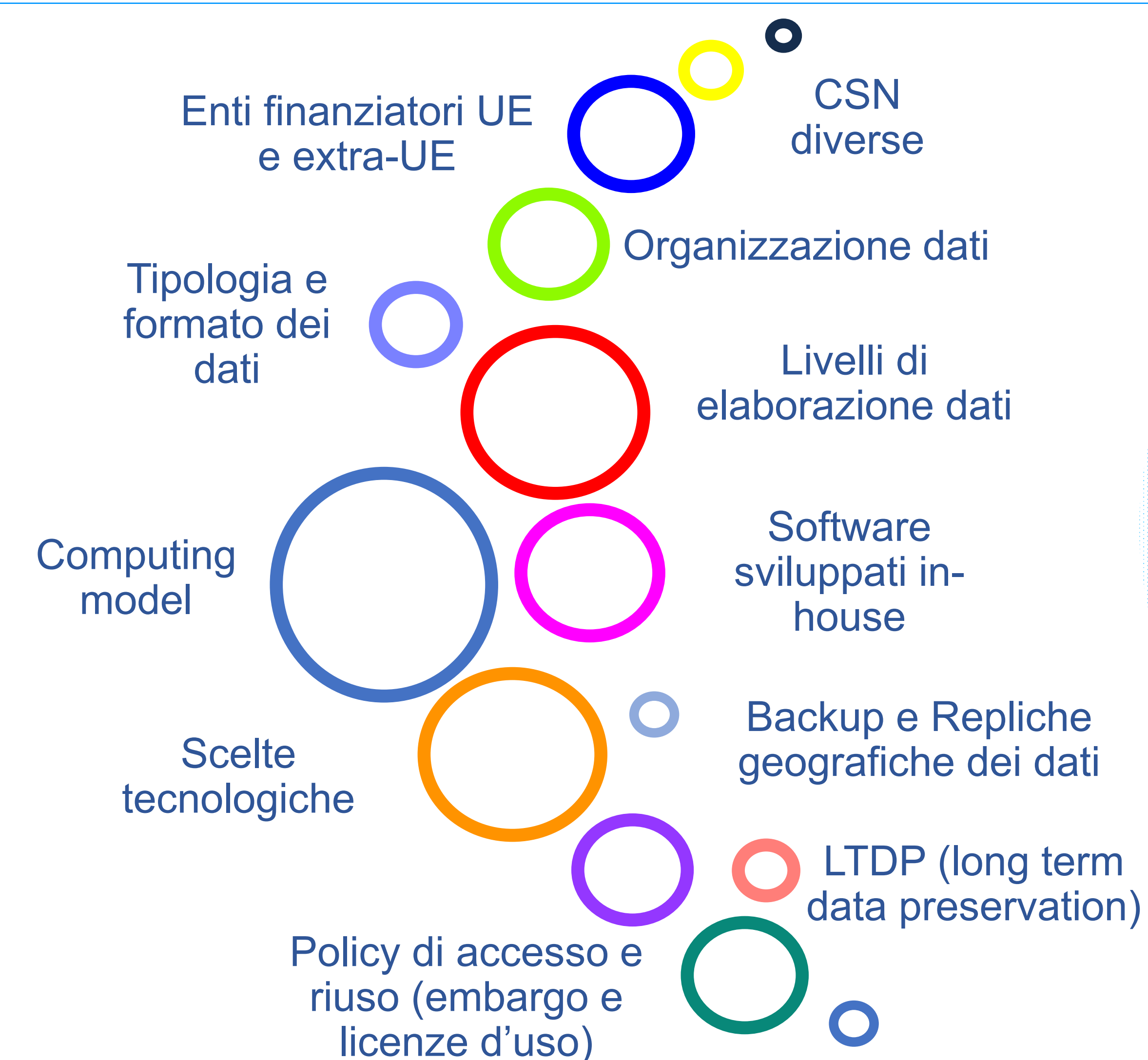
L'implementazione di questa infrastruttura ha il duplice scopo di **valorizzare i dati** come patrimonio della ricerca scientifica, essendo spesso unici, costosi e irripetibili, e di garantirne la **rintracciabilità**, la **disponibilità a lungo termine** e il riuso, in linea con i principi dell'Open Science. Inoltre, essa intende gettare le basi per una gestione accurata e responsabile dei dati prodotti dai grandi esperimenti ospitati presso i laboratori sotterranei dei LNGS.

2022 Proposta e Approvazione progetto PNRR LNGS-Future

2025 FAIRDATA repository @ LNGS

FASE DI RICOGNIZIONE

- Esperimenti ospitati @ LNGS e Servizi
- Software open-source
- Intervistati i responsabili della gestione dei dati di vari esperimenti, anche già conclusi, per avere feedback su esigenze specifiche
- PNSA asse di intervento 2 – Dati della ricerca



SISTEMI ETEROGENEI e COMPLESSI

- Definizione di **vocabolari** e ontologie da integrare nella piattaforma, nonché le licenze da associare ai diversi prodotti della ricerca.
- Necessaria redazione di un **DMP di laboratorio** (redatti DMP dei 4 laboratori nazionali INFN)
- **Metadati, licenze, identificativi persistenti** → visibilità, reperibilità ed impatto scientifico della propria ricerca

Customizzazione dei metadati



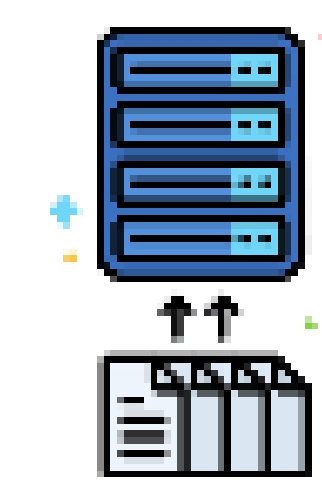
• Vocabolari controllati fit-for-purpose

- Affiliazione (nome istituzione)
- Nome + ORCID (autore + PID)
- ROR (identificativo univoco istituzione)
- **Subjects:** NASA Thesaurus, arXiv categories, CHMO (vocabolario metodi chimici), CESSDA, settori ERC, esperimenti e progetti LNGS

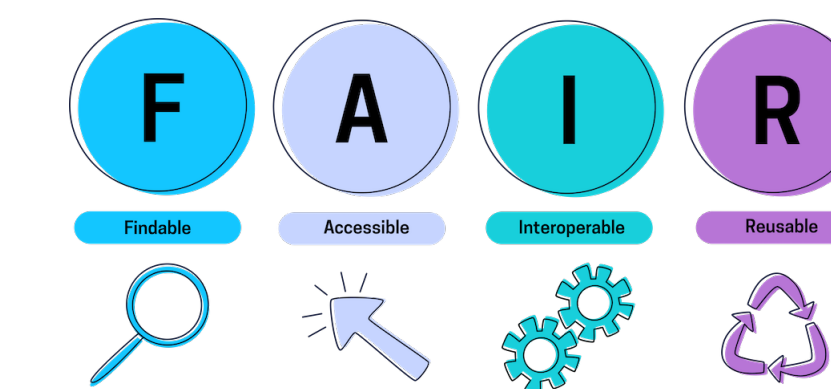
• Custom fields

- INFN experiment /project
- APC (article processing charges)
- Software

Piattaforma pilota FAIRDATA in produzione (<https://fairdata.lngs.infn.it>)



- Basata sul software open source InvenioRDM (versione 12.0)
- Interoperabile con l'archivio istituzionale dell'INFN Open Access Repository tramite il protocollo OAI-PMH
- Assegnazione di un DOI
- Possibile il deposito di tutti i prodotti della ricerca: dataset, immagini, audio, video, presentazioni, software, pubblicazioni, metadata only records. Popolata con dati di radiopurezza dei servizi *Tecniche Speciali* e *Chimica* dei LNGS, e alcuni dataset provenienti dalla rete di sensori per il monitoraggio ambientale dei laboratori sotterranei.
- **Vocabolari controllati specifici per disciplina**
- **Storage ad oggetti di tipo S3 con replica remota** ed estremamente scalabile, considerando che in futuro i volumi di dati coinvolti potrebbero essere dell'ordine dei petabyte.



Gestione campo autore e affiliazione

- In corso collaborazione con il gruppo INFN-DSI e il gruppo OpenScience per la creazione di un **Authority file** per gli autori INFN (intervento I. Piergentili DOI: 10.15161/oar.it/58e8g-jzk49)

Garantire la qualità dei metadati del repository